Lectures on Abstract Algebra

Preliminary Version

Richard Elman

Department of Mathematics,
University of California,
Los Angeles, CA 90095-1555, USA

Contents

Part 1. Preliminaries 1
1. Introduction 3

Chapter I. The Integers 9
2. Well-Ordering and Induction 9
3. Addendum: The Greatest Integer Function 13
4. Division and the Greatest Common Divisor 17

Chapter II. Equivalence Relations 27
5. Equivalence Relations 27
6. Modular Arithmetic 31
7. Surjective maps 36

Part 2. Group Theory 39

Chapter III. Groups 41
8. Definitions and Examples 41
9. First Properties 48
10. Cosets 53
11. Homomorphisms 57
12. The First Isomorphism Theorem 62
13. The Correspondence Principle 66
14. Finitely Generated Abelian Groups 70
15. Addendum: Divisible Groups 78
16. Addendum: Finitely Generated Groups 82
17. Series 85
18. Free Groups 90

Chapter IV. Group Actions 101
19. The Orbit Decomposition Theorem 101
20. Addendum: Finite Rotation Groups in $\mathbb{R}^3$. 104
21. Examples of Group Actions 107
22. Sylow Theorems 115
23. Addendum: Finite Solvable Groups 123
24. The Symmetric and Alternating Groups 129
25. Addendum: The Projective Special Linear Group 139

Part 3. Ring Theory 145

Chapter V. General Properties of Rings 147
26. Definitions and Examples 147
27. Factor Rings and Rings of Quotients 157
28. Zorn's Lemma 165
29. Localization 172

Chapter VI. Domains 177
30. Special Domains 177
31. Characterization of UFDs 185
32. Gaussian Integers 186
33. Addendum: The Four Square Theorem 192

Chapter VII. Polynomial Rings 199
34. Introduction to Polynomial Rings 199
35. Polynomial Rings over a UFD 205
36. Addendum: Polynomial Rings over a Field 210
37. Addendum: Algebraic Weierstraß Preparation Theorem 212

Part 4. Module Theory 217

Chapter VIII. Modules 219
38. Basic Properties of Modules 219
39. Free Modules 230

Chapter IX. Noetherian Rings and Modules 239
40. Noetherian Modules 239
41. Hilbert's Theorems 242
42. Addendum: Affine Plane Curves 249

Chapter X. Finitely Generated Modules Over a PID 253
43. Smith Normal Form 253
44. The Fundamental Theorem 258
45. Canonical Forms for Matrices 269
46. Addendum: Jordan Decomposition 281
47. Addendum: Cayley-Hamilton Theorem 284

Part 5. Field Theory 287

Chapter XI. Field Extensions 289
48. Algebraic Elements 289
49. Addendum: Transcendental Extensions 298
50. Splitting Fields 300
51. Algebraically Closed Fields 309
52. Constructible Numbers 312
53. Separable Elements 321

Chapter XII. Galois Theory 325
54. Characters 325

55. Computations 331
56. Galois Extensions 336
57. The Fundamental Theorem of Galois Theory 344
58. Addendum: Infinite Galois Theory 351
59. Roots of Unity 357
60. Radical Extensions 368
61. Addendum: On Hilbert Theorem 90 379
62. Addendum: Kummer Theory 381
63. Normal Basis Theorem 384
64. Addendum: Galois' Theorem 386
65. The Discriminant of a Polynomial 389
66. Purely Transcendental Extensions 393
67. Finite Fields 395
68. Addendum: Jacobson's Theorem 399
69. Addendum: Hilbert Irreducibility Theorem 402

Chapter XIII. Transcendental Numbers 415
70. Liouville Numbers 415
71. Transcendence of e 418
72. Symmetric Polynomials 420
73. Transcendence of π 422
74. Gelfond-Schneider Theorem 431

Chapter XIV. The Theory of Formally Real Fields 445
75. Orderings 445
76. Extensions of Ordered Fields 447
77. Characterization of Real Closed Fields 453
78. Hilbert's 17th Problem 457

Part 6. Commutative Algebra and Algebraic Number Theory 463

Chapter XV. Dedekind Domains 465
79. Integral Elements 465
80. Integral Extensions of Domains 469
81. Dedekind Domains 471
82. Extension of Dedekind Domains 478
83. Hilbert Ramification Theory 483
84. The Discriminant of a Number Field 486
85. Dedekind's Theorem on Ramification 490
86. The Quadratic Case 492
87. Addendum: Valuation Rings and Prüfer Domains 496

Chapter XVI. Algebraic Number Fields 505
88. Ideal and Counting Norms 505
89. Lattices in Number Fields 508
90. Units in a Ring of Algebraic Numbers 515

91. Minkowski Bound 520

Chapter XVII. Introduction to Commutative Algebra 527
92. Zariski Topology 527
93. Integral Extensions of Commutative Rings 536
94. Primary Decomposition 545
95. Addendum: Associated Primes of Modules 552
96. Akizuki and Krull-Akizuki Theorems 557
97. Affine Algebras 563
98. Regular Local Rings 575
99. Addendum: Fibers 587
100. Addendum: Japanese Rings 589
101. C_n-fields 590

Part 7. Semisimple Algebras and Representation Theory 597

Chapter XVIII. Division and Semisimple Rings 599
102. Wedderburn Theory 599
103. The Artin-Wedderburn Theorem 606
104. Finite Dimensional Real Division Algebras 610
105. Cyclic Algebras 611
106. Central Simple Algebras 615
107. The Brauer Group 626
108. Polynomial Rings over a Division Algebra 638

Chapter XIX. Introduction to Representation Theory 645
109. Representations 645
110. Split Group Rings 651
111. Addendum: Hurwitz's Theorem 653
112. Characters 657
113. Orthogonality Relations 659
114. Burnside's $p^a q^b$ Theorem 666
115. Addendum: Schur's Theorem 669
116. Induced Representations 671
117. Torsion Linear Groups 676

Part 8. Homological Algebra and Category Theory 683

Chapter XX. Universal Properties and Multilinear Algebra 685
118. Some Universal Properties of Modules 685
119. Tensor Products 690
120. Tensor, Symmetric, and Exterior Algebras 696
121. The Determinant 703

Chapter XXI. Introduction to Homological Algebra 711
122. Homology 711
123. Hom 717

124. Injective Modules 721
125. Ext 726
126. Projective Modules 738
127. Projective Modules over Commutative Rings 743
128. Ext II 750
129. Tensor Product Revisited 759
130. Limits 763
131. Flat Modules 769
132. Tor 773
133. Regular Local Rings II 776

Chapter XXII. Categories 789
134. Categories 789
135. Functors 793
136. Yoneda's Lemma 799
137. Adjoints 806
138. Limits 810
139. Additive and Abelian Categories 821
140. Derived Functors 827

Appendices 835

Appendix A. Axiom of Choice and Zorn's Lemma 837

Appendix B. Bertand's Hypothesis 841

Appendix C. Matrix Representations 849

Appendix D. Smith Normal Form over a Euclidean Ring 853

Appendix E. Symmetric Bilinear Forms 861

Appendix F. Primitive Roots 867

Appendix G. The Sign of the Gauss Sum 869

Appendix H. Pell's Equation 873

Bibliography 875

Notation 879

Index 883

Part 1

Preliminaries

1. Introduction

In this section, we introduce some of the notations and definitions that we shall use throughout this book. We shall do it by investigating a few mathematical statements. Consider the following:

Statements 1.1.

(1) The integer $2^{43112609}-1$, which has 12 978 189 digits, is a prime.
(2) There exist infinitely many (rational) primes.
(3) Let
$$\pi(x) := \text{the number of positive primes} \leq x.$$
Then
$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1.$$
(4) $\sqrt{2} \notin \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers.
(5) The real number π is not "algebraic" over $\mathbb{Q}$.
(6) There exist infinitely many real numbers (even "many more" than the number of elements in $\mathbb{Q}$) not algebraic over $\mathbb{Q}$.

To look at these statements, we need some definitions.

Let $a, b \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. We say that a *divides* b (in $\mathbb{Z}$) if
$$\text{there exists an integer } n \text{ such that } b = an\text{, i.e., if } a \neq 0\text{, then } \frac{b}{a} \in \mathbb{Z}.$$
We write
$$a \mid b \ \ (\text{in } \ \mathbb{Z}).$$

For example, $3 \mid 12$ as $12 = 4 \cdot 3$ but $5 \nmid 12$ in $\mathbb{Z}$, where $\nmid$ means does not divide.

An integer p is called a *prime* if $p \neq 0, \pm 1$ and
$$n \mid p \ \text{ in } \ \mathbb{Z} \text{ implies that } n \in \{1, -1, p, -p\}, \textit{ i.e.}, \ n = \pm 1, \pm p.$$
For example, $2, 3, 5, 7, 11, \ldots$ are prime. [The prime 2 is actually the "oddest prime of all"!] We shall need to know below that every integer $n > 1$ is divisible by some prime. (For the full statement see Theorem 1.3 below which we shall prove in Theorem 4.16.)

A complex number x is called *irrational* if x is not a rational number, i.e., $x \in \mathbb{C} \setminus \mathbb{Q} := \{z \in \mathbb{C} \mid z \notin \mathbb{Q}\}$, where $\mathbb{C}$ is the set of complex numbers. A complex number x is called *algebraic* (over $\mathbb{Q}$) if there exists a nonzero *polynomial* $f(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0$ with $a_0, \ldots, a_n \in \mathbb{Q}$ (some n) not all zero satisfying $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$. We shall write f for $f(t)$ where t *always represents a variable* and $f(x)$ means plug x into f. We let
$$\mathbb{Q}[t] := \text{ the set of polynomials with rational coefficients.}$$
So x is algebraic over $\mathbb{Q}$ if there exists $0 \neq f \in \mathbb{Q}[t]$ such that $f(x) = 0$. A complex number x that is not algebraic (over $\mathbb{Q}$) is called *transcendental* (over $\mathbb{Q}$), so $x \in \mathbb{C}$ is transcendental if there exists no nonzero polynomial $f \in \mathbb{Q}[t]$ satisfying $f(x) = 0$.

With the above definitions, we can look at our statements to see which ones are interesting, deep, etc.

We start with Statement 1 above. An integer $2^{43112609} - 1$ with $12\,978\,189$ digits was shown to be prime by a UCLA team using a primality test of Lucas on Mersenne numbers. This was the first known prime to have at least ten million digits. It is, on the face of it, not very interesting. After all, it is analogous to saying 97 is a prime. However, it is interesting historically. We call an integer $M_n := 2^n - 1$, n a positive integer, a *Mersenne number* and a *Mersenne prime* if it is a prime. In 1644, Mersenne conjectured that for $n \leq 257$, the Mersenne number M_n is prime if and only if $n = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$. [If M_n is a prime, then n must be a prime as $2^{ab} - 1$ factors if $a, b > 1$.] It was shown in the 1880's that M_{61} is a prime. In 1903 Frank Cole showed that $M_{67} = 193707721 \cdot 761838257287$. [He silently multiplied out these two numbers on a blackboard at a meeting of the American Mathematical Society.] There were three more errors to Mersenne's conjecture: M_{89} and M_{107} are prime and M_{257} is *composite*, i.e., not $0, \pm 1$, or a prime. The correct solution of Mersenne's conjecture is that M_n is a Mersenne prime if and only if $n = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127$ was completed in 1947.

Mersenne was interested in these numbers because of their connection to a study in antiquity. An integer $n > 1$ is called *perfect* if

$$n = \sum_{\substack{d \mid n \\ 0<d<n}} d \quad \text{or} \quad 2n = \sum_{\substack{d \mid n \\ 0<d\leq n}} d,$$

e.g., $6 = 1 + 2 + 3$. The first equation says that n is the sum of its *aliquot divisors*. A theorem in elementary number theory says:

Theorem 1.2. (Euclid/Euler) *An even number N is perfect if and only if $N = \frac{1}{2}p(p+1)$ with p a Mersenne prime (so $N = 2^{n-1}(2^n - 1)$ with $p = 2^n - 1$).*

It is still an open problem whether there exist infinitely many even perfect numbers (or equivalently, infinitely many Mersenne primes). It is also an open problem whether there exist any odd perfect numbers.

Statement 2 is very interesting and is due to Euclid. It may be the first deep mathematical fact that one learns. The proof is quite simple. If the result is false, let $p_1, \dots, p_n$ be a complete list of (positive) primes and set $N = p_1 \cdots p_n + 1$. We use (cf. Exercise 1.13(1)) that says:

$$\text{If } x \mid y \text{ and } x \mid z \text{ in } \mathbb{Z}, \text{ then for all } a, b \in \mathbb{Z}, \text{ we have } x \mid ay + bz,$$

i.e., if an integer divides two integers, then it divides any *$\mathbb{Z}$-linear combination* of those two numbers. From this, we conclude that if $p_i \mid N$ then $p_i \mid N - p_1 \cdots p_n$, so $p_i \mid 1$, which is impossible. But as mentioned above, every integer greater than one is divisible by a prime. This means that there exists a prime dividing N different from any of $p_1, \dots, p_n$, a contradiction.

Statement 3 is a deep theorem called the *Prime Number Theorem* or *PNT*. It quantifies Statement 2. It was first conjectured by Gauss and proven independently in 1896 by de la Vallée Poussin and Hadamard using complex analysis and fundamental work of Riemann. One shows that the *zeta function*, $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$, where s is a complex variable, has no root on the line $\mathrm{Re}(s) = 1$, i.e., $\zeta(1+\sqrt{-1}t)$ is never zero, and that this is equivalent to the PNT. [The most famous problem in mathematics is the Riemann Hypothesis, which says the analytic continuation of the zeta function for $\mathrm{Re}(s) > 0$ has all its roots on the line $\mathrm{Re}(s) = 1/2$.] An "elementary" proof in number theory means that it does not use complex analysis. An elementary proof may be very difficult. An elementary proof of PNT was discovered in 1949 by Selberg and Erdös. It is much harder than the non-elementary proof.

Statement 4 is not very interesting, although the Pythagoreans knew it but were afraid to leak the knowledge, fearing catastrophe if it got out. It did. They were apparently right. To prove this, we need not only positive integers $n > 1$ factor as a product of primes but that the product is essentially unique. More specifically, we need:

Theorem 1.3. (The Fundamental Theorem of Arithmetic) *Every integer $n > 1$ is a product of positive primes unique up to order, i.e., there exist unique primes $1 < p_1 < \cdots < p_r$ and unique integers $e_1, \ldots, e_r > 0$ such that*

$$n = p_1^{e_1} \cdots p_r^{e_r}. \tag{*}$$

We call (*) the *standard representation* or *standard factorization* of n. We show Statement 4 assuming that this theorem is true. [Euclid knew its proof and we shall prove it in Theorem 4.16.]

If $\sqrt{2}$ is rational, then $\sqrt{2} = \frac{m}{n}$ for some integers m, n with $n \neq 0$. This means that we have an equation of integers $2n^2 = m^2$. But the prime two occurs on the left hand side to an odd power and on the right hand side to an even power, contradicting the uniqueness part of the Fundamental Theorem of Arithmetic. Thus $\sqrt{2}$ is *irrational*, i.e., a complex number that is not rational.

Statement 5 was proven by Lindemann in 1882. This is historically interesting as it solves the famous Greek construction problem, "squaring the circle", which asks whether one can construct a circle with the same area as a given square using only a straight-edge and compass (according to specific rules). Its proof is not easy. One must reduce this geometric problem to an algebraic one in field theory, which is then proven using analysis. A proof is given in Section 73. It is easier to show that the real number e is transcendental. (Cf. Section 71.) The first real number shown to be transcendental was the number $\sum_{n=1}^{\infty} \frac{1}{10^{n!}}$ shown to be transcendental by Liouville [cf. Section 70]. [That it is transcendental is essentially due to the fact that this series converges much too rapidly.] An even deeper result, solved independently by Gelfond and Schneider [which we shall prove in Section 74], is

Theorem 1.4. (Hilbert's Seventh Problem) *Let α and β be algebraic over $\mathbb{Q}$ with $\alpha \neq 0, 1$ and β irrational. Then α^β is transcendental. In particular $\sqrt{2}^{\sqrt{2}}$ is transcendental.*

This theorem also proves that e^π is transcendental, since $e^\pi = (e^{\sqrt{-1}\pi})^{-\sqrt{-1}} = (-1)^{-\sqrt{-1}}$. It is unknown whether π^e is transcendental.

Lindemann actually proved that if α is algebraic and nonzero, then e^α is not algebraic. Since Euler showed that $-1 = e^{\sqrt{-1}\pi}$, it follows that $\sqrt{-1}\pi$ is not algebraic; and from this it follows that π is transcendental, using the fact that the product of algebraic numbers is algebraic, which we prove in Theorem 48.20.

Statement 6 is not as interesting as Statement 5 for the word "transcendental" is a misnomer. However, the proof of this by Cantor changed mathematics and caused much philosophical distress. Indeed, Cantor showed that there were "many more" reals than rationals. What does this mean?

Let A and B be two sets. [What is a set?] We say that A and B have the same *cardinality* and write $|A| = |B|$ if there exists a one-to-one and onto function, i.e., a *bijection*, $f : A \to B$. We call such an f a *bijective function.* Recall that a function or *map* $f : A \to B$ is one-to-one *or injective* if

$$f(a_1) = f(a_2) \quad \Longrightarrow \quad a_1 = a_2$$

(where $\Longrightarrow$ means *implies*). Let $f(A) := \{f(a) \mid a \in A\}$ denote the *image* of A. Then f is one-to-one is if and only if

$$f^{-1} : f(A) \to A \text{ given by } f(a) \mapsto a$$

is a function. We say f is onto or *surjective* if $B = f(A)$. [We also write a function as $A \xrightarrow{f} B$.]

Review 1.5. You should review the definitions of functions, inverses, etc. that we assume you have learned.

Examples 1.6.
1. $\{1, 2, \ldots, 26\}$ and $\{a, b, \ldots, z\}$ have the same cardinality.
2. $\mathbb{Z}$ and $2\mathbb{Z} := \{2n \mid n \in \mathbb{Z}\}$ have the same cardinality, as $f : \mathbb{Z} \to 2\mathbb{Z}$ given by $n \mapsto 2n$ is a bijection.
3. Let $\mathbb{Z}^+$ denote the set of positive integers and $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$ the set of non-negative integers. Then $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{Z}^+$ all have the same cardinality. Indeed, the maps
$$f : \mathbb{Z} \to \mathbb{N} \text{ by } n \mapsto \begin{cases} 2n - 1 & \text{if } n > 0 \\ -2n & \text{if } n \leq 0 \end{cases}$$
and $g : \mathbb{N} \to \mathbb{Z}^+$ by $n \mapsto n + 1$ are both bijections.

If A is a set, we call the symbol $|A|$ the *cardinality* of A and interpret it to mean the "number of elements" in A. For example, if A is *finite*, i.e., the set A has finitely many elements [we write $|A| < \infty$], then $|A|$ is the number of elements in A, e.g., we have $|\{1, \ldots, 26\}| = 26$. If A is not finite, we of course say that A is *infinite.* If a set A satisfies $|A| = |\mathbb{Z}|$, i.e., there exists a bijection $f : \mathbb{Z} \to A$ (equivalently, there exists a bijection $g : A \to \mathbb{Z}$), we say A is *countable.* (Some authors call such a set *countably infinite or denumerable.*)

The following facts can be shown [and we leave their proofs as exercises]:

Fact 1.7.
1. *A subset of a countable set is either finite or countable.*
2. $\mathbb{Q}$ *is countable.*
3. *Let* $\mathbb{C}$ *be the set of complex numbers. Then* $\{\alpha \in \mathbb{C} \mid \alpha \text{ algebraic }\}$, *the set of complex algebraic numbers, is countable.*

Theorem 1.8. (Cantor) *The set* $\mathbb{R}$ *of real numbers is not countable.*

PROOF. Suppose that $\mathbb{R}$ is countable. As $\mathbb{R}$ and the closed interval $[0,1]$ have the same cardinality by Exercise 1.13(11), the interval $[0,1]$ must also be countable. As $\mathbb{Z}^+ \subset \mathbb{Z}$ is not finite, it is also countable by the first fact (or by Example 1.6(3)). Therefore, there exists a bijection $f : \mathbb{Z}^+ \to [0,1]$, i.e., the closed interval $[0,1]$ is covered by a sequence. Write each $f(n)$ in (base ten) decimal form, say

$$\begin{aligned} f(1) &= .a_{11}a_{12}a_{13}\cdots a_{1n}\cdots \\ f(2) &= .a_{21}a_{22}a_{23}\cdots a_{2n}\cdots \\ &\ \vdots \\ f(n) &= .a_{n1}a_{n2}a_{n3}\cdots a_{nn}\cdots \\ &\ \vdots\ . \end{aligned}$$

For each n choose $b_n \in \{1,\dots,8\}$ with $b_n \neq a_{nn}$. [We omit 0 and 9 to prevent round off problems.] Let $b = b_1b_2\cdots b_n\cdots$. Then b and $f(n)$ differ in the nth digit as $b_n \neq a_{nn}$. Moreover, as $b_n \neq 0, 9$, the numbers b and $f(n)$ cannot be the same real number. Thus $b \neq f(n)$ for all $n \in \mathbb{Z}^+$, so f is not onto, a contradiction. (We have really used the fact that $\mathbb{R}$ is *complete*, i.e., that b is a real number.) □

Warning 1.9. Sets can be tricky – again we ask what is a set?

Indeed, consider the statement:

Statement 1.10. There is a universe, i.e., a set that contains all other sets.

Now you should consider the following:

Contemplate 1.11. (Russell's Paradox) Let

$$A := \{B \text{ is a set} \mid B \notin B\}.$$

Is $A \in A$? Is $A \notin A$?

What can you conclude?

Remark 1.12. Throughout these notes we shall assume the Axiom of Choice which can be found in Appendix A(A.8) and equivalent formulations of it. In particular, this allows us to define when two sets have the same cardinality in a useful way, called the Schroeder-Bernstein Theorem which says two sets A and B have the same cardinality if and only if there exist injective maps $f : A \to B$ and $g : B \to A$ (The proof can be found in Appendix A.13.)

Exercises 1.13.

1. Show if a nonzero integer a divides integers b and c, it divides $bx + cy$ for any integers x and y.
2. Show if $a^n - 1$ is a prime with $n > 1$, then $a = 2$ and n is a prime. [The converse is false as $23 \mid M_{11}$.]
3. If $2^n + 1$ is a prime, what can you say about n? Prove your assertion.

4. Let $f(t)$ be a polynomial with integer coefficients. Show that the set of integers $\{f(n) \mid 0 < n \in \mathbb{Z}\}$ contains infinitely many distinct prime divisors. (You may assume that nonzero polynomials with coefficients in $\mathbb{Z}$ have finitely many roots.)
5. Let $a, b, n \in \mathbb{Z}$ with $n > 1$ and $b \neq 0$. Determine when $n^{\frac{a}{b}}$ is a rational number. Prove your determination. (You can use the Fundamental Theorem of Arithmetic.)
6. Assume that the real number $\sqrt{\pi}$ is transcendental. Show that π is transcendental.
7. Let $f : A \to B$ be a map of sets. Prove that f is injective if and only if given any set C and any two set maps $g_i : C \to A$, $i = 1, 2$, with compositions $f \circ g_1 = f \circ g_2$, then $g_1 = g_2$.
8. Let $f : A \to B$ be a map of sets. Prove that f is surjective if and only if given any set C and any two set maps $h_i : B \to C$, $i = 1, 2$, with compositions $h_1 \circ f = h_2 \circ f$, then $h_1 = h_2$.
9. Show a subset of a countable set is either countable or finite.
10. If the sets A and B are countable, show that the *cartesian product*

$$A \times B := \{(a, b) \mid a \in A,\ b \in B\}$$

is also countable. Using induction (to be discussed), show if

$$A_1, \ldots, A_n \text{ are countable so is } A_1 \times \cdots \times A_n.$$

[Hint: Show $\mathbb{Z} \times \mathbb{Z}$ (or $\mathbb{Z}^+ \times \mathbb{Z}^+$ where $\mathbb{Z}^+ := \{n \in \mathbb{Z} \mid n > 0\}$) is countable by drawing a big (= infinite) matrix. In a similar way one proves Fact 1.7(2). In fact, the cartesian product of countable sets is countable, which is needed to prove Fact 1.7(3). Can you prove this?]
11. The closed interval $[0, 1]$ and the set of all real numbers $\mathbb{R}$ have the same cardinality.
12. Any two (finite) line segments (so each has more than one point) have the same cardinality. [Hint: Draw the two line segments parallel to each other.]

CHAPTER I

The Integers

In this chapter, we investigate the basic properties of the set of integers. In particular, we show that every integer can be factored into a product of primes, unique up to order. To do so we introduce concepts that shall be generalized. We assume that the reader has seen proofs by induction. We begin the chapter with an equivalent form of induction called the Well-Ordering Principle and use it to establish facts about division of integers. In particular, we prove the division algorithm. Of great historical importance is the notion of prime and the property discovered by Euclid that characterizes a prime that became a cornerstone of modern algebra.

2. Well-Ordering and Induction

We denote the *empty set* by $\emptyset$. In this section, we are interested in induction which you should have seen and turns out to be equivalent to the following:

The Well-Ordering Principle 2.1. Let $\emptyset \neq S \subset \mathbb{Z}^+$. Then S contains a *least* (also called a *minimal*) element, i.e.,

$$\text{There exists an } a \in S \text{ such that if } x \in S, \text{ then } a \leq x.$$

This is an axiom that we accept as being true. You can visualize it by drawing the positive real line and ticking off the integers. You move to the right until you get to the first element in S.

One can modify the Well-Ordering Principle to the seemingly more general:

The Modified Well-Ordering Principle 2.2. Let $\emptyset \neq T \subset \mathbb{Z}$. Suppose that there exists an $N \in \mathbb{Z}$ such that $N \leq x$ for all $x \in T$, i.e., T is *bounded from below*. Then T contains a least element.

Proof. Let $f : \mathbb{Z} \to \mathbb{Z}$ be the map defined by $x \mapsto x + |N| + 1$ where here $|N|$ means the absolute value of N. Then f is a bijection. [What is f^{-1}?] Hence the *restriction* of f to the subset T, which we denote by $f|_T : T \to \mathbb{Z}$, is also injective. As $f(T) = f|_T(T) \subset \mathbb{Z}^+$, there exists a least element $f(a) \in f(T)$, $a \in T$, by the Well-Ordering Principle. Then a is the least element of T as f preserves $\leq$. (You should check this.) $\square$

Remark 2.3. In an analogous way, if $\emptyset \neq T \subset \mathbb{Z}$ is *bounded from above*, i.e., there exists an integer N satisfying $s \leq N$ for all $s \in T$, then T contains a *largest* (also called a *maximal*) element, i.e., an element $a \in T$ such that $x \leq a$ for all x in T.

We have the following simple applications of well-ordering.

Application 2.4. There exists no integer N satisfying $0 < N < 1$ (or strictly between any integers n and $n + 1$).

Proof. Let $S = \{n \in \mathbb{Z} \mid 0 < n < 1\}$. If $\emptyset \neq S$ then there exists a least element $N \in S$ by well-ordering, so $0 < N < 1$ which implies that $0 < N^2 < N < 1$ – can you prove this? – and $N^2 \in \mathbb{Z}$ contradicting the minimality of N. $\square$

Application 2.5. Let $P(n)$ be a statement that is true or false depending on $n \in \mathbb{Z}^+$. If $P(n)$ is not true for some n then by the Well-Ordering Principle, there exists a least element $n \in \mathbb{Z}^+$ such that $P(n)$ is false. We call $P(n)$ a *minimal counterexample*.

Application 2.6. Suppose that $S \subset \mathbb{Z}^+$ and $1 \in S$. If S satisfies the condition that whenever $n \in S$ also $n + 1 \in S$, then $S = \mathbb{Z}^+$.

Proof. Let

$$T = \mathbb{Z}^+ \setminus S = \{n \in \mathbb{Z}^+ \mid n \notin S\},$$

i.e., $\mathbb{Z}^+ = T \cup S$, the union of T and S, and $\emptyset = T \cap S$, the intersection of T and S. (We say that $\mathbb{Z}^+$ is the *disjoint union* of S and T, and we write $\mathbb{Z}^+ = S \vee T$.)

Suppose that $T \neq \emptyset$. Then by well-ordering there exists a least positive element $n \in T$. In particular, $n - 1 \notin T$. As $1 \notin T$, $n > 1$, so we have $n - 1 \in \mathbb{Z}^+$. Hence by minimality, $n - 1 \in S$. The hypothesis now implies that $n \in S$, a contradiction. Thus $T = \emptyset$ so $\mathbb{Z}^+ = S$. $\square$

The applications yield the following:

Application 2.7. (First Principle of Finite Induction) For each positive integer n, let $P(n)$ be a statement that is true or false. Suppose we know that

(1) (*Base Case*) $P(1)$ is true.

(2) (*Induction Step*) If $P(n)$ is true then $P(n + 1)$ is true.

Then $P(n)$ is true for all positive integers n.

The hypothesis in the induction step is called the *induction hypothesis*.

Often the following equivalent variant of the First Principle of Finite Induction is more useful. That it is equivalent is left as an exercise.

Application 2.8. (Second Principle of Finite Induction) For each positive integer n, let $P(n)$ be a statement that is true or false. Suppose that the assumption that $P(m)$ is true for all positive integers $m < n$ implies that $P(n)$ is true. Then $P(n)$ is true for all positive integers n.

Remark 2.9. If $n = 1$ then $\{m < 1 \mid m \in \mathbb{Z}^+\}$ is empty, so you still must show that $P(1)$ is true if you wish to use the Second Principle of Finite Induction, as your proof would fail for $n = 1$ otherwise. Remember that well-ordering needs a nonempty set.

Remark 2.10. Note that using the Modified Well-Ordering Principle, you can start induction at any integer and prove things from that point on, e.g., you can start at 0.

We assume that you have seen induction proofs in the past, e.g., in your linear algebra course. Induction proofs can be very complicated. We give such a complicated induction proof. Note in the proof the formal steps versus the mathematical ones.

Example 2.11. The product of any $n \geq 1$ consecutive positive integers is divisible by $n!$.

Proof. Let m and n be two positive integers and set

$$(m)_n := m(m+1)\cdots(m+n-1),$$

the product of n consecutive integers starting from m.

Claim 2.12. We have $n! \mid (m)_n$ for all positive integers m and n.

Of course, the claim is exactly what we want to prove. Note that there are two integers in the claim. We have

If $m \in \mathbb{Z}^+$ is arbitrary and $n = 1$ then $(m)_n = m$ and $n! = 1! \mid m$.

We assume

Induction Hypothesis I. The claim holds for fixed $n = N - 1$ and for all m.

This is the induction step on n, using N for clarity. As we know that this holds for $n = 1$ and for all m, we must show that the result holds for N and all m, i.e., we must show

$$(N-1)! \mid (m)_{N-1} \text{ for all } m \quad \Longrightarrow \quad N! \mid (m)_N \text{ for all } m.$$

Let $m = 1$. Then $(1)_N = N!$ and $N! \mid (1)_N$. Note that this is the first step of an induction on m. So we can assume

Induction Hypothesis II. The claim holds for fixed $n = N$ and $m = M$. (Notationally it is easier to use M rather than $M - 1$).

We must show that Induction Hypotheses I and II imply the result for $n = N$ and $m = M + 1$, i.e.,

$$\text{if } N! \mid (M)_N \text{ and } (N-1)! \mid (M+1)_{N-1} \text{ then } N! \mid (M+1)_N.$$

Note that if we show this, then we have completed the induction step for the Induction Hypothesis II hence the result would be true for $n = N$ and for all m. This in turn completes the induction step for Induction Hypothesis I and hence would prove the claim.

Note also, so far even though a lot has been written, everything has been completely formal and no real work has been done except for the facts that $1 \mid m$ and $n! \mid n!$. We finally must do the mathematics. This comes from the equation

$$\begin{aligned}
&(M+1)_N - (M)_N \\
&\qquad = (M+1)(M+2)\cdots(M+N) - M(M+1)\cdots(M+N-1) \\
&\qquad = (M+1)\cdots(M+N-1)[(M+N) - M] \qquad \text{(factoring)} \\
&\qquad = N(M+1)\cdots(M+N-1) = N(M+1)_{N-1}.
\end{aligned}$$

By Induction Hypothesis I, we have

$$(N-1)! \mid (M+1)_{N-1} \text{ so } N! \mid N(M+1)_{N-1},$$

and by Induction Hypothesis II, we have $N! \mid (M)_N$. This means that

$$N! \mid N(M+1)_{N-1} + (M)_N, \ \textit{i.e.}, \ N! \mid (M+1)_N$$

by Exercise 1.13(1), completing the induction step for Induction Hypothesis II, which completes the induction step for Induction Hypothesis I as needed. $\square$

Corollary 2.13. *Let $n \in \mathbb{Z}^+$. Then there exist infinitely many n consecutive composite (i.e., non-prime and not 0 or ± 1) positive integers.*

Proof. Let $m \in \mathbb{Z}^+$ and $N = (m)_{n+1} = m(m+1)\cdots(m+n)$. Then by the example, we have $(n+1)! \mid N$. If follows that for any integer s satisfying $2 \le s \le n+1$, we have $s \mid N+s$. As $n+1 < N+s$, the positive integers

$$(*) \qquad N+2,\ N+3,\ \ldots,\ N+n+1$$

are all composite. $\square$

Of course, the same proof works with $m = 1$, so we did not need the example to prove this. However, (*) says there are arbitrarily large gaps between primes. On the other hand, primes are not 'sparse', as Chebyshev proved Bertrand's Hypothesis, which says

If $n \in \mathbb{Z}^+$ then there exists a prime p satisfying $n \le p \le 2n$.

Of course, $2n$ in the above can only be prime if $n = 1$. In fact, earlier Euler showed that the infinite sum $\sum_{p \text{ a positive prime}} \frac{1}{p}$ diverges, so there are 'many more' primes than say squares (in some sense).

Corollary 2.14. *Let m and n be positive integers with $n \le m$. Define the binomial coefficients*

$$\binom{m}{n} := \frac{m!}{(m-n)!n!} = \frac{m(m-1)\cdots(m-n+1)}{n!}$$

and

$$\binom{-m}{n} := (-1)^n \frac{m(m+1)\cdots(m+n-1)}{n!}.$$

Then $\binom{m}{n}$ and $\binom{-m}{n}$ are integers.

Proof. We know that $\binom{m}{n} = \frac{(m-n+1)_n}{n!}$ is an integer by the example. The second statement follows from the identity

$$\binom{-m}{n} = (-1)^n \binom{m+n-1}{n}.$$

$\square$

Corollary 2.15. *Let $p > 1$ be a prime. Then*

$$\binom{p}{1}, \binom{p}{2}, \ldots, \binom{p}{p-1}$$

are all divisible by p, i.e., $p \mid \binom{p}{n}$ for $1 \le n \le p-1$.

Proof. If $1 \le n \le p-1$, then the example says that

$$n! \mid p(p-1)\cdots(p-n+1) = (p-n+1)_n$$

We also know that s and p have *no common nontrivial factors* if $1 < s < p$. [Proof?] We say that s and p are *relatively prime*. (Cf. 4.4.) We shall show this below (cf. Theorem 4.10) but assume for now the following:

Theorem 2.16. (General Form of Euclid's Lemma) *Let a, b, c be nonzero integers with a and b relatively prime integers. If $a \mid bc$ then $a \mid c$.*

Applying Euclid's Lemma, we conclude that

$$n! \mid (p-1)\cdots(p-n+1) \text{ so } pn! \mid p(p-1)\cdots(p-n+1)$$

for all n satisfying $1 \le n \le p-1$. (Why?) □

Exercises 2.17.

1. Prove that the number of subsets of a set with n elements is 2^n.
2. When Gauss was ten years old he almost instantly recognized that $1+2+\cdots+n = \frac{n(n+1)}{2}$. [Actually, what he did was a bit harder.] What is a formula for the sum of the first n cubes? Prove your result.
3. The first nine Fibonacci numbers are $1, 1, 2, 3, 5, 8, 13, 21, 34$. What is the nth Fibonacci number F_n? Show that $F_n < 2^n$.
4. Note that Euclid's proof of the infinitude of primes clearly shows that if p_n is the nth prime, then $p_{n+1} \le p_n^n + 1$. Be more careful and show that $p_{n+1} \le 2^{2^{n+1}}$? Using this, show that $\pi(x) \ge \log\log(x)$, where $\pi(x)$ is the number of primes less than x if $x \ge 2$. [This is a bad estimate.]
5. Prove that the Well-Ordering Principle, the First Principle of Finite Induction, and the Second Principle of Finite Induction are all equivalent.
6. State and prove the binomial theorem. What algebraic properties do you need for your proof to work?
7. Fill in the missing details in the proof of Corollary 2.15.

3. Addendum: The Greatest Integer Function

Our (double) induction showing that binomial coefficients are integers was complicated to write, although the mathematics was simple. In this addendum, we shall give another proof, mathematically more complicated, but of greater interest. We shall use some results that we shall prove in the next section, in particular, the Division Algorithm 4.2 and the Fundamental Theorem of Arithmetic 4.16. This alternate proof uses the following function:

Definition 3.1. The *greatest integer function*

$$[\ ] : \mathbb{R} \to \mathbb{Z} \text{ is given by } [x] := \text{ the greatest integer } n \text{ satisfying } n \le x,$$

i.e., if $x \in \mathbb{R}$, using Application 2.4, we can write

$$x = n + x_0 \text{ with } 0 \le x_0 < 1,\ n \in \mathbb{Z} \text{ then } [x] = n.$$

This function satisfies:

Properties 3.2. Let m be a positive integer. Then the following are true:

(1) $[x] \le x < [x] + 1$.
(2) $[x + m] = [x] + m$.
(3) $[\frac{x}{m}] = [\frac{[x]}{m}]$.

(4) $[x] + [y] \le [x + y] \le [x] + [y] + 1$.

(5) If $n, a \in \mathbb{Z}^+$, then $[\frac{n}{a}]$ is the number of integers among $1, \ldots, n$ that are divisible by a.

PROOF. (1) and (2) are easy to see.

(3): Write $x = n + x_0$ with $0 \le x_0 < 1$ and $n \in \mathbb{Z}$. Using the Division Algorithm 4.2, write $n = qm + r$ with $q, r \in \mathbb{Z}$ and $0 \le r < m$. Then we have

$$[\frac{x}{m}]=[\frac{n + x_0}{m}]=[\frac{qm + r + x_0}{m}]=[q + \frac{r + x_0}{m}]= q+[\frac{r + x_0}{m}].$$

As $r \in \mathbb{Z}$ and $0 \le r < m$, we have $0 \le r \le m-1$ (cf. Application 2.4); and as $0 \le x_0 < 1$, we have $r + x_0 < m$, so

$$[\frac{x}{m}] = q = [q + \frac{r}{m}] = [\frac{n}{m}] = [\frac{[x]}{m}].$$

(4): Write

$$\begin{array}{llllll} x & = n + x_0 & \text{with} & 0 \le x_0 < 1 & \text{and} & n \in \mathbb{Z} \\ y & = q + y_0 & \text{with} & 0 \le y_0 < 1 & \text{and} & q \in \mathbb{Z}. \end{array}$$

Then $0 \le x_0 + y_0$, so

$$[x + y] = [n + q + x_0 + y_0] = n + q + [x_0 + y_0] = [x] + [y] + [x_0 + y_0] < [x] + [y] + 2$$

and the result follows easily.

(5): Let $a, 2a, \ldots, ja$ denote all the positive integers $\le n$ and divisible by a. We must show $[\frac{n}{a}] = j$. Clearly, $ja \le n < (j + 1)a$. Therefore, $j \le \frac{n}{a} < j + 1$. The result now follows. $\square$

Let $n \in \mathbb{Z}^+$ and $p > 1$ be a prime. By the Fundamental Theorem of Arithmetic 4.16, we know that there exists a unique integer $e \ge 0$ such that $p^e \mid n$ but $p^{e+1} \nmid n$. We shall write this as $p^e \mid\mid n$.

Theorem 3.3. *Let n be a positive integer and p a positive prime. Suppose that $p^e \mid\mid n!$. Then*

$$e = \sum_{i=1}^{\infty} [\frac{n}{p^i}].$$

PROOF. If $p^i > n$, then $[\frac{n}{p^i}] = 0$. So the sum is really a finite sum. We prove the result by induction on n.

$n = 1$. There is nothing to prove.

We can, therefore, make the following:

Induction Hypothesis. Let $e' = \sum_{i=1}^{\infty} \left[\frac{n-1}{p^i}\right]$, then $p^{e'} \mid\mid (n-1)!$.

Let $p^f \mid\mid n$. As $n! = n(n-1)!$, by the induction hypothesis, we have $p^e = p^{e'} p^f = p^{e'+f} \mid\mid n!$, hence $e = f + e'$ or $f = e - e'$. So it suffices to prove that

$$f = \sum_{i=1}^{\infty} [\frac{n}{p^i}] - \sum_{i=1}^{\infty} [\frac{n-1}{p^i}].$$

Claim 3.4. $[\frac{n}{p^i}] - [\frac{n-1}{p^i}] = \begin{cases} 1 & \text{if } p^i \mid n \\ 0 & \text{if } p^i \not| \, n. \end{cases}$

Suppose that $p^i \mid n$, then $n = p^i u$ with $u \in \mathbb{Z}$. As $0 < \frac{1}{p^i} < 1$, we have

$$[\frac{n}{p^i}] = u \text{ and } [\frac{n-1}{p^i}] = [u - \frac{1}{p^i}] = u - 1$$

as needed.

Suppose that $p^i \not| \, n$. The result is immediate if $p^i > n$, so suppose not. Then we can write $n = p^i u + r$ with u and r integers with r satisfying $1 \leq r < p^i$ using the Division Algorithm 4.2. Hence

$$\frac{n}{p^i} = u + \frac{r}{p^i}, \quad \text{so} \quad [\frac{n}{p^i}] = u$$

and

$$\frac{n-1}{p^i} = \frac{p^i u + r - 1}{p^i} = u + \frac{r-1}{p^i} \quad \text{for} \quad 0 \leq r - 1 < p^i,$$

so we also have $[\frac{n-1}{p^i}] = u$ as needed. This proves the claim. But the claim means that

$$\sum_{i=1}^{\infty} [\frac{n}{p^i}] - \sum_{i=1}^{\infty} [\frac{n-1}{p^i}] = \sum_{p^i \mid n} 1 = f. \qquad \square$$

Example 3.5. Using the properties of [ ], we compute e such that $7^e \mid\mid 1000!$.

$$[\frac{1000}{7}] = 142$$
$$[\frac{1000}{7^2}] = [\frac{[\frac{1000}{7}]}{7}] = [\frac{142}{7}] = 20$$
$$[\frac{1000}{7^3}] = [\frac{[\frac{1000}{7^2}]}{7}] = [\frac{20}{7}] = 2$$
$$[\frac{1000}{7^4}] = [\frac{[\frac{1000}{7^3}]}{7}] = [\frac{2}{7}] = 0,$$

so $e = 142 + 20 + 2 = 164$, i.e., $7^{164} \mid\mid 1000!$.

We can generalize our result about binomial coefficients to show that *multinomial coefficients* are integers.

Corollary 3.6. *Suppose that $a_1, \ldots, a_r$ are non-negative integers satisfying $a_1 + \cdots + a_r = n$. Then the multinomial coefficient $\dfrac{n!}{a_1! \cdots a_r!}$ is an integer.*

PROOF. By the Fundamental Theorem of Arithmetic 4.16 and the theorem, it suffices to prove that for each prime p, we have

$$\sum_{i=1}^{\infty} [\frac{n}{p^i}] \geq \sum_{i=1}^{\infty} [\frac{a_1}{p^i}] + \cdots + \sum_{i=1}^{\infty} [\frac{a_r}{p^i}].$$

Applying Property 3.2(4) shows that

$$[\frac{a_1}{p^i}] + \cdots + [\frac{a_r}{p^i}] \leq [\frac{a_1 + \cdots + a_r}{p^i}] = [\frac{n}{p^i}]$$

for all i. Summing over i yields the result. $\square$

That the product of n consecutive positive integers is divisible by $n!$ and that binomial coefficients are integers is now easy to see as $\dfrac{(m)_n}{n!} = \dbinom{m+n-1}{n}$ for all $m \in \mathbb{Z}^+$ and $\dbinom{a}{b} = \dfrac{a!}{b!(a-b)!}$ for all $a, b \in \mathbb{Z}^+$ with $b \leq a$.

Chebyshev cleverly used the binomial coefficient $\dbinom{2n}{n} = \dfrac{(2n)!}{n!n!}$. For each positive prime $p \leq 2n$, let $r_p \in \mathbb{Z}$ satisfy

$$p^{r_p} \leq 2n < p^{r_p+1}$$

and $p^e \mid\mid \binom{2n}{n}$. Then

$$e = \sum_{i=1}^{\infty} [\frac{2n}{p^i}] - 2\sum_{i=1}^{\infty} [\frac{n}{p^i}].$$

We have

$$[\frac{2n}{p^i}] = [\frac{n}{p^i} + \frac{n}{p^i}] \leq [\frac{n}{p^i}] + [\frac{n}{p^i}] + 1$$

by Property 3.2(4), so $e \leq \sum_{i=1}^{r_p} 1 = r_p$. This is used to prove

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} \text{ divides } \prod_{\substack{p \leq 2n \\ p \text{ prime}}} p^{r_p} \tag{3.7}$$

and if further $n < p \leq 2n$, then $p \mid (2n)!$ but $p \nmid (n!)^2$. Therefore, we see

$$n^{\pi(2n)-\pi(n)} \leq \prod_{\substack{n<p\leq 2n \\ p \text{ prime}}} p \leq \binom{2n}{n} \leq \prod_{\substack{p\leq 2n \\ p \text{ prime}}} p^{r_p} \leq (2n)^{\pi(2n)}. \tag{3.8}$$

Using this Chebyshev showed that there were positive real numbers c_1 and c_2 satisfying

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}$$

(cf. The Prime Number Theorem). He also showed if $n \geq 3$ and $2n/3 < p < n$, with p a prime, then $p \nmid \binom{2n}{n}$ and used this to prove Bertrand's Hypothesis that there always exists a prime p such that $n \leq p \leq 2n$. For a full proof see Appendix B. Besides Theorem 3.3, one only needs the main results in the next section §4.

Exercises 3.9.

1. Let a and b be positive integers. Show that $\dfrac{(ab)!}{a!(b!)^a}$ is an integer.
2. Verify equations (3.7) and (3.8).

3. Let $0 < x < 1$. Let $a_1 > 0$ be the smallest positive integer satisfying $x_1 = x - a_1^{-1} \geq 0$, $a_2 > 0$ be the smallest positive integer satisfying $x_2 = x_1 - a_2^{-1} \geq 0$, etc. Show that $x = \sum_{i=1}^{\infty} \frac{1}{a_i} = \sum_{i=1}^{n} \frac{1}{a_i}$ some n if and only if x is a rational number. [This says that every rational between 0 and 1 is an sum of Egyptian numbers.] (You can use the Fundamental Theorem of Arithmetic 4.16.)

4. Division and the Greatest Common Divisor

We turn to further applications of well-ordering. We shall need the following:

Properties 4.1. For integers r, n, and m the following division properties hold:

(1) If $r \mid m$ and $r \mid n$, then $r \mid am + bn$ for all integers a and b.
(2) If $r \mid n$, then $r \mid mn$.
(3) If If $r \mid n$ and $n \neq 0$, then the absolute value $|n| \geq |r| \geq r$.
(4) If $m \mid n$ and $n \mid m$, then $n = \pm m$.
(5) If $mn = 0$, then $m = 0$ or $n = 0$.
(6) If $mr = nr$, then either $m = n$ or $r = 0$.

Proof. Exercise, but do not use the Fundamental Theorem of Arithmetic. $\square$

The most important elementary property about the integers is the next result, which intertwines multiplication and addition of integers. In the sequel, we shall investigate when it holds in more general situations. As it is one of our first proofs about integers, we shall put in a lot of detail. Notice that one can often assume properties about integers that we have not shown and perhaps are not clear. For example, in the proof below, we assume the transitivity of $>$. Can this be proved or must it be axiomized?

Theorem 4.2. (Division Algorithm) *Let m and n be integers with m positive. Then there exist unique integers q and r satisfying:*

(i) $n = qm + r$.
(ii) $0 \leq r < m$

[Of course, (ii) is the crux.]

Proof. We have two things to show: existence and uniqueness. We first show

Uniqueness: Let (q, r) and (q', r') be two pairs of integers satisfying the conclusion. We must show $q = q'$ and $r = r'$. We have

$$(*) \qquad \begin{gathered} qm + r \;=\; n \;=\; q'm + r' \\ 0 \leq r,\, r' < m \end{gathered}$$

Without loss of generality, we may assume that $r \leq r'$. By (*), we have

$$0 \leq r' - r = (q - q')m.$$

If $q - q' = 0$, i.e., $q = q'$, then $r' - r = 0$ and $r' = r$ and we are done. So we may assume that $q - q' \neq 0$, i.e., $q \neq q'$. Then $r' - r = (q - q')m \neq 0$ by Property 4.1(5) as $m \neq 0$. Therefore, we have

$$r' - r > 0 \text{ and } m \mid r' - r.$$

Thus by Property 4.1(3), we have

$$m \leq r' - r < r' < m,$$

a contradiction. So $q - q' = 0$ as needed.
[Note that we have used the important Property 4.1(5):

$$ab = 0 \quad \Longrightarrow \quad a = 0 \text{ or } b = 0,$$

which is equivalent to

$$ab = ac \quad \Longrightarrow \quad a = 0 \text{ or } b = c.]$$

Existence:
Case 1. $n > 0$:
Intuitively we know that the number n must lie in some half open interval $[qm, (q+1)m)$, i.e., there exists an integer q such that $qm \leq n < (q+1)m$. Let us show this rigorously. Let

$$S = \{s \in \mathbb{Z}^+ \mid sm > n\} \subset \mathbb{Z}^+.$$

As $m > 0$, we have $m \geq 1$. (Recall we have shown there is no integer properly between 0 and 1.) So $(n+1)m = mn + m \geq n + m > n$. [Can you show that $mn > n$?] Thus $n+1 \in S$ so $S \neq \emptyset$. By the Well-Ordering Principle, there exists a least integer $q+1 \geq 1$ which means $qm \leq n < (q+1)m$. Let $r = n - qm \geq 0$. We then have

$$0 \leq r = n - qm < (q+1)m - qm = m,$$

so these q, r work.
Case 2. $n < 0$:
By Case 1, there exist integers q_1 and r_1 satisfying

$$-n = |n| = q_1 m + r_1 \quad \text{and} \quad 0 \leq r_1 < m.$$

If $r_1 = 0$, then $q = -q_1$ and $r = 0$ work. If $r_1 > 0$, then

$$\begin{aligned} n &= -q_1 m - r_1 = -q_1 m - m + m - r_1 \\ &= (-q_1 - 1)m + (m - r_1). \end{aligned}$$

Since $0 \leq m - r_1 < m$, we have $q = -q_1 - 1$ and $r = m - r_1$ work. □

Question 4.3. What if $m < 0$ in the above?

As mentioned above, later we shall be interested in generalizing the Division Algorithm for integers. You already have used an analogue of it in the case of the division of polynomials with real coefficients in the real numbers. Does this analogue hold if we only allow integer coefficients?

We turn to another important property about the integers.

Definition 4.4. Let n and m be integers with at least one nonzero. An integer d is called a *greatest common divisor* or *gcd* of m and n if d satisfies the following:

(i) $d > 0$.
(ii) $d \mid m$ and $d \mid n$.
(iii) If e is an integer satisfying $e \mid m$ and $e \mid n$ then $e \mid d$.

If $d = 1$ is a gcd of n and m, we say that n and m are *relatively prime*.
[Note if a is nonzero, then that a is a gcd of a and 0 and 1 is a gcd of 1 and a.]

Theorem 4.5. *Let $n \neq 0$ and m be integers. Then a gcd of m and n exists and is unique.*

PROOF. **Uniqueness**: If both d and d' satisfy (i), (ii), and (iii), then $d \mid d'$ and $d' \mid d$. Therefore, $d = \pm d'$ by Property 4.1(4), hence $d = |d| = |d'| = d'$ by (i).

Existence: Let

$$S = \{am + bn \mid am + bn > 0 \text{ with } a, b \in \mathbb{Z}\}.$$

[This is the tricky part of this proof. Where did this come from?]

As $n \neq 0$, we have $|n| \in S$, so $S \neq \emptyset$. By the Well-Ordering Principle, there exists a least element $d = am + bn$ in S for some integers a, b.

Claim. This d works, i.e., d satisfies (i), (ii), and (iii).

By choice, d satisfies (i). To show (ii), we use the Division Algorithm to produce integers q and r satisfying $n = qd + r$ with $0 \leq r < d$. We show $r = 0$, i.e., $d \mid n$. As

$$n = dq + r = (am + bn)q + r, \quad \text{we have} \quad 0 \leq r = (1 - bq)n + (-aq)m.$$

This means that either $r \in S$ or $r = 0$. Since $r < d$, the minimality of d implies that $r \notin S$, so $r = 0$ and $d \mid n$. Similarly, $d \mid m$.

As for (iii), suppose that the integer e satisfies $e \mid m$ and $e \mid n$. Then by Property 4.1 (1), we have $e \mid am + bn = d$. □

Notation 4.6. If m and n are integers with at least one nonzero, we shall denote their gcd by (m, n).

Note that the proof even gives the following

Bonus. Let m and n be integers, at least one nonzero, and $d = (m, n)$. Then there exist integers x and y satisfying

$$d = (m, n) = mx + ny.$$

In particular, the *Diophantine equation*

$$(m, n) = mX + nY$$

(with X and Y variables) has a solution, where the word Diophantine means that we only want integer solutions, e.g., the Diophantine equation $X^2 = 2$ has no solution.

Warning 4.7. In general, the x and y in the Bonus are not unique, e.g.,

$$\begin{aligned}(2,4) &= 2 = 1 \cdot 4 - 1 \cdot 2 = 0 \cdot 4 + 1 \cdot 2.\\ (2,3) &= 1 = 1 \cdot 3 - 1 \cdot 2 = -1 \cdot 3 + 2 \cdot 2.\end{aligned}$$

See Exercise 4.24(7) for the general solution to the Diophantine equation $d = mX + nY$.

Unfortunately, our proof does not show how to find any solution to the Diophantine equation $(m, n) = mX + nY$. Euclid knew how to do so, and his proof constructs such a solution.

Theorem 4.8. (Euclidean Algorithm) *Let a and b be positive integers with $b \nmid a$. [If $b \mid a$, we have $(a,b) = b$.] Then there exists an integer $k > 0$ and equations*

$$\begin{aligned}
a &= bq_1 + r_1 && \text{with} \quad 0 < r_1 < b \\
b &= r_1 q_2 + r_2 && \text{with} \quad 0 < r_2 < r_1 \\
r_1 &= r_2 q_3 + r_3 && \text{with} \quad 0 < r_3 < r_2 \\
&\ \vdots \\
r_{k-2} &= r_{k-1} q_k + r_k && \text{with} \quad 0 < r_k < r_{k-1} \\
r_{k-1} &= r_k q_{k+1}
\end{aligned}$$

for some integers $q_1, \dots, q_{k+1}$ and $r_1, \dots, r_k$.

Proof. Exercise 4.24(4). □

Moreover, $r_k = (a,b)$ and the Diophantine equation $(a,b) = aX + bY$ has a solution.

Finding the gcd of two integers is much easier and faster than factoring numbers. Indeed, using the internet to make purchases or send confidential information is (relatively) safe because of the difficulty in factoring large numbers.

Properties 4.9. Let a, b, and c be integers with $a \neq 0$ and $d = (a,b)$.

(1) We have $d = 1$ if and only if $1 = ax + by$ for some integers x, y.
(2) $(\frac{a}{d}, \frac{b}{d}) = 1$. [Note that $\frac{a}{d}$ and $\frac{b}{d}$ are integers.]
(3) If $d = 1$ and $a \mid bc$ (in $\mathbb{Z}$), then $a \mid c$.
(4) If $a \mid bc$ (in $\mathbb{Z}$), then $\frac{a}{d} \mid c$.
(5) If $m > 0$, then $(ma, mb) = md$.

We have seen Property 4.1(3) above. As it is so important, we write it one more time.

Theorem 4.10. (General Form of Euclid's Lemma) *Let a, b be relatively prime integers with a nonzero. If $a \mid bc$ with c an integer, then $a \mid c$.*

A special case of this theorem is:

Theorem 4.11. (Euclid's Lemma) *Let a, b be integers and p be a prime satisfying $p \mid ab$. Then $p \mid a$ or $p \mid b$.*

Of course this means

Consequence 4.12. If a prime p satisfies $p \mid a_1 \cdots a_r$, with a_i integers for $i = 1, \dots, r$, then there exists an i such that $p \mid a_i$.

Proof. (of the properties and Euclid's Lemma.)

(1): We already know if $d = 1$ then $1 = ax + by$ for some $x, y \in \mathbb{Z}$. Conversely, suppose that $1 = ax + by$ for some $x, y \in \mathbb{Z}$. As $d \mid a$ and $d \mid b$, we have $d \mid 1$ by Property 4.1(1). Since $d > 0$, we have $d = 1$.

(2): By the Bonus, there exist integers x and y satisfying $d = ax + by$ hence $1 = \frac{a}{d}x + \frac{b}{d}y$ so (2) follows by (1).

(3): Whenever you can write 1 or 0 in a nontrivial way do so! It is a very useful technique. So by (1), we have

$$1 = ax + by \text{ for some } x, y \in \mathbb{Z} \qquad \text{(Key Observation)} \tag{4.13}$$

and given such an equation, you can always multiply it, say by the integer c, to get

$$c = cax + cby. \qquad \text{(Key Trick)} \tag{4.14}$$

As $a \mid a$ and $a \mid cb$, we conclude that $a \mid c$.

(4): The hypothesis means that $bc = an$ for some integer n, so $\frac{a}{d}n = \frac{b}{d}c$. By (2), we know that $(\frac{a}{d}, \frac{b}{d}) = 1$, hence $\frac{a}{d} \mid c$ by (3).

(5): We leave this is an exercise.

(proof of) Euclid's Lemma. If p is a prime and $p \mid ab$ but $p \nmid a$, then $(p, a) = 1$ as only $\pm 1, \pm p$ divide p. Since $p \mid ab$, we conclude that $p \mid b$ by (3). □

The converse of Euclid's Lemma is also true, viz.,

Proposition 4.15. *Let p be an integer with $|p| > 1$. Then p is a prime if and only if whenever $p \mid ab$, with a and b integers, then $p \mid a$ or $p \mid b$.*

This proposition, which we leave as an exercise (cf. Exercise 4.24(8)) is a key to much of (commutative) ring theory. It says that we have two ways to define the notion of a prime element in the integers. For more general structures, called rings, we can look at both of these conditions. Unfortunately, they need not be equivalent and it turns out that the more useful condition is the condition $p \mid ab$ then $p \mid a$ or $p \mid b$. In fact, its generalization has major repercussions not only in algebra but also in geometry. We shall investigate some of these later on.

We had a further loose end to establish, viz., a proof of the Fundamental Theorem of Arithmetic, which we turn to next. We first state it again.

Theorem 4.16. (The Fundamental Theorem of Arithmetic) *Every integer $n > 1$ is a product of positive primes unique up to order, i.e., there exist unique primes $1 < p_1 < \cdots < p_r$ and integers $e_1, \dots, e_r > 0$ such that*

$$n = p_1^{e_1} \cdots p_r^{e_r}. \tag{*}$$

[We call (*) the *standard representation* or *standard factorization* of n, and this representation is unique.]

Proof. **Existence**: Let

$$S = \{n > 1 \text{ in } \mathbb{Z} \mid n \text{ is not a product of primes}\}.$$

We must show $S = \emptyset$. Suppose this is false. By the Well-Ordering Principle, there exists a minimal element $n \in S$. Clearly, no prime lies in S, so n is not a prime. Hence there exist integers n_1, n_2 satisfying $n = n_1 n_2$ and $1 < n_i < n$, $i = 1, 2$ (as there exists an integer $1 < n_1 < n$ dividing n). By minimality, $n_1, n_2 \notin S$, so each is a product of primes. Hence so is $n = n_1 n_2$, a contradiction.

[This is a wonderful argument, and we shall see it quite often.]

Uniqueness: Suppose that

$$p_1^{e_1} \cdots p_r^{e_r} = n = q_1^{f_1} \cdots q_s^{f_s}$$

with $1 < p_1 < \cdots < p_r$ and $1 < q_1 < \cdots < q_s$ primes and all the e_i, f_j positive integers. We may assume that $p_1 \leq q_1$. As $p_1 \mid n = q_1^{f_1} \cdots q_s^{f_s}$, we must have $p_1 \mid q_i$ for some i by Euclid's Lemma. But $p_1 \leq q_i$ are primes, so we must have $i = 1$ and $p_1 = q_1$. Thus, by cancellation,

$$p_1^{e_1-1} \cdots p_r^{e_r} = q_1^{f_1-1} \cdots q_s^{f_s}$$

and we are done by induction. What is the induction hypothesis and why are we done? $\square$

The proof of the existence statement in the Fundamental Theorem of Arithmetic is used repeatedly in mathematics and its consequence occurs much more frequently than the uniqueness statement. If we are studying a class of objects in which there is a set of basic objects (atoms), the basic question that arises is whether we can break up an element in the class into a finite number of these atoms (in some way). In the Fundamental Theorem of Arithmetic the atoms are the primes.

We next show how Euler reformulated the Fundamental Theorem of Arithmetic. This led to the subject called Analytic Number Theory by showing how the zeta function defined previously is related to primes.

Definition 4.17. An *arithmetic function* is a function $f : \mathbb{Z}^+ \to \mathbb{C}$. If f is a nonzero arithmetic function, it is called *multiplicative* if $f(mn) = f(m)f(n)$ if $(m,n) = 1$ and *completely multiplicative* if, in addition, $f(mn) = f(m)f(n)$ for all $m, n \in \mathbb{Z}^+$.

Note. If f is a multiplicative function, then $f(1) = 1$, since $f(1)f(n) = f(n)$ is true for some n with $f(n) \neq 0$.

As examples, we give a few of the useful arithmetic functions used in number theory.

Examples 4.18. Let p be an arbitrary positive prime and $n > 1$ an arbitrary integer with standard factorization $n = p_1^{e_1} \cdots p_r^{e_r}$. The following are arithmetic functions $f : \mathbb{Z}^+ \to \mathbb{C}$ defined as follows:

1. The *von Mangoldt function*: $\lambda(1) = 0$ and $\Lambda(n) = \log p$ if $n = p^e$ some $e \geq 1$. It is not multiplicative.
2. The *Möbius μ-function*: $\mu(1) = 1$ and $\mu(n) = (-1)^r$. It is multiplicative, but not completely multiplicative.
3. The *Euler phi-function*: $\varphi(1) = 1$ and $\varphi(n) = |\{d \mid 1 \leq d < n \text{ with } (d,n) = 1\}|$. It is multiplicative but not completely multiplicative (as we shall see later).
4. The *division functions*: $\sigma_i(1) = 1$ and $\sigma_i(n) = \sum_{d \mid n} d^i$. (Here and below the sum means d is a positive integer dividing the positive integer n.) It is completely multiplicative.
5. The *identity arithmetic function*: $I(1) = 1$ and $I(n) = [\frac{1}{n}]$. It is completely multiplicative.
6. The *Liouville function*: $\lambda(1) = 1$ and $\lambda(n) = (-1)^{e_1 + \cdots + e_r}$. It is completely multiplicative.

We set up some notation (used before). If p is a positive prime, let $\sum\limits_p$, $\prod\limits_p$ mean the sum (or product) over all positive primes p. We can also condition the p. In the theorem, below we will have absolutely convergent infinite sums and products. This will allow us to reorder such sums and products. Therefore, it will be useful to have the following notation with $1 < x$ in $\mathbb{R}$:

$$\begin{aligned} A_x &:= \{n \mid \text{ all prime factors of } n \text{ are less than } x\} \text{ and} \\ B_x &:= \{n \mid \text{ there exists a } p > x \text{ such that } p \not| \, n\}. \end{aligned}$$

By the Fundamental Theorem of Arithmetic 4.16, each element in A_x and B_x is uniquely represented and $\mathbb{Z}^+ = A_x \cup B_x$ for all $x > 0$. Although the union is not a disjoint union, $\mathbb{Z}^+ = \lim_{x\to\infty} A_x$ with every element uniquely a product of its prime factors. Using theorems from calculus, we prove:

Theorem 4.19. (Euler) *Let $f : \mathbb{Z}^+ \to \mathbb{C}$ be a multiplicative function. Suppose one of the following two conditions hold:*

(i) *The infinite sum* $\sum\limits_{n=1}^{\infty} |f(n)|$ *converges, i.e.,* $\sum\limits_{n=1}^{\infty} f(n)$ *converges absolutely.*

(ii) *The infinite product* $\prod\limits_p (1 + |f(p)| + |f(p^2)| + \cdots)$ *converges, i.e.,*

$$\prod_p (1 + f(p) + f(p^2) + f(p^3) + \cdots) \text{ converges absolutely.}$$

Then

$$\sum_{n=1}^{\infty} f(n) = \prod_p (1 + f(p) + f(p^2) + f(p^3) + \cdots).$$

If, in addition, f is completely multiplicative, then

$$\sum_{n=1}^{\infty} f(n) = \prod_p \frac{1}{1 - f(p)}.$$

The formulas in the conclusion of the theorem are called *Euler Formulas*.

Proof. Suppose that Condition (i) holds. Say $\sum_{n=1}^{\infty} |f(n)|$ converges to $\overline{S}$ and $\sum_{n=1}^{\infty} f(n)$ converges to S. Let $P(x) = \prod_{p\le x}(1 + f(p) + f(p^2) + f(p^3) + \cdots)$. As $P(x)$ is a finite product of absolutely convergent series, we can arrange terms and arrive at the same value. In particular, in the notation set up as above, we have

$$|S - P(x)| \le |\sum_{B_x} f(n)| \le \sum_{B_x} |f(n)| \le \sum_{n>x} |f(n)|.$$

As $x \to \infty$, we have $\sum_{n\ge x} |f(n)| \to 0$. Consequently $P(x) \to S$ as $x \to \infty$. Therefore, $\prod_p (1 + |f(p)||f(p^2)| + +|f(p^3)| + \cdots)$ converges to $\overline{S}$.

Suppose Condition (2) holds: Let $P(x) = \prod_{p\le x}(1+|f(p)|+|f(p^2)|+|f(p^3)|+\cdots)$. Then

$$P(x) = \sum_{A_x} |f(n)| \geq \sum_{n\le x} |f(n)|.$$

Therefore, $\sum_{n=1}^{\infty} |f(n)|$ converges. So Condition (i) holds. This proves the first statement. If f is completely multiplicative, then $1+f(p)+f(p^2)+f(p^3)+\cdots = 1+f(p)+f(p)^2)+f(p^3)+\cdots$ is a geometric series. The second statement follows. □

Corollary 4.20. *Let f be a multiplicative function satisfying the conditions in Theorem 4.19. Then*

$$\sum_{n=1}^{\infty} f(n) = \prod_p \frac{1}{1-f(p)}.$$

Corollary 4.21. *The sum $\displaystyle\sum_p \frac{1}{p}$ diverges. In particular there exists infinitely many primes.*

Proof. Let $f : \mathbb{Z}^+ \to \mathbb{C}$ be the function $f(n) = \dfrac{1}{n}$. If $\displaystyle\sum_p \frac{1}{p}$ converges, then both $\displaystyle\prod_p(1-\frac{1}{p})$ and $\displaystyle\prod_p(1-\frac{1}{p})^{-1}$ converge. Therefore, $\displaystyle\sum_n \frac{1}{n}$ converges, a contradiction. □

Not only does this corollary say that there are infinitely many primes, but also lots of them relative to the number of positive integers. For example, there exist infinitely many square integers yet $\sum_{n=1}^{\infty} \frac{1}{n^r}$ converges for any $e > 1$ in $\mathbb{R}$. However, the harmonic series, i.e., with $e = 1$ of such a sum, leads by Theorem 4.19 to the relationship between positive primes and integers that we want.

Corollary 4.22. *The zeta function $\displaystyle\zeta(s) = \sum_{i=1} \frac{1}{n^s}$ with s a complex variable converges absolutely for all* $\mathrm{Re}(s) > 1$. *In particular,*

$$\zeta(s) = \prod_p (1-\frac{1}{p^s})^{-1}.$$

To prove Theorem 4.19, we used the Fundamental Theorem of Arithmetic. We now show the converse holds.

Theorem 4.23. *The Fundamental theorem is equivalent to $\displaystyle\zeta(s) = \sum_{i=1} \frac{1}{n^s}$ with s a complex variable converges for all* $\mathrm{Re}(s) > 1$.

Proof. Let N be the arithmetic function defined by $N(1) = 1$ and for $n > 1$, $N(n)$ is the number of all ways of writing n as a product of primes up to order, e.g., $N(p^2) = 2$, since pp, p^2 are all of them and if q is another prime, $N(p^2q) = 2$, since pqp, qp^2 are all of them. Since $\zeta(s)$ converges absolutely for $\mathrm{R}(s) > 1$,

$$\zeta(s) = \prod_p (1-\frac{1}{p})^{-s} = \sum_{n=1}^{\infty} \frac{N(n)}{n^s}.$$

It follows that $N(n) = 1$ if $N(n) > 1$, i.e., every positive integer is a product of primes in precisely one way. □

Exercises 4.24.

1. Prove all the properties in Properties 4.1 without using the Fundamental Theorem of Arithmetic.
2. Let $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{Q}$ [or any *field*, or even any *ring*, cf. Definition 8.3]. Let $F[t]$ be the set of polynomials with coefficients in F with the usual addition and multiplication. [Which are?] State and prove the analogue of the Division Algorithm for integers. [Use your knowledge of such division. Use the degree of a polynomial as a substitute for statement (ii) in the Division Algorithm for integers. (One usually does not define the degree of the zero polynomial, so if r is the remainder, write $r = 0$ or ...).] What can you do if you take polynomials with coefficients in $\mathbb{Z}$?
3. Prove the following modification of the Division Algorithm: If m and n are two integers with m nonzero, then there exist unique integers q and r satisfying
$$n = mq + r \text{ with } -\frac{1}{2}|m| < r \leq \frac{1}{2}|m|.$$
Moreover, if $m > 0$ is odd, then we can find unique integers q and r satisfying
$$n = mq + r \text{ with } 0 \leq |r| < \frac{m}{2}.$$
[Note that in this case, $m/2$ is not an integer.]
4. Prove the Euclidean Algorithm 4.8, i.e., show that such a k exists and that the Diophantine equation $(m, n) = mX + nY$ has a solution using the Euclidean Algorithm.
5. In the Euclidean Algorithm show each nonzero remainder r_i ($i \geq 2$, where we view b as r_0) satisfies $r_i < \frac{1}{2}r_{i-2}$. Deduce that the number of steps in the algorithm is less than
$$\frac{2\log b}{\log 2},$$
where b is the larger of the numbers a, b. Using the modified Division Algorithm, show the number of steps there is at most $\frac{\log b}{\log 2}$.
6. Solve the Diophantine equation
$$(39493, 19853) = 39493\,X + 19853\,Y.$$
7. Let m, n, and d be nonzero integers. Show that the Diophantine equation $d = mX+nY$ has a solution if and only if $(m, n) \mid d$. If this is the case and (x, y) is a solution in integers, then the general solution is $(x + \frac{n}{(m,n)}k,\ y - \frac{m}{(m,n)}k)$ with $k \in \mathbb{Z}$.
8. Let p be an integer. Then p is a prime if and only if whenever $p \mid ab$, with a and b integers, then $p \mid a$ or $p \mid b$.
9. Carefully write up the induction step and its proof for the uniqueness part of the Fundamental Theorem of Arithmetic in two different ways, i.e., by inducting on two different things.

10. Let n and m be integers with at least one nonzero. An integer l is called a *least common multiple* or *lcm* of m and n if l satisfies the following:
    (i) $l \geq 0$.
    (ii) $m \mid l$ and $n \mid l$.
    (iii) If k is an integer satisfying $m \mid k$ and $n \mid k$, then $l \mid k$.
    Show that an lcm of m and n exists and is unique, denote it by $[m, n]$. In addition, show that $mn = [m, n](m, n)$.
11. Define $\sigma : \mathbb{Z}^+ \to \mathbb{Z}^+$ by $\sigma(n) = \sum_{d|n} d$, the sum of the (positive) divisors of n. Show
    (i) If m and n are relatively prime (positive) integers, then $\sigma(mn) = \sigma(m)\sigma(n)$.
    (ii) If p is a (positive) prime integer and n an integer, then $\sigma(p^n) = (p^{n+1}-1)/(p-1)$.
12. In the notation of the previous result, a positive integer n is called a *perfect number* if $\sigma(n) = 2n$. Prove Theorem 1.2.
13. Show for $n \geq 1$ that $\log n = \sum_{d|n} \Lambda(d)$.
14. Show for $n \geq 1$ that $I(n) = [\frac{1}{n}] = \sum_{d|n} \mu(d)$.
15. Show for $n \geq 1$ that
$$\sum_{d|n} \lambda(d) = \begin{cases} 1 & \text{if } n \text{ is a square} \\ 0 & \text{otherwise} \end{cases}$$
    and also show that $\lambda^{-1}(n) = |\mu(n)|$ for all $n \leq 1$.
16. Show by the following steps that we get an alternate proof that the infinite series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{p_n}$ diverges where p_r is the rth positive prime.
    (i) We may assume that the result is false and $\sum_{n=k+1}^{\infty} \frac{1}{p_n} < \frac{1}{2}$.
    (ii) Let $N = p_1 \cdots p_k$. Then $p_i \not| \, 1 + nN$ for any prime $i = 1, \dots, k$ and
$$\sum_{n=1}^{m} \frac{1}{1+nN} \leq \sum_{i=1}^{\infty} \Big(\sum_{r=k+1}^{\infty} \frac{1}{p_r} \Big)^i \leq \sum_{i=1}^{\infty} (\frac{1}{2})^i.$$
    (iii) This leads to a contradiction.
17. Show that $\displaystyle\prod_{p \leq x} (1 - \frac{1}{p}) < \frac{1}{\log x}$ for all $x \geq 2$.
18. Let $A(x, r)$ denote the number of positive primes not exceeding $x \in \mathbb{R}$ and not divisible by the first r primes $2, 3, 5, \dots, p_r$. Show all of the following:
    (i) $\displaystyle A(x, r) = [x] - \sum_{1 \leq i \leq r} [\frac{x}{p_i}] + \sum_{1 \leq i < j \leq r} [\frac{x}{p_i p_j}] - \cdots$.
    (ii) $\pi(x) \leq A(x, r) + r$.
    (iii) $\displaystyle\lim_{x \to \infty} \frac{\pi(x)}{x} = 0$.
19. Prove that there exist arbitrarily large gaps in the sequence of positive primes.

CHAPTER II

Equivalence Relations

This chapter is the most important foundational chapter in Part One. It introduces the notion of an equivalence relation and determines equivalent formulations of it. In particular, an equivalence relation is shown to be the same as a surjective map. The importance of equivalence relations is that they lead to 'quotients'. This is probably the hardest concept to understand in Part One and Part Two. It is used to coarsen problems that may be more easily be solved, and hopefully leads to the solution of specific problems in which we are interested. The basic difficulty is that it may easy to define an equivalence relation on collection of sets, but it probably is not useful if it must reflect additional properties (e.g., algebraic properties) of those maps, i.e., if the surjective map arising from the equivalence relation (or defining it) does not reflect these additional properties. For example, what information does a surjective linear transformation of vector spaces give us?

5. Equivalence Relations

In this section, we study one of the basic concepts in mathematics, equivalence relations. You should have seen this concept. Recall the following:

Definition 5.1. Let A and B be two sets. A *relation* on A and B is a subset $R \subset A \times B$. It is customary to write aRb if $(a, b) \in R$, and we shall always do so. If $A = B$ then we call such a relation a *relation on* A.

Question 5.2. What kind of relation is a function $f : A \to B$?

Besides functions, we are mostly interested in equivalence relations, which we also recall.

Definition 5.3. A relation R on A is called an *equivalence relation* on A if for all a, b, c in A,

(1)	aRa.	(*Reflexivity*)
(2)	If aRb then bRa.	(*Symmetry*)
(3)	If aRb and bRc then aRc.	(*Transitivity*)

Remark 5.4. We often denote an equivalence relation by $\sim$ or $\approx$.

Question 5.5. Why do we need (1) in the definition, i.e., why don't (2) and (3) imply (1)?

Whenever a new term is defined, you should try to find examples. It is usually difficult to have an intuitive idea of an abstract concept without examples that you know. We give a few examples of sets having equivalence relations on them.

Examples 5.6. The following are equivalence relations:

1. Any set A under $=$ where if $a, b \in A$ then $a = b$ in A means that we have equality of sets $\{a\} = \{b\}$.
2. Triangles in $\mathbb{R}^2$ under congruence (respectively, similarity).
   [Question: Is the mirror image of a triangle congruent to the original triangle?]
3. The set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ under $\sim$ where
$$(a, b) \sim (c, d) \ \text{ if } \ ad = bc \, (\text{in } \ \mathbb{Z}).$$
4. $\mathbb{Z}$ under $\equiv \mod 2$ where
$$m \equiv n \mod 2 \ \text{ if } \ m - n \text{ is even, i.e., } \ 2 \mid m - n.$$
   Equivalently, if both m and n are odd or both are even in this example.
5. Let R be $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or any field (or, in fact, any ring, e.g., $\mathbb{Z}$, cf. Definition 8.3). Set
$$\mathbb{M}_n(R) := \{n \times n \, \text{matrices with entries in } R\}.$$
   Then $\sim$ is an equivalence relation on $\mathbb{M}_n(R)$, where
$$A \sim B \ \text{ if there exists } \ C \in \mathbb{M}_n(R) \ \text{ invertible such that } \ A = CBC^{-1}.$$
   We call this equivalence relation *similarity of matrices*. We call two matrices A and B *similar* if $A \sim B$. [Cf. Change of Basis Theorem in linear algebra.]
6. The Change of Basis Theorem in linear algebra, in fact, leads to the following equivalence relation. Let R be any field (or ring). Define
$$R^{m \times n} := \{m \times n \ \text{ matrices with entries in } R\}.$$
   Then $\approx$ is an equivalence relation on $R^{m \times n}$, where
$$\begin{aligned} &A \approx B \ \text{ if there exists invertible matrices} \\ &\qquad C \in \mathbb{M}_m(R) \ \text{ and } \ D \in \mathbb{M}_n(R) \text{ satisfying } \ A = CBD. \end{aligned}$$
   We say two matrices A and B in $R^{m \times n}$ are *equivalent* if $A \approx B$.
7. Let R be a field (or ring). Then $\backsim$ is an equivalence relation on $\mathbb{M}_n(R)$, where
$$A \backsim B \ \text{ if there exists } \ C \in \mathbb{M}_n(R) \ \text{ invertible such that } \ A = CBC^t,$$
   where C^t is the transpose of C.
8. Define $\backsim_u$ on $\mathbb{M}_n(\mathbb{C})$ by
$$A \backsim_u B \ \text{ if there exists } \ C \in \mathbb{M}_n(\mathbb{C}) \ \text{ invertible such that } \ A = CBC^*,$$
   where C^* is the adjoint of C. Then $\backsim_u$ is an equivalence relation on $\mathbb{M}_n(\mathbb{C})$.

Definition 5.7. Let $\sim$ be an equivalence relation on a set A. For each a in A, the set

$$[a] = [a]_\sim := \{b \in A \mid a \sim b\}$$

is called the *equivalence class* of a *relative to* $\sim$. We shall usually write

$$\overline{a} \text{ for } [a]$$

and call the following set of subsets of A,

$$\overline{A} = A/\sim := \{\overline{a} \mid a \in A\},$$

the *set of equivalence classes* of $\sim$ on A.

Warning 5.8. The equivalence class of a is a subset of A not (in general) an element of A, i.e., $\overline{a} \subset A$, not $\overline{a} \in A$.

Let $\sim$ be an equivalence relation on A. Then we have a map

$$\overline{} : A \to \overline{A} \text{ given by } a \mapsto \overline{a}.$$

This map is clearly surjective and is called the *natural* or *canonical surjection*. [It is the "obvious" map as it depends on no choices.] Thus if α is an element in $\overline{A}$, there exists an element a in A such that $\alpha = \overline{a}$.

Examples 5.9. 1. Let $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ be given by

$$(a, b) \sim (c, d) \text{ if } ad = bc \,(\text{in } \mathbb{Z}).$$

Then we have

$$\overline{(a,b)} \text{ is just } \frac{a}{b} \text{ in } \mathbb{Q} = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})/\sim .$$

2. Consider $\mathbb{Z}$ under $\equiv \mod 2$. Then we have

$$\{\text{all even integers}\} = \overline{0} = \overline{\pm 0} = \overline{\pm 2} = \overline{\pm 4} = \cdots = \overline{\pm 2n} = \cdots .$$

$$\{\text{all odd integers}\} = \overline{1} = \overline{\pm 1} = \overline{\pm 3} = \overline{\pm 5} = \cdots = \overline{\pm(2n+1)} = \cdots .$$

We shall write $\overline{\mathbb{Z}} = \mathbb{Z}/\equiv \mod 2$ as $\mathbb{Z}/2\mathbb{Z}$, so

$$\mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}.$$

Recall the following:

Definition 5.10. Let A_i, $i \in I$, be sets. [We call I an *indexing set*.] The *union* of the sets A_i is the set

$$\bigcup_{i \in I} A_i := \{x \mid \text{ there exists an } i \in I \text{ such that } x \in A_i\}.$$

We shall usually denote this set by $\bigcup_I A_i$. This union is a *disjoint union* if $A_i \cap A_j = \emptyset$ for all $i, j \in I$ with $i \neq j$. If this is the case, we denote it by $\bigvee_I A_i$. Of course, the *intersection* of the sets A_i is the set

$$\bigcap_{i \in I} A_i := \{x \mid x \in A_i \text{ for all } i \in I\},$$

and we denote by $\bigcap_I A_i$.

Proposition 5.11. *Let $\sim$ be an equivalence relation on A. Then*

$$A = \bigvee_{\overline{A}} \overline{a}.$$

In particular, if $a, b \in A$, then

$$\textit{either } \overline{a} = \overline{b} \ \textit{ or } \ \overline{a} \cap \overline{b} = \emptyset,$$

hence

$$\overline{a} = \overline{b} \ \textit{ if and only if } \ a \sim b.$$

Proof. As $a \sim a$ for all $a \in A$ by Reflexivity and $a \in \overline{a}$ by definition, we have $A = \bigcup_A \overline{a}$. By Symmetry, we have $a \sim b$ if $\overline{a} = \overline{b}$. If $c \in \overline{a} \cap \overline{b}$, then $c \sim a$ and $c \sim b$, so $a \sim c$ by Symmetry hence $a \sim b$ by Transitivity, so $a \in \overline{b}$, i.e., $\overline{a} \subset \overline{b}$. Similarly, we have $\overline{b} \subset \overline{a}$, hence $\overline{a} = \overline{b}$. □

Definition 5.12. Let $\sim$ be an equivalence relation on A. An element $x \in \overline{a}$ is called a *representative* of $\overline{a}$. For example, a is a representative of $\overline{a}$. So by Proposition 5.11, if x is a representative of $\overline{a}$, then $\overline{x} = \overline{a}$. A *system of representatives* for A relative to $\sim$ is a set

$$\mathcal{S} = \{\text{precisely one element from each equivalence class}\},$$

so $A = \bigvee_{\mathcal{S}} \overline{x}$. For example, if $\sim$ is $\equiv \mod 2$ then $\{-36, 15\}$ is a system of representatives of $\equiv \mod 2$ and $\mathbb{Z} = \overline{-36} \vee \overline{15}$.

In later sections, this shall be very useful, so we give it the name of

Mantra 5.13. of Equivalence Relations. In the above setup, we have

$$A = \bigvee_{\mathcal{S}} \overline{x}.$$

In particular, if $|A| < \infty$, then

$$|A| = \sum_{\mathcal{S}} |\overline{x}|.$$

This is only useful if we can compute the size of the equivalence class $\overline{x}$.

Exercises 5.14.

1. A *partition* of a set A is a collection $\mathcal{C}$ of subsets of A such that $A = \bigvee_{\mathcal{C}} B$. Let R be an equivalence relation on A. Show that $\overline{A}$ partitions A. Conversely, let $\mathcal{C}$ partition A. Define a relation $\sim$ on A by $a \sim b$ if a and b belong to the same set in $\mathcal{C}$. Show $\sim$ is an equivalence relation on A. So an equivalence relation and a partition of a set are essentially the same.
2. Draw lines in $\mathbb{R}^2$ perpendicular to the X-axis through all integer points and the analogous lines parallel to the Y-axis. Define a partition of the plane that arises. [Be careful about the points on the lines.]
   [**Question**. Can you do this in such a way that when viewed in $\mathbb{R}^3$ the corresponding equivalence classes looks like a torus?]

6. Modular Arithmetic

We introduce one of the most important equivalence relations in elementary arithmetic. It generalizes $\equiv \mod 2$.

Definition 6.1. Fix an integer $m > 1$ and let $a, b \in \mathbb{Z}$. We say that a is *congruent* to b *modulo* m and write $a \equiv b \mod m$ if $m \mid a - b$ in $\mathbb{Z}$. The set

$$\begin{aligned}\overline{a} = [a]_m &:= \{x \in \mathbb{Z} \mid x \equiv a \mod m\} \\ &= \{x \in \mathbb{Z} \mid x = a + km \text{ some } k \in \mathbb{Z}\}\end{aligned}$$

is a subset of $\mathbb{Z}$ called the *residue class* of *a modulo m*. We shall denote this set as above or as $a + m\mathbb{Z}$. For example,

$$m\mathbb{Z} = 0 + m\mathbb{Z} = \{km \mid k \in \mathbb{Z}\},$$

all the multiples of m.

Question 6.2. In the above, what would happen if we let $m = 1$ or $m = 0$? What if we let $m < 0$?

Let a and b be integers and $m > 1$ an integer. We have three ways of saying the same thing. They are:

$$m \mid a - b \quad \text{or} \quad a \equiv b \mod m \quad \text{or} \quad \overline{a} = \overline{b}.$$

One usually uses the one that is most convenient, although algebraically the last is the most interesting.

Proposition 6.3. *Let $m > 1$ in $\mathbb{Z}$. Then $\equiv \mod m$ is an equivalence relation. In particular,*

$$\mathbb{Z} = \overline{0} \vee \overline{1} \vee \cdots \vee \overline{m-1}.$$

Write $\mathbb{Z}/m\mathbb{Z}$ for $\mathbb{Z}/\equiv \mod m$. Then

$$\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \cdots, \overline{m-1}\} \text{ and } |\mathbb{Z}/m\mathbb{Z}| = m.$$

Let $a, b, c, d \in \mathbb{Z}$ satisfy

$$a \equiv b \mod m \text{ and } c \equiv d \mod m.$$

Then

$$a + c \equiv b + d \mod m \text{ and } a \cdot c \equiv b \cdot d \mod m.$$

Equivalently,

$$\text{if } \overline{a} = \overline{b} \text{ and } \overline{c} = \overline{d}, \text{ then } \overline{a + c} = \overline{b + d} \text{ and } \overline{a \cdot c} = \overline{b \cdot d}.$$

Define

$$+ : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ by } \overline{a} + \overline{b} := \overline{a + b}$$

and

$$\cdot : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ by } \overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

Then $+$ and $\cdot$ are well-defined.

[To be *well-defined* means that they are functions. In this case, this means that if $\overline{a} = \overline{a_1}$ and $\overline{b} = \overline{b_1}$, then $\overline{a} + \overline{b} = \overline{a_1} + \overline{b_1}$ and $\overline{a} \cdot \overline{b} = \overline{a_1} \cdot \overline{b_1}$.]

Moreover, $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$ *satisfies the axioms of a commutative ring, i.e., for all* $a, b, c \in \mathbb{Z}$, *we have*

(1) $(\overline{a} + \overline{b}) + \overline{c} = \overline{a} + (\overline{b} + \overline{c})$.
(2) $\overline{0} + \overline{a} = \overline{a} = \overline{a} + \overline{0}$.
(3) $\overline{a} + (\overline{-a}) = \overline{0} = (\overline{-a}) + \overline{a}$.
(4) $\overline{a} + \overline{b} = \overline{b} + \overline{a}$.
(5) $(\overline{a} \cdot \overline{b}) \cdot \overline{c} = \overline{a} \cdot (\overline{b} \cdot \overline{c})$.
(6) $\overline{1} \cdot \overline{a} = \overline{a} = \overline{a} \cdot \overline{1}$.
(7) $\overline{a} \cdot \overline{b} = \overline{b} \cdot \overline{a}$.
(8) $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$.
(9) $(\overline{b} + \overline{c}) \cdot \overline{a} = \overline{b} \cdot \overline{a} + \overline{c} \cdot \overline{a}$.

PROOF. Exercise. □

The element $\overline{0}$ is called the *zero* or the *unity* of $\mathbb{Z}/m\mathbb{Z}$ under $+$ and $\overline{1}$ is called the *one* or the *unity* of $\mathbb{Z}/m\mathbb{Z}$ under $\cdot$.

Notation 6.4. As is customary, we shall usually drop the multiplication symbol $\cdot$ in $a \cdot b$ when convenient.

Remark 6.5. Let $\sim$ be an equivalence relation on A. If B is a set, to show that $f : \overline{A} \to B$ is *well-defined*, i.e., f is a function, one must show that

$$\overline{a} = \overline{a'} \quad \Longrightarrow \quad f(\overline{a}) = f(\overline{a'}),$$

i.e., is independent of the representative for $\overline{a}$.

The proposition then also says the canonical surjection

$$^{-} : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ given by } x \mapsto \overline{x}\,(= [x]_m)$$

satisfies

$$\overline{a + b} = \overline{a} + \overline{b} \text{ and } \overline{a \cdot b} = \overline{a} \cdot \overline{b}$$

and

$$0 \mapsto \overline{0} \text{ and } 1 \mapsto \overline{1}.$$

Definition 6.6. A *commutative ring* is a set R together with two maps

$$+ : R \times R \to R \text{ and } \cdot : R \times R \to R,$$

write $+(a, b)$ as $a + b$ and $\cdot(a, b)$ as $a \cdot b$, called *addition* and *multiplication*, respectively, satisfying for all $a, b, c, \in R$ the axioms $(1) - (9)$ above hold with R replacing $\mathbb{Z}/m\mathbb{Z}$. In (2), there exists an element 0 satisfying (2) and (3) and in (6) an element 1 satisfying (6). In particular, R is not empty. If we do not necessarily assume that R satisfies (7), we call R a *ring*.

Examples 6.7. 1. The sets $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/m\mathbb{Z}$ are commutative rings.

2. If R is a ring, then so are $\mathbb{M}_n(R)$ and $R[t]$ under the usual $+$ and $\cdot$ of matrices and polynomials, respectively.

We call a map $f : R \to S$ of rings a *ring homomorphism* if it preserves $+$ and $\cdot$ and takes unities to unities. For example,

$$\bar{\ } : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ is a surjective ring homomorphism.}$$

[A surjective ring homomorphism is also called an *epimorphism*.]

To investigate the interrelationship of congruences of different moduli is an appropriate study. We turn to the important and useful case of congruences of pairwise relatively prime moduli. We begin with the following lemma.

Lemma 6.8. *Let m, n, and a_i, $1 \le i \le r$, be integers.*

(1) If $(a_i, m) = 1$ for $i = 1, \dots, r$, then $(a_1 \cdots a_r, m) = 1$.
(2) If $(a_i, a_j) = 1$ for $i \ne j$ and $a_i \mid n$ for $i, j = 1, \dots, r$, then $a_1 \dots a_r \mid n$.

Proof. (1): By induction, it suffices to do the case $r = 2$. (Why?) By Key Observation (4.13), we have equations

$$x_1a_1 + y_1m = 1 = x_2a_2 + y_2m,$$

for some $x_1, x_2, y_1, y_2 \in \mathbb{Z}$, so

$$1 = (x_1a_1 + y_1m)(x_2a_2 + y_2m) = x_1x_2a_1a_2 + \text{something} \cdot m$$

and we are done.
[**Note**. We only used induction for the convenience of notation – do you see why?]

(2): The case $r = 1$ is immediate. By induction, we have $a_1 \cdots a_{r-1} \mid n$. By (1), we have $(a_1 \cdots a_{r-1}, a_r) = 1$ (setting $m = a_r$). By Key Observation (4.13), there exists an equation

$$a_1 \cdots a_{r-1}x + a_ry = 1 \text{ for some } x, y \in \mathbb{Z}.$$

By Key Trick (4.14), we have

$$a_1 \cdots a_{r-1}nx + a_rny = n.$$

As $a_1 \cdots a_{r-1}a_r \mid a_1 \cdots a_{r-1}nx$ and $a_1 \cdots a_{r-1}a_r \mid na_ry$, we conclude that $a_1 \cdots a_r \mid n$, as needed. □

Theorem 6.9. (Chinese Remainder Theorem) *Let m_i be integers with $(m_i, m_j) = 1$ for $1 \le i, j \le r$ and $i \ne j$. Set $m = m_1 \cdots m_r$ and suppose that $c_1, \dots, c_r$ are integers. Then there exists an integer x satisfying all of the following:*

$$(*) \qquad \begin{aligned} x &\equiv c_1 \mod m_1 \\ &\vdots \\ x &\equiv c_r \mod m_r. \end{aligned}$$

Moreover, the integer x is unique modulo m, i.e., if an integer y also satisfies $y \equiv c_i \mod m_i$ for $1 \le i \le r$, then $x \equiv y \mod m$.

Proof. **Existence**: Let $n_i = \frac{m}{m_i} = m_1 \cdots \widehat{m_i} \cdots m_r$, where $\widehat{\ }$ means omit. By Lemma 6.8(1), we have $(m_i, n_i) = 1$ for $1 \le i \le r$, so by Key Observation (4.13), there exist equations

$$1 = d_im_i + e_in_i,$$

for some integers d_i, e_i, $i = 1, \dots, r$. Set $b_i = e_i n_i$ for $i = 1, \dots, r$. Then

$$1 = d_i m_i + b_i \text{ and } m_j \mid b_i \text{ if } i \neq j.$$

This means that

$$1 \equiv b_i \mod m_i \text{ and } 0 \equiv b_i \mod m_j \text{ if } i \neq j.$$

Consequently,

$$x_0 := c_1 b_1 + \cdots + c_r b_r \equiv c_i b_i \equiv c_i \mod m_i, \text{ with } i = 1, \dots, r$$

and x_0 works.

Uniqueness: Suppose that y_0 also works. Then $x_0 \equiv y_0 \mod m_i$ for $1 \leq i \leq r$, i.e., $m_i \mid x_0 - y_0$ for $1 \leq i \leq r$. By Lemma 6.8 (2), we have $x_0 \equiv y_0 \mod m$. $\square$

We want to interpret what we did above in the language of “rings” and “ring homomorphisms”. Let $m > 1$ in $\mathbb{Z}$ and $(a, m) = 1$. By Key Observation (4.13), there are integers x and y satisfying $1 = ax + my$. Applying the canonical surjection $^{-} : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ to this yields

$$\overline{1} = \overline{ax + my} = \overline{a}\,\overline{x} + \overline{m}\,\overline{y} = \overline{a}\,\overline{x} + \overline{0}\,\overline{y} = \overline{a}\,\overline{x}\,(= \overline{x}\,\overline{a}),$$

i.e., $\overline{a}$ has a *multiplicative inverse* in $\mathbb{Z}/m\mathbb{Z}$. An element $\overline{a}$ that has a multiplicative inverse is called a *unit*. Let

$$(\mathbb{Z}/m\mathbb{Z})^\times := \{\overline{a} \in \mathbb{Z}/m\mathbb{Z} \mid \overline{a} \text{ is a unit}\}.$$

So if $a \in \mathbb{Z}$ and $(a, m) = 1$, then $\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times$.

Conversely, suppose $a \in \mathbb{Z}$ satisfies $\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times$. Then there exists $b \in \mathbb{Z}$ satisfying $\overline{a}\overline{b} = \overline{1}$. In particular, we have $m \mid ab - 1$, so $ab - 1 = mk$ for some integer k, or equivalently, $ab - mk = 1$. We therefore have

Conclusion 6.10. Let a and $m > 1$ be integers. Then

$$\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times \text{ if and only if } (a, m) = 1.$$

We can now interpret Lemma 6.8(1) and Properties 4.9(1) as follows: If x and y are integers then

$$\overline{x}, \overline{y} \in (\mathbb{Z}/m\mathbb{Z})^\times \text{ if and only if } \overline{xy} \in (\mathbb{Z}/m\mathbb{Z})^\times.$$

In particular, the set of units $(\mathbb{Z}/m\mathbb{Z})^\times$ of $\mathbb{Z}/m\mathbb{Z}$ is *closed* under multiplication (but not addition), i.e., we can write

$$\cdot : (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/m\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \text{ taking } (\overline{a}, \overline{b}) \to \overline{a} \cdot \overline{b}.$$

We wish to generalize this.

Definition 6.11. Let R be a ring with $1 \neq 0$ and set

$$R^\times := \{r \in R \mid \text{there exists an } a \in R \text{ such that } ar = 1 = ra\},$$

the *units* of the ring R. Note that this set is closed under multiplication.

Question 6.12. What are $\mathbb{Z}^\times$, $\mathbb{Q}^\times$, $\mathbb{R}^\times$, $\mathbb{C}^\times$, $\mathbb{M}_n(\mathbb{R})^\times$, $\mathbb{M}_n(\mathbb{Z})^\times$?

We now interpret the Chinese Remainder Theorem in this new language. As before, let m_i, $1 \le i \le r$, be integers with $(m_i, m_j) = 1$ if $i \ne j$, and $m = m_1 \cdots m_r$. Then $m_j \mid m$ for all j. In particular, if $m \mid a - b$, then $m_j \mid a - b$ for all j, i.e., if $a \equiv b \mod m$, then $a \equiv b \mod m_j$ for all j. This means that the map

$$\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m_j\mathbb{Z} \text{ given by } [a]_m \mapsto [a]_{m_j}$$

is well-defined (i.e., a function). For example, we have $7 \equiv 1 \mod 6$, so $7 \equiv 1 \mod 3$ and $7 \equiv 1 \mod 2$. This leads to a map

$$\begin{gathered} \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_r\mathbb{Z} \\ \text{given by } [a]_m \mapsto ([a]_{m_1}, \ldots, [a]_{m_r}). \end{gathered} \tag{6.13}$$

Giving the right hand set component-wise operations turns it into a commutative ring and we see immediately that this map is a ring homomorphism. (Note that the zero of the right hand side is $([0]_{m_1}, \ldots, [0]_{m_r})$ and the one is $([1]_{m_1}, \ldots, [1]_{m_r})$). The Chinese Remainder Theorem says that this map is bijective. The inverse of this map is also checked to be a ring homomorphism. So the Chinese Remainder Theorem says the map above is a *ring isomorphism*, i.e., a bijective ring homomorphism whose inverse is also a ring homomorphism. (We shall see that this last condition is unnecessary.) This means that the rings

$$\mathbb{Z}/m\mathbb{Z} \text{ and } \mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_r\mathbb{Z} \text{ are } \textit{isomorphic},$$

i.e., they look the same algebraically. For example, if m and n are relatively prime integers greater than 1, then $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ are isomorphic. In particular, if $n = p_1^{e_1} \cdots p_s^{e_s}$ is a standard factorization, then $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_s^{e_s}\mathbb{Z}$ are isomorphic. This last fact is useful as it reduces solving many equations modulo n to congruences modulo the various prime power constituents in its standard factorization.

It is left as an exercise to show the isomorphism in (6.13) above induces a bijection between

$$(\mathbb{Z}/m\mathbb{Z})^\times \text{ and } (\mathbb{Z}/m_1\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/m_r\mathbb{Z})^\times.$$

It follows that *Euler phi-function* (also called the *Euler totient function*) $\varphi : \mathbb{Z}^+ \to \mathbb{Z}^+$ by $\varphi(1) = 1$ and $\varphi(m) = |(\mathbb{Z}/m\mathbb{Z})|^\times|$ for $m > 1$ defined before, is an arithmetic multiplicative function, i.e., $\varphi(mn) = \varphi(m)\varphi(n)$ whenever m and n are relatively prime positive integers.

Exercises 6.14.

1. (Fermat's Little Theorem) Let p be a prime. Then for all integers a, we have

$$a^p \equiv a \mod p.$$

   [You can use the Binomial Theorem if you state it carefully and then prove the *Children's Binomial Theorem* that says for all integers a and b, we have $(a+b)^p \equiv a^p + b^p \mod p$. We will have another proof of this later.]
2. Prove that there exist infinitely many primes p satisfying $p \equiv 3 \mod 4$. [It is true that there exist infinitely many primes p satisfying $p \equiv 1 \mod 4$ but this is harder.]
   [Hint: Look at the proof for the infinitude of primes.]
3. Prove Proposition 6.3.

4. Find the smallest positive integer x satisfying the congruences
$$x \equiv 3 \mod 11 \quad x \equiv 2 \mod 12 \quad \text{and} \quad x \equiv 3 \mod 13.$$
5. What are the units of the rings $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}[t]$, $\mathbb{Z}[t]$, $(\mathbb{Z}/4\mathbb{Z})[t]$, and $\mathbb{M}_n(\mathbb{C})$?
6. Show if m and n are relatively prime positive integers, then the map
$$(\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times \text{ given by } [a]_{mn} \mapsto ([a]_m, [a]_n)$$
is a bijection. In particular, $\varphi(mn) = \varphi(m)\varphi(n)$ whenever m and n are relatively prime positive integers.
7. Let n and r be positive integers and p a positive prime. Show that $\varphi(p^e) = p^{e-1}(p-1)$ and $\varphi(n) = n \prod_{\substack{p \mid n \\ p \text{ prime}}} (1 - \frac{1}{p})$.

7. Surjective maps

In the previous sections, we introduced equivalence relations, using them to construct the integers mod m and the corresponding surjection from the integers to the integers mod n. We want to expand the usefulness of equivalence relations. In mathematics, one tries to reduce the study of functions that preserve a given structure into injective and surjective maps that preserve that structure. The reason for studying injective maps is that if one can *embed*, i.e., find an injection of an object into one with more structure, one can often pull back information. The reason to study surjective maps is that usually the target is simpler, e.g., $\mathbb{Z}/m\mathbb{Z}$ is simpler than $\mathbb{Z}$. If we are just studying sets, i.e., without any additional structure, this reduction can always be accomplished. Equivalence relations give the surjective maps.

Exercise 5.14(1) showed that equivalence relations are essentially the same as partitioning. Let A be a set with an equivalence relation $\sim$ on it. Then the set of equivalence classes, $\overline{A} := \{\overline{a} \mid a \in A\}$, partitions A, and the exercise says that partitions of A induce equivalence relations on A. Let $\mathcal{S}$ be a system of representatives for A relative to $\sim$, so $A = \bigvee_{\mathcal{S}} \overline{a}$. We then have the canonical surjection
$$^- : A \to \overline{A} \text{ given by } a \mapsto \overline{a},$$
i.e., every equivalence relation leads to a surjective map. We want the converse. Let
$$f : A \to B \text{ be a surjective map.}$$
We define an equivalence relation $\sim$ on A by
$$a \sim a' \text{ if } f(a) = f(a').$$
This is clearly an equivalence relation on A, as $=$ is an equivalence relation on B. We compute the equivalence class of a to be
$$\overline{a} = \{x \in A \mid x \sim a\} = \{x \in A \mid f(x) = f(a)\},$$
called the *fiber* of f at $f(a)$, and let
$$f^{-1}\bigl(f(a)\bigr) = \{x \in A \mid f(x) = f(a)\}.$$

[In general, f^{-1} is not a function, but we use the general notation if $f : A \to B$ is any map and $D \subset B$, then $f^{-1}(D) := \{x \in A \mid f(x) \in D\}$. If $D = \{b\}$ (the *preimage* of D), then the *fiber* $f^{-1}(b)$ at b is $f^{-1}(\{b\})$.]

We now know, if $\sim$ is the equivalence relation defined by the surjective map f, then we have

(*) $\qquad \overline{a} = \overline{a'}$ if and only if $a \sim a'$ if and only if $f(a) = f(a')$.

But this means if we define

$$\overline{f} : \overline{A} \to B \text{ by } \overline{a} \mapsto f(a),$$

then

$$\overline{f}(\overline{a}) = f(a),$$

so $\overline{f}$ is a well-defined injective map by (*). We say that f *induces* $\overline{f}$. This induced map is also surjective since $\overline{f}(\overline{A}) = f(A)$, hence it is a bijection. So set theoretically we cannot tell the sets $\overline{f}(\overline{A})$ and B apart. Moreover, we have a *commutative diagram*

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "-"'] & B \\ \overline{A} \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

i.e., $f = \overline{f} \circ {}^{-}$. We call this the *First Isomorphism Theorem of Sets*. [In general, we say a diagram *commutes* if following any composition of any maps (arrows) from the same place to the same target gives equal maps.]

This is so useful that we summarize the above.

Summary 7.1. Let $f : A \to B$ be a surjective map. Define an equivalence relation $\sim$ on A by $a \sim a'$ if $f(a) = f(a')$. Then

(1) $\sim$ is an equivalence relation.
(2) $\overline{A} = \{f^{-1}\big(f(a)\big) \mid a \in A\}$.
(3) ${}^{-} : A \to \overline{A}$ given by a is mapped to $\overline{a}$ is a surjective map.
(4) $\overline{f} : \overline{A} \to B$ given by $\overline{a}$ maps to $f(a)$ is a well-defined bijection.
(5) The diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "-"'] & B \\ \overline{A} \arrow[ur, "\overline{f}"'] & \end{tikzcd} \quad \text{commutes.}$$

We can push this a bit further. Let $g : A \to C$ be a map and set $B = g(A)$. Let $f : A \to B$ be given by $f(a) = g(a)$ for all $a \in A$, i.e., we change the target. Then g is the composition

$$A \xrightarrow{f} B \xrightarrow{\text{inc}} C,$$

with inc the inclusion map.

Notation 7.2. In the future, we shall usually abuse notation and write the same letter for a function and the function arising by changing its target (if it makes sense) when no confusion can arise.

For clarity, we still write g for this f here. Let $\sim$ be the equivalence relation given by $a \sim a'$ if $f(a) = f(a')$. Then we have a commutative diagram

$$\begin{tikzcd} A \arrow[r, "g"] \arrow[d, "-"'] \arrow[dr, "f"] & C \\ \overline{A} \arrow[r, "\overline{f}"'] & B. \arrow[u, hook, "\mathrm{inc}"'] \end{tikzcd}$$

If we set $\overline{g} = \mathrm{inc} \circ \overline{f}$, we get a commutative diagram

(7.3)
$$\begin{tikzcd} A \arrow[r, "g"] \arrow[d, "-"'] & C \\ \overline{A} \arrow[ur, "\overline{g}"'] & \end{tikzcd}$$

with $^{-}$ a surjection and $\overline{g}$ an injection. This is the *First Isomorphism Theorem (alternate version)*, showing that every map *factors* as a composition injection $\circ$ surjection. Note that the equivalence classes of $\sim$ are precisely the nonempty fibers of the map g.

If we give our sets additional structure, we would be interested in maps that preserve this additional structure. In general, the map $\overline{g}$ may not preserve the additional structure, and the set $\overline{A}$ may not have a compatible structure. We shall see in the sequel that in certain cases we can achieve this.

The map $g : A \to C$ is also a composition of surjection $\circ$ injection. Indeed let $\Gamma_g : A \to A \times C$ be given by $a \mapsto \big(a, g(a)\big)$ (called the *graph of* g) and $\pi_C : A \times C \to C$ by $(a, c) \mapsto c$, the *projection* of $A \times C$ onto C. Then $g = \pi \circ \Gamma_g$. This shows that to study (set) maps, it suffices to study injections and surjections.

Part 2

Group Theory

CHAPTER III

Groups

In this chapter, we begin our study of abstract algebra. The basic object in this study is a *group*, a set G with one (binary) operation $\circ$ subject to three axioms: associativity, existence of a unity, and existence of inverses. An example that you know from linear algebra is the set $\mathrm{GL}_n(\mathbb{R})$ of all $n \times n$ matrices over the reals with nonzero determinant which is called the $n \times n$ *general linear group* over $\mathbb{R}$. Although groups are easy to define, their structure is far from simple. A primary reason for this is that, in general, we do not assume that $x \circ y = y \circ x$ in G.

In algebra, as in other fields of mathematics, one studies sets with additional structure by investigating maps between those sets that preserve the additional structure. In particular, one would like to know when two objects with additional structure are essentially the same. Here the prototypes that you have encountered in linear algebra are linear transformations between vector spaces and isomorphisms of vector spaces. One idea to keep in mind is that in modern mathematics maps between objects are more important than the objects themselves. In this chapter, we give many examples of groups and maps between them and establish the basic theorems needed to investigate groups.

It is important that you learn the many examples given below (as well as others), as you can only theorize what is true based on examples that you know – of course, your guess may very well not be true. In this course, we shall mostly apply our theorems to groups having finitely many elements because this allows us to use the power of equivalence relations in an effective way – we can count. The general theory of arbitrary groups is very difficult, but it is extremely important not only in algebra but in other fields of study because of the general linear group (and its subgroups).

8. Definitions and Examples

A map $\cdot : G \times G \to G$ is called a *binary operation*. We shall always write $a \cdot b$ for $\cdot(a, b)$.

Definition 8.1. Let G be a set with a binary operation $\cdot : G \times G \to G$. We call $(G, \cdot)$ a *group* if it satisfies the following axioms:

Associativity. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$.

Unity. There exists an element $e \in G$ such that for all $a \in G$, we have $a \cdot e = a = e \cdot a$. The element e is called a *unity* or an *identity*.

Inverses. There exists a unity e in G; and, for any $x \in G$, there exists a $y \in G$ satisfying $xy = e = yx$.

If, in addition, G satisfies

Commutativity. $a \cdot b = b \cdot a$ for all a, b in G, we call G an *abelian* group.

We usually write G for $(G, \cdot)$ and ab for $a \cdot b$ if $\cdot$ is clear.

The most common algebraic property on a set with a binary operation is associativity, although there are interesting algebraic objects that do not satisfy this property. However, all the algebraic objects that we shall study will satisfy an analogue of associativity. If our set G under its binary operation only satisfies Associativity and Unity, then it is called a *monoid*.

Remarks 8.2. Let $\cdot : G \times G \to G$ be a binary operation.

1. If G satisfies associativity and $a_1, \dots, a_n \in G$, then $a_1 \cdots a_n$ makes sense, i.e., is independent of parentheses. In particular, if $n \in \mathbb{Z}^+$ then a^n makes sense with $a^1 = a$ and $a^n = a(a^{n-1})$ for $n > 1$. If G is a monoid, we let $a^0 = e$. The nasty uninteresting induction on the complexity of parentheses is left to the diligent reader.
2. If G satisfies Unity, then the unity e is unique. Indeed if e' is another unity then $e = ee' = e'$.
   [The unity of an abstract group will usually be written e_G or e but for many specific types of groups the unity will be written as 0 or 1.]
3. If G is a monoid, then $a \in G$ has at most one inverse. If it has an inverse, then it is denoted by a^{-1}, in which case a is the inverse of a^{-1}. Indeed if b and c are inverses of a then
   $$b = b \cdot e = b \cdot (a \cdot c) = (b \cdot a) \cdot c = e \cdot c = c.$$
   [In fact, the proof shows if there exist $b, c \in G$ satisfying $ba = e = ac$ then $b = c$ is the inverse of a.]
4. If G is a monoid and $a, b \in G$ have inverses then so does ab and $(ab)^{-1} = b^{-1}a^{-1}$. In addition, if $n \in \mathbb{Z}^+$ then $a^{-n} = (a^{-1})^n$.

   Question. Let a, b be elements in a monoid such that ab has an inverse. Is it true that a and b have inverses?
5. If G is a group, then the *cancellation laws* hold, i.e., for all $a, b, c \in G$, we have
   $$ab = ac \implies b = c \qquad \text{and} \qquad ba = ca \implies b = c.$$
   In particular, if $a = xy$ in G, then $x = ay^{-1}$ and $y = x^{-1}a$.
6. If G is a group under a binary operation notated by $+ : G \times G \to G$, then G will always be an abelian group. We then call G an *additive* group and write 0 (or 0_G) for e_G and $-a$ for a^{-1}.

Definition 8.3. Using this new language, it is easy to define a *ring* more carefully. Let R be a set with two binary operations $\cdot : R \times R \to R$ and $+ : R \times R \to R$. Then R is a ring under *addition* $+$ and *multiplication* $\cdot$ if $(R, +)$ is an additive group, $(R, \cdot)$ is a monoid, and R satisfies the *Distributive Laws*
$$a \cdot (b + c) = a \cdot b + a \cdot c \text{ and } (b + c) \cdot a = b \cdot a + c \cdot a,$$
for all $a, b, c \in R$. The distributive laws interrelate the two binary operations. The multiplicative unity is written 1 (or 1_R if there can be confusion). The ring is called a *commutative ring* if $(R, \cdot)$ is a commutative monoid. One checks that $1 = 0$ in R if and

only if $R = \{0\}$, the *zero* or *trivial ring*. If R is not the zero ring it is called a *division ring* if $(R \setminus \{0\}, \cdot)$ is a group. A commutative division ring is called a *field*.

It is worth mentioning again that when one has defined a new concept, one should always try to find many examples. Theorems usually arise from knowledge of explicit examples. We present many examples of groups.

Examples 8.4. 1. A *trivial* group is a group consisting of a single element, necessarily the unity.

2. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/m\mathbb{Z}\,(m > 1)$, or any ring is an additive group under $+$.
3. $\mathbb{R}^+$, the set of positive real numbers, is an abelian group but $\mathbb{Z}^+$ is only an abelian (i.e., commutative) monoid under multiplication with unity $e = 1$, and not a group. Neither are monoids under addition, although they still satisfy associativity.
4. If F is a field, e.g., if $F = \mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$, then by definition $F^\times = F \setminus \{0\}$ is an abelian group under multiplication, e.g., if $F = \mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$. More generally, if R is any ring then its set of units $R^\times$, i.e., the set of elements having a multiplicative inverse, is a group under $\cdot$, abelian if R is commutative, called the *group of units of R*. This leads to many examples of groups.
5. Let V be a vector space over a field F. Then $(V, +)$ is an additive group.
6. Let S be a nonempty set. Then
$$\Sigma(S) := \{f : S \to S \mid f \text{ is a bijection}\}$$
is a group under the composition of functions. The unity is the identity map on S which we shall write as 1_S. A bijection $f : S \to S$ is called a *permutation*. [If $f \in \Sigma(S)$, what is f^{-1}?] We call $\Sigma(S)$ the *group of all permutations* of S. It is also a *transitive* group on S. That is, for all $x, y \in S$, there exists a permutation $f \in \Sigma(S)$ satisfying $f(x) = y$. We view the group $\Sigma(S)$ *acting* on the set S via
$$\Sigma(S) \times S \to S \text{ by } (f, s) \mapsto f(s).$$
The theory of abstract groups arose by axiomatizing how such "concrete" groups act on sets, often, as in this case, as functions acting on the set. If $S = \{1, \ldots, n\}$ then $\Sigma(S)$ is denoted S_n and called the *symmetric group on n letters* . Note that $|S_n| = n!$.
Question. Let $S = \{a, b, \ldots, z\}$. Can you tell the difference between $\Sigma(S)$ and S_{26} algebraically?

For the next examples, we need the definition of a subgroup of a group. A subset H of a group G is called a *subgroup* of $(G, \circ)$ if $H = (H, \star)$ is a group under the restriction map $\star = \circ|_{H\times H} H \times H \to H$, i.e., H becomes a group under the restriction of the binary operation on G to H, i.e., $\cdot|_{H\times H} : H \times H \to H$ makes sense (meaning that the image of $\cdot|_{H\times H}$ lies in H), and has the same unity as G (although this last condition will be seen to be unnecessary).

7. Let S be a nonempty set and $x_0 \in S$. The set
$$\Sigma(S)_{x_0} := \{f \in \Sigma(S) \mid f(x_0) = x_0\} \subset \Sigma(S)$$
is a group called the *stabilizer* of x_0 in $\Sigma(S)$. We say elements of $\Sigma(S)_{x_0}$ *fix* x_0. We shall see that groups often arise as functions acting on sets and a major problem is to

find *fixed points* of these actions, e.g., in the above x_0 is a *fixed point* of the action of $\Sigma(S)_{x_0}$ on S. Note that $(S_n)_n$ looks algebraically like S_{n-1}.

$\Sigma(S)_{x_0}$ is also a *subgroup* of $\Sigma(S)$.

We can generalize the stabilizer of x_0 as follows: Let $x_0, \dots, x_n$ be elements of S. Then

$$\Sigma(S)_{x_0} \cap \cdots \cap \Sigma(S)_{x_n} = \{f \in \Sigma(S) \mid f(x_i) = x_i, \text{ for } i = 1, \dots, n\}$$

is a subgroup of $\Sigma(S)$ (as is $\Sigma(S)_{x_i}$ for all i) stabilizing (fixing) $x_0, \dots, x_n$.

8. Let G be a group, H_i with $i \in I$ be subgroups of G. Then $\bigcap_I H_i$ is a subgroup of G. In general, $\bigcup_I H_i$ is not a subgroup of G. [Can you find a condition when it is?]
9. Let G be a group and $W \subset G$ a subset. Set

$$\mathcal{W} = \{H \subset G \mid H \text{ is a subgroup of } G \text{ with } W \subset H\}.$$

As $W \subset G$, we have $\mathcal{W} \neq \emptyset$. Let

$$\langle W \rangle := \bigcap_{\mathcal{W}} H = \bigcap_{\substack{W \subset H \subset G \\ H \text{ subgroup of } G}} H.$$

This is the unique smallest subgroup of G containing W. [Can you show this?] We say that W *generates* $\langle W \rangle$ and that W is a *set of generators* for $\langle W \rangle$. A set of generators is not unique. If $W = \{a_1, \dots, a_n\}$ is finite, we also write $\langle a_1, \dots, a_n \rangle$ for $\langle \{a_1, \dots, a_n\} \rangle$. We say G is *finitely generated* if there exists a finite set W such that $G = \langle W \rangle$ and *cyclic* if there exists an element $a \in G$ such that $G = \langle a \rangle$. If this is the case then $G = \{a^n \mid n \in \mathbb{Z}\}$ and is abelian. For example, $(\mathbb{Z}, +)$ and $(\mathbb{Z}/m\mathbb{Z}, +)$, $m > 1$, are cyclic with generators 1 and $\overline{1}$, respectively.

Question. What are the analogues of the above for vector spaces? The analogy will not go too far, as the concept of linear independence almost always fails.

Warning. If G is a group that can be generated by two elements, G need not be abelian. As an example consider a figure eight in the plane and take the group generated by the two counterclockwise rotations of 360^o, each one rotating around its respective circle starting from the intersection point. [Can you see why this group is not abelian?]

10. Let

$$T := \{z \in \mathbb{C} \mid |z| = 1\},$$

where $|z| = \sqrt{z\overline{z}}$ with $\overline{z}$ the complex conjugate of z. This is an abelian group under multiplication, called the *circle group*. It is a subgroup of $\mathbb{C}^\times$. If $n \in \mathbb{Z}^+$ then

$$\mu_n := \{z \in T \mid z^n = 1\} = \langle e^{2\pi\sqrt{-1}/n} \rangle$$

is a cyclic subgroup of T called the *group of nth roots of unity*. Another subgroup of T is

$$\{z \in T \mid z \in \mu_n \text{ for some } n \in \mathbb{Z}^+\} = \bigcup_{\mathbb{Z}^+} \mu_n.$$

Note that a subgroup of an abelian group is abelian.

11. The symmetries of a geometric object often form a group. We look at some special cases. Consider an equilateral triangle in the plane with a side on the X-axis labeling the ordered vertices A, B, C, with line segment AB the base. Let r be a counterclockwise rotation of 120^o. The triangle looks the same but the vertices are now ordered C, A, B, with base CA. Composing rotations shows $r^2 = r \circ r$ takes the triangle to B, C, A, and r^3 back to A, B, C, i.e., we have the *relation* r^3 is the identity. So $\langle r \rangle$ is a cyclic group of three elements. [Should we be able to compare this group to μ_3?] Now, viewing the plane in $\mathbb{R}^3$, let f denote the flip along the line of the apex of the triangle perpendicular to the base. It takes the ordered vertices A, B, C to B, A, C. Under composition, we have the *relation* f^2 is the identity which we write as 1 and $\langle f \rangle$ is a cyclic group of two elements. [Should we be able to compare this group to μ_2 and to $(\mathbb{Z}/2\mathbb{Z}, +)$ and to $(\mathbb{Z}^\times, \cdot)$?] Composing f and r leads to the *relations* $f^{-1} \circ r \circ f = r^2 = r^{-1}$ and a non-abelian group with six elements, viz.,

$$\{1, r, r^2, f, fr, rf\}.$$

[Show this.] We say that r and f generate this group subject to the relations $r^3 = 1$, $f^2 = 1$, and $f^{-1}rf = r^{-1}$. Note that it follows that $rf = fr^2$ and $fr = r^2 f$. It is customary to write this as

$$\langle r, f \mid r^3 = 1,\ f^2 = 1, f^{-1}rf = r^{-1} \rangle,$$

i.e., in the form

$$\langle \text{generators} \mid \text{relations} \rangle.$$

This group is called the symmetries of an equilateral triangle or the *dihedral group of order six* and denoted by D_3. [Compare D_3 with S_3.] More generally, consider a regular n-gon, $n \geq 3$, with r a counterclockwise rotation of $2\pi/n$ radians and f a flip along the perpendicular at the bisection point of the base. Then under composition, we get a non-abelian group with $2n$ elements. It is a group that is defined by two *generators* r and f satisfying the three *relations*

$$r^n = 1 \qquad f^2 = 1 \qquad f^{-1} \circ r \circ f = r^{-1}.$$

It follows that $f^{-1} \circ r \circ f = r^{n-1}$. It is called the symmetries of the regular n-gon or the *dihedral group of order* $2n$ and denoted by D_n. As above, this is usually written

$$D_n := \langle r, f \mid r^n = 1,\ f^2 = 1, f^{-1}rf = r^{-1} \rangle.$$

Note if $n > 3$ then D_n and S_n have a different number of elements.

12. Let $Q = \{1, -1, i, -i, j, -j, k, -k\}$, a set with eight elements. This becomes a non-abelian group called the *quaternion group* if we invoke the relations $(-1)^2 = 1$, $k = ij = -ji$ and $i^2 = j^2 = -1$.

13. Let $F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, or any field. Then

$$\mathrm{GL}_n(F) := \{A \in \mathbb{M}_n(F) \mid \det A \neq 0\}$$

is a group under matrix multiplication called the *general linear group of degree n*. It and its subgroups are probably the most important groups in mathematics. If $n = 1$ then $\mathrm{GL}_1(F) = F^\times$ but if $n > 1$, this group is not abelian. [Proof?]

More generally, let R be any ring. Then the set of units $\big(\mathbb{M}_n(R)\big)^\times$ is a group under multiplication of matrices. If R is commutative, it turns out that the determinant of a matrix still makes sense and $A \in \big(\mathbb{M}_n(R)\big)^\times$ if and only if $\det(A) \in R^\times$, so we let

$$\mathrm{GL}_n(R) := \big(\mathbb{M}_n(R)\big)^\times$$

also called the *general linear group of degree n*. If $A \in \mathbb{M}_n(R)$, we denote its ijth entry by A_{ij}. Then as usual, its *transpose* is A^t, where $(A^t)_{ij} = A_{ji}$ and if $R = \mathbb{C}$, its *adjoint* or complex conjugate transpose is A^* where $(A^*)_{ij} = \overline{A_{ji}}$. If A is a diagonal matrix, we also write $A = \mathrm{diag}(a_1, \dots, a_n)$, e.g., the identity is the matrix $I = I_n = \mathrm{diag}(1, \dots, 1)$.

We define some of the interesting subgroups of the general linear group over a commutative ring R [For the subgroups listed below having the restriction to matrices of determinant 1, we also assume that 2 is a unit in the ring.] (the same definitions hold for over an arbitrary ring except when the determinant is involved.)

$$\begin{aligned}
\mathrm{SL}_n(R) &= \{A \in \mathrm{GL}_n(R) \mid \det A = 1\} && \textit{special linear group}\\
\mathrm{O}_n(R) &= \{A \in \mathrm{GL}_n(R) \mid A^t A = I\} && \textit{orthogonal group}\\
\mathrm{SO}_n(R) &= \mathrm{SL}_n(R) \cap \mathrm{O}_n(R) && \textit{special orthogonal group}\\
\mathrm{D}_n(R) &= \{A \in \mathrm{GL}_n(R) \mid A \ \text{ diagonal}\} && \textit{diagonal group}\\
\mathrm{T}_n(R) &= \{A \in \mathrm{GL}_n(R) \mid A_{ij} = 0 \ \text{ if } \ i > j\} && \textit{upper triangular group}\\
\mathrm{ST}_n(R) &= \{A \in \mathrm{T}_n(R) \mid A_{ii} = 1 \ \text{ all } \ i\} && \textit{strictly upper triangular group}\\
&\text{Let } J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \ \text{ in } \ \mathrm{GL}_{2n}(R). \text{ Then}\\
\mathrm{Sp}_{2n}(F) &= \{A \in \mathrm{GL}_n(R) \mid A^t J A = J\} && \textit{symplectic group}
\end{aligned}$$

We also have the more specialized subgroups.

$$\begin{aligned}
&\text{Let } I_{3,1} = \mathrm{diag}(1,1,1,-1) \ \text{ in } \ \mathrm{GL}_4(\mathbb{R}). \text{ Then}\\
\mathrm{O}_{3,1}(\mathbb{R}) &= \{A \in \mathrm{GL}_4(\mathbb{R}) \mid A^t I_{3,1} A = I_{3,1}\} && \textit{Lorenz Group}\\
\mathrm{U}_n(\mathbb{C}) &= \{A \in \mathrm{GL}_n(\mathbb{C}) \mid A^* A = I\} && \textit{unitary group}\\
\mathrm{SU}_n(\mathbb{C}) &= \mathrm{SL}_n(\mathbb{C}) \cap \mathrm{U}_n(\mathbb{C}) && \textit{special unitary group.}
\end{aligned}$$

14. Let V be a vector space over a field F. Then

$$\mathrm{Aut}_F(V) := \{T : V \to V \mid T \ \text{ a linear (i.e., vector space) isomorphism}\}$$

is a group under composition called the *automorphism group* of V. An isomorphism of a vector space to itself is called an *automorphism*. This generalizes greatly, and its generalization probably is the most important type of group, e.g., cf. $\mathrm{Aut}_F(V)$ and $\mathrm{GL}_n(F)$ if V is an n-dimensional vector space.

15. Let G_i, $i \in I$, be groups and $G = \bigtimes_I G_i$ the *cartesian product* of the sets G_i, $i \in I$. We write elements in G by $(g_i)_I = (g)i_{i\in I}$. [Technically,

$$\bigtimes_I G_i := \{f : I \to \bigcup_I G_i \mid f(i) \in G_i \text{ for all } i \in I\},$$

and if I is infinite, $\bigtimes_I G_i$ is nonempty by the Axiom of Choice (Appendix A (A.8)).] Then G is a group under *componentwise operation*, e.g., $e_G = (e_{G_i})_I$. This group is called the *external direct product* of the G_i's. If all the G_i are abelian, so is $\bigtimes_I G_i$.

For example, writing $\mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}$, we have an abelian group

$$V := \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} = \{(\overline{0},\overline{0}), (\overline{0},\overline{1}), (\overline{1},\overline{0}), (\overline{1},\overline{1})\}.$$

It satisfies:

$$\begin{gathered}(a,b)+(a,b) = (\overline{0},\overline{0}) \text{ if } a,b \in \mathbb{Z}/2\mathbb{Z}\\ (\overline{1},\overline{0})+(\overline{0},\overline{1}) = (\overline{1},\overline{1})\\ (\overline{1},\overline{1})+(\overline{0},\overline{1}) = (\overline{1},\overline{0})\\ (\overline{1},\overline{1})+(\overline{1},\overline{0}) = (\overline{0},\overline{1})\\ \textit{etc.}\end{gathered}$$

This group is called the *Klein Four (Vier) Group*. Note that this is an abelian group that is not cyclic. Also it is worth noting that V is a two-dimensional vector space over the field $\mathbb{Z}/2\mathbb{Z}$.

16. Let $m > 0$. Then we know that $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic and $\big((\mathbb{Z}/m\mathbb{Z})^\times, \cdot\big)$ is abelian but the latter may not be cyclic. What is $(\mathbb{Z}/8\mathbb{Z})^\times$ or more generally $(\mathbb{Z}/2^n\mathbb{Z})^\times$ if $n > 2$? The answer is interesting, as it caused a well-known theorem in number theory called Grünwald's Theorem to be in fact false. The correction by Wang had to take the answer to this question into account.
17. If $a, b \in \mathbb{Z}^+$ have d as their gcd, then the additive group $\langle a, b\rangle = \langle d\rangle$.

Exercises 8.5.

1. Let a, b be elements in a monoid such that ab has an inverse. Is it true that a and b have inverses? Prove this if true or give a counterexample if false.
2. If H_i, $i \in I$, are subgroups of a group G, show that $\bigcap_I H_i$ is a subgroup of G, but $\bigcup_I H_i$ is not a subgroup in general (i.e., give a counterexample). Find a nontrivial condition on the H_i such that $\bigcup_I H_i$ is a subgroup.
3. Show if G is a group in which $(ab)^2 = a^2b^2$ for all $a, b \in G$, then G is abelian.
4. Determine all groups having at most six elements.
5. Show that the quaternion group is a group of eight elements.
6. Show that the dihedral group D_n has $2n$ elements.
7. Let G be a group and H a subgroup of G. Define the *core* of H in G by $\mathrm{Core}_G(H) := \bigcap_G xHx^{-1}$. Show $\mathrm{Core}_G(H)$ is the unique largest subgroup N of H satisfying $xNx^{-1} = N$ for all x in G.

8. Let G be a group and H a subgroup of G. Define the *normalizer* of H in G by $N_G(H) := \{x \in G \mid xHx^{-1} = H\}$. Show that $N_G(H)$ is a subgroup of G containing H and is the unique largest subgroup K of G containing H satisfying $H = xHx^{-1}$ for all x in K.
9. Show if G is a group in which $(ab)^n = a^n b^n$ for all $a, b \in G$ and $n = i, i+1, i+2$ for some positive integer i, then G is abelian.
10. Let $W \subset G$. Show that $\langle W \rangle$ is the unique smallest subgroup of G containing W.
11. If G is a group and W is a nonempty subset of G, a subset, show that
$$\langle W \rangle = \{g \in G \mid \text{There exist } w_1, \dots, w_r \in W \text{ (some } r) \text{ not necessarily distinct such that } g = w_1^{e_1} \cdots w_r^{e_r},\ e_1, \dots, e_r \in \mathbb{Z}\},$$
i.e., W is an "alphabet" for the "words" (elements) in $\langle W \rangle$. Unfortunately, spelling may not be unique in G for any "alphabet".
12. Let F be a field (or any nontrivial ring). Show that $\mathrm{GL}_n(F)$ is not abelian for any $n > 1$.
13. Let p be a prime and $F = \mathbb{Z}/p\mathbb{Z}$. Show that F is a field.
[Hint: First show that F is a *domain*, i.e., a commutative ring satisfying: whenever $a, b \in F$ satisfy $ab = 0$ in F then $a = 0$ or $b = 0$. Then show that any domain with finitely many elements is a field.]
14. Compute $(\mathbb{Z}/8\mathbb{Z})^\times$. Find a different group isomorphic to it. Can you do the same for $(\mathbb{Z}/2^n\mathbb{Z})^\times$ if $n = 4, 5$? if n is any integer at least three?
15. Let $\mathcal{F} := \{f \mid f : \mathbb{Z}^+ \to \mathbb{C}\}$ and I in $\mathcal{F}$ be given by $I(1) = 1$ and $I(n) = 0$ for all $n > 1$. Show that $\mathcal{F}$ is an abelian monoid and $\mathcal{A} := \{f \mid f \in \mathcal{F} \text{ with } f(1) \neq 0\}$ is an abelian group with unity I under the *Dirichlet product* given by
$$(f \star g)(n) := \sum_{d \mid n} f(d) g(\frac{n}{d}).$$

9. First Properties

Recall that we defined a subgroup of a group as follows:

Definition 9.1. A subset H of a group G is called a *subgroup* of G if it becomes a group under the restriction of the binary operation on G to H, i.e., $\cdot|_{H \times H} : H \times H \to H$ makes sense (meaning that the image of $\cdot|_{H \times H}$ lies in H), and has the same unity as G.

We begin this section with a criterion for a subset of a group to be a subgroup. This can be compared to the analogous criterion for sets to be subspaces of a vector space.

Proposition 9.2. *Let G be a group and H be a nonempty subset of G. Then H is a subgroup of G if and only if the following two conditions hold:*

(i) *If a and b are elements of H, then ab is an element of H.*
[We say that H is *closed* under $\cdot$, i.e., this means that the restriction
$$\cdot|_{H \times H} : H \times H \to H$$

is defined. Note we have changed the target from G to H.]

(*ii*) *If* $a \in H$ *then* $a^{-1} \in H$.

Moreover, these two conditions are equivalent to

$$(*) \qquad \text{If } a, b \in H \text{ then } ab^{-1} \in H.$$

PROOF. ($\Rightarrow$) follows from the definition of subgroup, since by definition, $e_G = e_H$.
($\Leftarrow$): We first note that (i) and (ii) imply (*) for if $a, b \in H$, then $a, b^{-1} \in H$, hence $ab^{-1} \in H$. Conversely, suppose that (*) holds. As $a \in H$ implies $a, a \in H$, we have $e_G = aa^{-1} \in H$. By definition, $e_G a = a = a e_G$ for all $a \in H$, so $e_G = e_H$. As $e_G, a \in H$, we have $e_G a^{-1} = a^{-1}$ lies in H. We conclude that if $a, b \in H$, then $a, b^{-1} \in H$, hence $ab = a(b^{-1})^{-1} \in H$. Finally, since associativity holds in G, it holds in the subset H. □

Note that property (*) implies that we do not need to assume in the definition of a subgroup that it has the same unity as the group.

Corollary 9.3. *Let G be a group and $H \subset G$ a nonempty subset. If H is a finite set, then H is a subgroup if and only if H is closed under $\cdot$.*

PROOF. We need only check:
($\Leftarrow$): It suffices to show if $a \in H$ then $a^{-1} \in H$. Our hypothesis implies that

$$S := \{a^n \mid n \in \mathbb{Z}^+\} \subset H$$

and is also a finite set. We need the following, whose proof we leave as an exercise:

Dirichlet's Pigeonhole Principle: If $N \geq n+1$ objects are placed in $\leq n$ boxes, then at least one box contains at least two elements.

In our case, $|S| < \infty$, so there exist integers $n > m \geq 1$ satisfying $a^n = a^m$. Therefore, $a^{n-m} = a^n a^{-m} = e_G = aa^{n-m-1}$ in G. It follows that e_G lies in H and $a^{n-m-1} = a^{-1}$ lies in H if $n - m - 1 > 0$. □

Note that this proof shows that if G is a group with $a \in G$ and $\langle a \rangle$ finite, then there exists $N \in \mathbb{Z}^+$ such that $a^N = e$ and by well-ordering there exists a minimal such N.

If G is a group, we call $|G|$ the *order* of G, so G is a *finite* group if it has finite order, *infinite* otherwise. If $a \in G$, we call $|\langle a \rangle|$ the *order* of a and say a has *finite order* if $|\langle a \rangle|$ is finite.

Examples 9.4. 1. $(\mathbb{Z}, +) = \langle 1 \rangle$ is an infinite cyclic group.

2. If $n \in \mathbb{Z}^+$ then $(\mathbb{Z}/n\mathbb{Z}, +) = \langle \overline{1} \rangle$ is a cyclic group of order n as is $\mu_n = \langle e^{2\pi\sqrt{-1}/n} \rangle$.
3. $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is an abelian group of order four and not cyclic.
4. If $n > 2$ then D_n is a non-abelian group of order $2n$.
5. The quaternion group is a non-abelian group of order 8. [Does it look the same algebraically as D_4?]
6. If $n > 2$, then S_n is a non-abelian group of order $n!$.
7. $(\mathbb{Z}/n\mathbb{Z})^\times$ is an abelian group of order $\varphi(n)$.
8. $\mathrm{GL}_n(\mathbb{R})$, $n > 1$, is an infinite non-abelian group.

We need to know how to determine when two groups are the same algebraically.

Let $\varphi : G \to G'$ be a map of groups. We call φ a *group homomorphism* if

$$\begin{aligned}\varphi(a \cdot b) &= \varphi(a) \cdot \varphi(b) \text{ for all } a, b \in G\\ \varphi(e_G) &= e_{G'}.\end{aligned}$$

Note in the first equation $\cdot$ on the left hand side is in G and on the right hand side is in G'. We shall also see that we really do not need the second equation. We put it in because "homomorphisms" of algebraic structures should preserve those structures, and the unity is a special element of a group. That it is preserved given the first equation is fortuitous but should not *a priori* be assumed.

If φ above is a (group) homomorphism and if, in addition, φ is:

(1) injective, we say φ is a *(group) monomorphism* or *monic*.

(2) surjective, we say φ is a *(group) epimorphism* or *epic*.

(3) bijective and φ^{-1} is a homomorphism, we say φ is a *(group) isomorphism*.

We let

$$\ker \varphi := \{a \in G \mid \varphi(a) = e_{G'}\}$$

called the *kernel* of φ and

$$\operatorname{im} \varphi := \{\varphi(a) \in G' \mid a \in G\}$$

called the *image* of φ.

If G and G' are groups and there exists an isomorphism $\varphi : G \to G'$, which we also write as $\varphi : G \xrightarrow{\sim} G'$, we say that G and G' are *isomorphic* and write $G \cong G'$.

Remark 9.5. Let $\varphi : G \to G'$ be a group homomorphism. We leave it as an exercise to show that φ is a monomorphism if and only if given any group homomorphisms $\psi_1, \psi_2 : H \to G$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$; and φ is an epimorphism if and only if given any group homomorphisms $\theta_1, \theta_2 : G' \to H$ with compositions satisfying $\theta_1 \circ \varphi = \theta_2 \circ \varphi$, then $\theta_1 = \theta_2$. (Cf. Exercise 1.13(7) and (8).)

Remark 9.6. You should know that a bijective linear transformation of vector spaces is an isomorphism, i.e., the inverse is automatically linear. The same is true for groups, i.e., a bijective homomorphism of groups is an isomorphism. We shall see this is also true in other algebraic cases. The proof is essentially the same as that for bijective linear transformations.

A group homomorphism satisfies the following:

Properties 9.7. Let $\varphi : G \to G'$ be a group homomorphism.

1. $\varphi(e_G) = e_{G'}$ (i.e., this is not needed in the definition of a subgroup).
2. $\varphi(a^{-1}) = \varphi(a)^{-1}$ for all $a \in G$.
3. $\ker \varphi \subset G$ and $\operatorname{im} \varphi \subset G'$ are subgroups.
4. φ is monic if and only if $\ker \varphi = \{e_G\}$.
5. φ is epic if and only if $\operatorname{im} \varphi = G'$.

PROOF. (1): $e_{G'}\varphi(e_G) = \varphi(e_G) = \varphi(e_G e_G) = \varphi(e_G)\varphi(e_G)$.
By Cancellation, $e_{G'} = \varphi(e_G)$.
(2): $e_{G'} = \varphi(e_G) = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1})$. Similarly, $e_{G'} = \varphi(a^{-1})\varphi(a)$.
(3), (5) are left as exercises.
(4): We have $\varphi(a) = \varphi(b)$ if and only if

$$e_{G'} = \varphi(a)\varphi(b)^{-1} = \varphi(a)\varphi(b^{-1}) = \varphi(ab^{-1})$$

if and only if $ab^{-1} \in \ker\varphi$. So $\ker\varphi = \{e_G\}$ if and only if $\varphi(a) = \varphi(b)$ implies $a = b$. □

Examples 9.8. 1. Let G and H be groups. Then the map $\varphi : G \to H$ given by $x \mapsto e_H$ is a group homomorphism, called the *trivial homomorphism*.

2. Let H be a subgroup of G. Then the inclusion map of H in G is a group homomorphism. This observation leads to a better way to define the notion of a subgroup, viz., let $H \subset G$ be a subset with both H and G groups. Then H is a subgroup of G if and only if the inclusion map is a monomorphism.
3. Let F be a field. Then the map

$$\det : \mathrm{GL}_n(F) \to F^\times \text{ given by } A \mapsto \det A$$

is an epimorphism with $\ker\det = \mathrm{SL}_n(F)$. (The same is true if F is just a commutative ring.)
4. Let $n \in \mathbb{Z}^+$. Then the map $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $x \mapsto \overline{x}$ is an epimorphism of additive groups with $\ker{}^- = n\mathbb{Z} := \{kn \mid k \in \mathbb{Z}\}$.
5. Let $n \in \mathbb{Z}^+$. Then the map $\varphi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}^\times$ given by $\overline{x} \mapsto e^{2\pi\sqrt{-1}x/n}$ is a well-defined (why?) map, and is a monomorphism with $\operatorname{im}\varphi = \mu_n$.
6. Let G be an abelian group. Then the map $\varphi : G \to G$ given by $x \mapsto x^{-1}$ is a group homomorphism. If G is not abelian, this map is never a group homomorphism.

Classifying objects up to isomorphism is usually impossible and difficult when possible. Indeed some of the most important theorems arise when this can be accomplished. However, in a few cases it is quite elementary. For example, two finite sets are isomorphic (i.e., bijective) if and only if they have the same number of elements, and two finite dimensional vector spaces over a field F are isomorphic if and only if they have the same dimension. Another is the case for cyclic groups that we now show.

Theorem 9.9. (Classification of Cyclic Groups) *Let $G = \langle a\rangle$ be a cyclic group and $\theta : \mathbb{Z} \to G$ the map given by $m \mapsto a^m$. Then θ is a group epimorphism. It is an isomorphism if and only if G is infinite. If G is finite, then $|G| = n$ if and only if $\ker\theta = n\mathbb{Z}$. If this is the case, then the map $\overline{\theta} : \mathbb{Z}/n\mathbb{Z} \to G$ given by $\overline{m} \mapsto a^m$ induced by θ is a group isomorphism.*

PROOF. As $\theta(i + j) = a^{i+j} = a^i a^j = \theta(i)\theta(j)$, the map is a group homomorphism. It is clearly an epimorphism.
We must show that θ is an isomorphism if and only if $|G|$ is infinite and if not then $\ker\theta \neq \{e\}$ to establish the first part of the theorem.

We have θ is one-to-one if and only if $a^i \neq a^j$ whenever $i \neq j$ if and only if $a^m \neq e_G$ for all $m \neq 0$ if and only if θ is an isomorphism (as it is onto). In particular, if θ is one-to-one, then G is infinite.

So we may assume that there exists $N \in \mathbb{Z}^+$ with $a^N = e_G$, hence θ is not one-to-one. We show that this means that G is finite. By the Well-Ordering Principle, there exists a least $n \in \mathbb{Z}^+$ such that $a^n = e_G$. We show

Claim 9.10. *$a^i = a^j$ if and only if $i \equiv j \mod n$. In particular, $a^m = e_G$ if and only if $n \mid m$.*

Note that the claim implies that $|G| = n < \infty$ and $\ker\theta = n\mathbb{Z}$. So suppose that the integers i and j are not equal. We may assume that $j > i$. Set $m = j - i \in \mathbb{Z}^+$ and write $m = kn + r$ with $0 \leq r < n$ and k integers using the Division Algorithm. We know that

$$\begin{aligned} a^i = a^j &\text{ if and only if } a^m = a^{j-i} = e_G \\ &\text{ if and only if } a^r = (a^n)^k a^r = a^m = e_G. \end{aligned}$$

By the minimality of n, this can occur if and only if $n \mid m$ if and only if $m \equiv 0 \mod n$ if and only if $i \equiv j \mod n$ as required.

We know by the First Isomorphism of Sets 7.1 that θ induces a bijection $\overline{\theta}$. [Make sure that you understand this.] It is clearly a homomorphism, so a group isomorphism. □

Subgroups of a cyclic group are classified by the following, whose proof we leave as an exercise [that you should do]:

Theorem 9.11. (Cyclic Subgroup Theorem) *Let $G = \langle a \rangle$ be a cyclic group, $H \subset G$ a subgroup. Then the following are true:*

(1) $H = \{e\}$ or $= \langle a^m \rangle$ with $m \geq 1$ the least positive integer such that $a^m \in H$. If $|G| = n\,(< \infty)$ then $m \mid n$. If G is infinite then $|H| = 1$ or H is infinite.
(2) If $|G| = n$ and $m \mid n$ then $\langle a^m \rangle$ is the unique subgroup of G of order $\frac{n}{|m|}$.
(3) If $|G| = n$ and $m \nmid n$ then G has no subgroup of order m.
(4) If $|G| = n$, the number of subgroups of G is equal to the number of positive divisors, $\sum_{d|n} 1$, of n (where in such a sum, we always assume $d > 0$).
(5) If $|G|$ is a prime then $\{e\}$ and G are the only subgroups of G.

Exercises 9.12.

1. Let p be a prime and $F = \mathbb{Z}/p\mathbb{Z}$. Then F is a field by Exercise 8.5 (13). Compute $|G|$ for $G = \mathrm{GL}_n(F)$, $\mathrm{SL}_n(F)$, $\mathrm{T}_n(F)$, $\mathrm{ST}_n(F)$, and $D_n(F)$.
2. Prove Dirichlet's Pigeonhole Principle.
3. Prove Remark 9.5.
4. Prove Property 9.7(3) and Remark 9.6, i.e., a bijective group homomorphism is a group isomorphism. In particular, if $\varphi : G \to G'$ is a monomorphism then G is isomorphic to $\varphi(G)$.
5. Let G_i, $i \in I$, be groups. Show that $\bigtimes_I G_i$ satisfies the following property relative to the group homomorphisms $\pi_j : \bigtimes_I G_i \to G_j$ defined by $\{g_i\}_I \mapsto g_j$ for all $j \in I$:

Given a group G and group homomorphisms $\phi_j : G \to G_j$, there exists a unique group homomorphism $\psi : G \to \bigtimes_I G_i$ satisfying $\phi_j = \pi_j \circ \psi$ for all $j \in I$.

6. Prove that every group of order four is either isomorphic to $\mathbb{Z}/4\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z} \times Z/2\mathbb{Z}$.
7. Prove the Cyclic Subgroup Theorem (Theorem 9.11).
8. Let G be an abelian group. Suppose that the elements a, b of G are of relatively prime orders m and n respectively. Show ab has order mn. Is this true if G is not abelian? Prove or give a counterexample.
9. Find a group isomorphic to $\mathrm{ST}_2(\mathbb{R})$ besides itself. Prove your answer.
10. Show that the quaternion group Q has a unique subgroup of order two, in particular, it is not isomorphic to D_4.
11. Let G be the subgroup of $\mathbb{M}_2(\mathbb{C})$ generated by
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$
Show G is isomorphic to the quaternion group Q.

10. Cosets

We now define an equivalence on a group G that mimicks congruence modulo m for integers. Although the similarity is precise, the reason for doing so is not so clear. Later we shall see how this equivalence arises in a natural way.

Let H be a subgroup of G. Define $\equiv \mod H$ by
$$\text{if } a, b \in G \text{ then } a \equiv b \mod H \text{ if and only if } b^{-1}a \in H.$$
Then $\equiv \mod H$ is an equivalence relation. [Show this.] If $a \in G$, the equivalence class $\overline{a}$ of a is called the *left coset* of a relative to H. Analogous to $\equiv \mod m$ for integers, we shall usually write aH for $\overline{a}$ — remember the binary operation for $\mathbb{Z}$ is $+$ — and shall write G/H for $\overline{G} = G/\equiv \mod H$. So we have the natural set surjection
$$\overline{} : G \to G/H \text{ given by } a \mapsto \overline{a} = aH.$$

Warning 10.1. In general, G/H is not a group, so the natural surjection above is not a group homomorphism. [Can you give an example?]

Observation 10.2. Let $H \subset G$ be a subgroup and $H = eH$ $(e = e_G = e_H)$ and $a \in G$. Then
$$aH = H \text{ if and only if } aH = H = eH \text{ if and only if } a = e^{-1}a \in H,$$
i.e., in our other notation
$$\overline{a} = \overline{e} \text{ if and only if } e \in \overline{a} \text{ if and only if } a \in \overline{e}.$$
We also have
$$\begin{aligned} aH &= \{b \in G \mid b \equiv a \mod H\} = \{b \in G \mid a^{-1}b \in H\} \\ &= \{b \in G \mid a^{-1}b = h \text{ some } h \in H\} \\ &= \{b \in G \mid b = ah \text{ some } h \in H\} = \{ah \mid h \in H\}. \end{aligned}$$
(Cf. $a + m\mathbb{Z}$ for the equivalence $\equiv \mod m$.)

We now use the Mantra of Equivalence Relations 5.13. Let H be a subgroup of G and $\mathcal{H}$ be a system of representatives for $\equiv \mod H$. We call $\mathcal{H}$ a *left transversal* of H in G. Then we have

Mantra 10.3. for Cosets. In the above setup, we have

$$G = \bigvee_{\mathcal{H}} aH.$$

In particular, if $|G| < \infty$, then

$$|G| = \sum_{\mathcal{H}} |aH|.$$

As mentioned before, this is only useful if we can compute the size of the equivalence class aH. We call $|\mathcal{H}|$ the *index* of H in G and denote it by $[G : H]$. The Mantra allows us to establish our first important counting result:

Theorem 10.4. (Lagrange's Theorem) *Let G be a finite group and H a subgroup of G. Then*

$$|G| = [G : H]|H|. \text{ In particular, } |H| \mid |G| \text{ and } [G : H] \mid |G|.$$

PROOF. We begin with the following:

Claim 10.5. *If G is an arbitrary group (i.e., without assuming it is finite) and H a subgroup, then $|aH| = |H|$ for all $a \in G$.*

Define the *left translation map*

$$\lambda_a : H \to aH \text{ by } h \mapsto ah.$$

By our computation above, we know that λ_a is onto. But it is one-to-one also, as $ah = ah'$ in G with $h, h' \in G$ implies that $h = h'$ by cancellation. Thus λ_a is a bijection, establishing the claim.

If $\mathcal{H}$ is a left transversal for H in G, the mantra together with the claim imply that

$$|G| = \sum_{\mathcal{H}} |aH| = \sum_{\mathcal{H}} |H| = |\mathcal{H}||H|$$

as needed. □

Notation 10.6. If G is a group and $H = \{e_G\}$, we shall write 1 for H unless G is additive in which case we shall write 0 for H. The subgroup $\{e_G\}$ is called the *trivial group* in G.

With this notation, $|G| = [G : 1]$, so if G is finite, Lagrange's Theorem can be written

$$[G : 1] = [G : H][H : 1].$$

We, of course, have an analogous Lagrange's Theorem for *right cosets*. [Definition?] It follows by both Lagrange's Theorems (i.e., for left and right cosets) that if G is a finite group and H a subgroup then $|G|/|H|$ is equal to the (left) index of H in G which is equal to the right index of H in G, i.e., the number of right cosets of H in G is the same as the number of left cosets of H in G. This means that $[G : H]$ makes sense without prescribing right or left cosets. It does not mean, however, that if $a \in G$, then $aH = Ha$. This is usually false. We shall find an important condition on H for this to be true for all a in G.

Warning 10.7. In general, the converse to Lagrange's Theorem is false, i.e., if G is a finite group with $m \in \mathbb{Z}^+$ satisfying $m \mid |G|$, then there may not be a subgroup H in G such that $|H| = m$, if $m \neq 1$ or $|G|$. Of course, if G is a finite cyclic group then the converse does hold by the Cyclic Subgroup Theorem 9.11.

Lagrange's Theorem has many immediate consequences.

Corollary 10.8. *Let G be a finite group and $a \in G$. Then the order of a divides $|G|$.*

Corollary 10.9. *Let G be a group and K and H two finite subgroups of G of relatively prime order. Then $K \cap H = \{e_G\}$.*

PROOF. We leave this as an exercise. □

Corollary 10.10. *Let G be a finite group of prime order p. Then $G \cong \mathbb{Z}/p\mathbb{Z}$. In particular, 1 and G are the only subgroups of G.*

Notation 10.11. If A is a subset of B but $A \neq B$, we shall write $A < B$.

The corollary says if G is a finite group of prime order, then G contains no *proper* subgroups, i.e., G contains no subgroup H satisfying $1 < H < G$.

Corollary 10.12. *Let G be a finite group and $a \in G$. Then $a^{|G|} = e$.*

PROOF. If n is the order of $\langle a \rangle$, then $|G| = nm$ for some $m \in \mathbb{Z}^+$ so $e = (a^n)^m = a^{|G|}$. □

Corollary 10.13. (Euler's Theorem) *Let m, n be relatively prime integers with $m > 1$. Then $n^{\varphi(m)} \equiv 1 \mod m$, where φ is the Euler φ-function.*

PROOF. We have shown that the unit group $(\mathbb{Z}/mZ)^\times$ is given by

$$(\mathbb{Z}/mZ)^\times = \{\overline{a} \mid a \in \mathbb{Z}, (a, m) = 1\}$$

and by definition, its cardinality is $\varphi(m)$. □

Corollary 10.14. (Fermat's Little Theorem) *Let p be a positive prime integer. Then $n^p \equiv n \mod p$ for all integers n. If $p \not| \, n$, then $n^{p-1} \equiv 1 \mod p$.*

PROOF. As $\varphi(p) = p - 1$, this follows from Euler's Theorem together with the observation that $0^p = 0$. □

We give a simple application of Euler's Theorem that shows the usefulness of elementary number theory in coding theory. All messages to be encoded are assumed to have been translated into (sequences of) integers. This application is called the *RSA code* that allows encoding and decoding using a public key, i.e., an integer m publicly known, together with a secret key, a number not publicly known. It and its variants are a major way of sending important data, e.g., credit card numbers, over the internet. To be effective, the secret key is based on the fact that the public key cannot be factored (in a reasonable amount of time). To set up the code, one chooses the public key m to be an integer that is the product of two large distinct primes p and q. We know that

$$\begin{aligned}\varphi(m) &= \varphi(p)\varphi(q) = (p-1)(q-1)\\ &= pq - (p+q) + 1 = m - (p+q) + 1.\end{aligned}$$

So to know $\varphi(m)$, we must know p and q, hence we must be able to factor m. A public encoder e is chosen and fixed. This is an integer satisfying $\big(e, \varphi(m)\big) = 1$. Then d is chosen $\big($using the knowledge of $m = pq$, hence $\varphi(m)\big)$ to be an integer satisfying $de \equiv 1 \mod \varphi(m)$, i.e., the inverse of e modulo $\varphi(m)$. Write $de = 1 + k\varphi(m)$ with $k \in \mathbb{Z}$. d is the (non-public) decoder. Messages are written in numbers $x < m$. So RSA code works as follows:

$$\begin{aligned}&\textbf{Encode}: \quad x \mapsto x^e \mod m \quad [e \text{ is publicly known}]\\ &\textbf{Decode}: \quad y \mapsto y^d \mod m \quad [d \text{ is secret}].\end{aligned}$$

Claim 10.15. $x^{ed} \equiv x \mod m$, i.e., for all positive integers $x < m$, we get x back.

If $(x, m) = 1$, then by Euler's Theorem, we have

$$x^{ed} = x \cdot x^{k\varphi(m)} \equiv x \mod m$$

as needed, so we may assume that $(x, m) \neq 1$. As x satisfies $1 < x < m$, either $p \mid x$ or $q \mid x$ but not both. We may assume that $p \mid x$. Therefore, $x = pn$ for some integer n and $(x, q) = 1$. Since $p \mid x$,

$$x^{ed} \equiv 0 \equiv x \mod p.$$

By Fermat's Theorem,

$$x^{\varphi(m)} = (x^{q-1})^{p-1} \equiv 1^{p-1} \equiv 1 \mod q.$$

Since p and q are relatively prime and both p and q divide $x^{ed} - x$, we have $m = pq \mid x^{ed} - x$, so $x^{ed} \equiv x \mod m$. This establishes the claim.

Some care must be taken in the choice of the large primes p and q. For example, if all primes less than N are known and either $p-1$ or $q-1$ factors into products of such primes, then m can be factored using what is called the Pollard $(p-1)$-Factoring Algorithm. Note that the idea of the RSA code is also based on group theory. Other public codes have been developed using group theory.

Exercises 10.16.

In the following exercises, let G be a group with H and K subgroups.

1. Show that $\equiv \mod H$ is an equivalence relation.
2. Let G be a group of order p^n where p is a prime and $n \geq 1$. Prove that there exists an element of order p in G.
3. Classify all groups of order eight up to isomorphism.
4. Let G be a group and $H \subset G$ a subgroup. Let $\mathcal{H}$ be a left transversal for H in G. Find a right transversal for H in G using $\mathcal{H}$.

5. Let $HK := \{hk \mid h \in H,\, k \in K\}$. Then (clearly) $H/(H\cap K)$ is a subset of $G/(H\cap K)$ and HK/K is a subset of G/K. Show that $f : H/(H\cap K) \to HK/K$ by $h(H\cap K) \mapsto hK$ is a well-defined bijection. In particular, if G is a finite group, then $|HK||H\cap K| = |H||K|$.
6. Suppose that G is a finite group. Show $[G : H\cap K] \leq [G : K][G : H]$ with equality if $[G : H]$ and $[G : K]$ relatively prime.
7. (Poincaré) Suppose both H and K have finite index in G. Show that $H\cap K$ has finite index in G.
8. If $K \subset H \subset G$, show that $[G : K] = [G : H][H : K]$ (even if any are infinite if read correctly).
9. If K and H are finite subgroups of G of relatively prime degree, show that $K\cap H = \{e_G\}$.
10. Let $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ have order m and $b \in (\mathbb{Z}/n\mathbb{Z})^\times$ have order n.
    (i) Show that the order of a^r is $r/(m,r)$.
    (ii) Show if $(m,n) = 1$, then the order of ab is mn.
11. Let p be an odd prime. Prove that $(\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$. [You may use the fact that the polynomial $t^n - 1$ has at most n roots in $(\mathbb{Z}/p\mathbb{Z})[t]$ for any $n > 0$.]
12. Let m and n be positive integers that are not relatively prime. Show that $\mathbb{Z}/m\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}$ is abelian but not cyclic.
13. Prove that a group of order 30 can have at most 7 subgroups of order 5.
14. (Wilson's Theorem) Let p be a positive integer. Then p is a prime if and only if $(p-1)! \equiv -1 \mod p$. (This is not a practical criterion.)
    [Hint: If $1 \leq j \leq p-1$, when can $j^2 \equiv 1 \mod p$?]

11. Homomorphisms

We begin with a useful fact whose proof is left as an exercise:

Observation 11.1. If $\varphi : G \to G'$ is a group homomorphism and H is a subgroup of G, then $\varphi(H) \subset G'$ is a subgroup.

If G is a group and $x \in G$, then the map

$$\theta_x : G \to G \ \text{ given by } \ g \mapsto xgx^{-1}$$

is called *conjugation* by x. If $H \subset G$ is a subgroup, then we let

$$xHx^{-1} := \theta_x|_H(H) = \theta_x(H) = \{xhx^{-1} \mid h \in H\}.$$

Lemma 11.2. *Let G be a group and $x \in G$. Then $\theta_x : G \to G$ is an isomorphism. In particular, if $H \subset G$ is a subgroup, so is $\theta_x(H)$ and $H \cong \theta_x(H) = xHx^{-1}$. In particular, $|H| = |xHx^{-1}|$. Since $\theta_x\theta_{x^{-1}} = 1_G = \theta_{x^{-1}}\theta_x$,*

Proof. The map θ_x is a bijection with inverse $\theta_{x^{-1}}$. Let $g, g' \in G$. Then the equation

$$\theta_x(gg') = xgg'x^{-1} = xgeg'x^{-1} = xgx^{-1}xg'x^{-1} = \theta_x(g)\theta_x(g') \tag{11.3}$$

shows that θ_x is a homomorphism. □

This idea of writing an identity in a clever way is a most useful device, which we shall call *Great Trick*. You have undoubtedly used it quite often for the elements 1 and 0 in $\mathbb{R}$. By Observation 11.1, we know that $\theta_x(H) \subset G$ is a subgroup, so we can view $\theta_x : H \to \theta_x(H)$. As it is a bijective homomorphism, it is an isomorphism.

A group isomorphism $\varphi : G \to G$ is called a *group automorphism*. So θ_x above is a group automorphism for all $x \in G$. If G is an abelian group, then θ_x is the identity on G for all $x \in G$. If G is not abelian, then there must exist elements x and y in G satisfying $xy \neq yx$, so $\theta_x \neq 1_G$. In general, if H is a subgroup, $xHx^{-1} \neq H$, i.e., the isomorphism $\theta_x|_H : H \to xHx^{-1}$ (replacing the target with $\theta_x(H)$) is not an automorphism. [Can you come up with examples when $\theta_x|_H$ is an automorphism?] Of course, if $x \in H$ then $xHx^{-1} = \theta_x(H) = H$.

We say that a subgroup H of a group G is *normal* and write $H \triangleleft G$ if $xHx^{-1} = \theta_x(H) = H$ for all $x \in G$.

Remarks 11.4. Let H be a subgroup of G.

1. $H \triangleleft G$ does not mean that $\theta_x|_H = 1_H : H \to H$ for all $x \in G$ (or even all $x \in H$). In fact, it usually is not. For example, (see below)
$$H := \langle r \rangle \triangleleft D_3 \text{ but } frf^{-1} = r^{-1}, \text{ so } \theta_f|_H \neq 1_H.$$
2. Let
$$\mathrm{Aut}(G) := \{\sigma : G \to G \mid \sigma \text{ is an automorphism}\}.$$
Then $\mathrm{Aut}(G)$ is a group under composition. [Check this.] In fact, $\mathrm{Aut}(G)$ is a subgroup of $\Sigma(G)$. A conjugation $\theta_x : G \to G$, with $x \in G$, is also called an *inner automorphism* of G. Set
$$\mathrm{Inn}(G) := \{\theta_x \mid x \in G\}.$$
Check that this is a subgroup of $\mathrm{Aut}(G)$, called the *inner automorphism group* of G. Moreover, the following are equivalent:
   (a) $H \triangleleft G$.
   (b) $\theta(H) = H$ for all $\theta \in \mathrm{Inn}(G)$.
   (c) $\theta|_H \in \mathrm{Aut}(H)$ for all $\theta \in \mathrm{Inn}(G)$.
   (d) The restriction map $1_{\mathrm{Aut}(G)}|_{\mathrm{Inn}(G)} : \mathrm{Inn}(G) \to \mathrm{Aut}(H)$ is defined (i.e., the image lies in $\mathrm{Aut}(H)$).
3. $H \triangleleft G$ if and only if $xH = Hx$ for all $x \in G$. (Why?)
4. If $xHx^{-1} \subset H$ for all $x \in G$, then $y^{-1}Hy \subset H$ for all $y \in G$ (let $y = x^{-1}$), so $H \subset yHy^{-1}$ for all $y \in G$. Hence
$$H \triangleleft G \text{ if and only if } xHx^{-1} \subset H \text{ for all } x \in G.$$

Examples 11.5. Let G be a group.

1. $1 \triangleleft G$ and $G \triangleleft G$. If G is a nontrivial group and these are the only normal subgroups of G, then we say G is a *simple group*. We have seen that $\mathbb{Z}/p\mathbb{Z}$ is a simple group for all primes p. One of the great theorems in mathematics is the classification of all finite simple groups. The proof is thousands of pages long.

2. If G is abelian, every subgroup is normal. The converse is false, i.e, there exist non-abelian groups in which every subgroup is normal.
3. Let
$$Z(G) := \{x \in G \mid xg = gx \text{ for all } g \in G\}.$$
This set is called the *center of* G. It is a normal subgroup of G. In fact, any subgroup of $Z(G)$ is a normal subgroup of G. Moreover, G is abelian if and only if $G = Z(G)$.
4. If F is a field (even a commutative ring), then $\mathrm{SL}_n(F) \triangleleft \mathrm{GL}_n(F)$.
5. Let $\theta : G \to G'$ be a group homomorphism. Then $\ker\theta \triangleleft G$.
[In general, $\operatorname{im}\theta$ is not a normal subgroup of G. [Can you give an example?]
6. If G is an abelian group, then $\varphi : G \to G$ given by $x \mapsto x^{-1}$ is an automorphism. It is the identity map if and only if $x^2 = e$ for all $x \in G$, otherwise it is of order two. In particular, if G is a finite abelian group of order at least three, then $2 \mid |\operatorname{Aut}(G)|$. [Can you prove this?]
7. $\operatorname{Inn}(G) \triangleleft \operatorname{Aut}(G)$.
8. If H is a subgroup of G of index two then $H \triangleleft G$. Indeed, if $a \in G \setminus H$ then $G = H \vee aH = H \vee Ha$, so we must have $aH = Ha$, hence $aHa^{-1} = H$.

   For the next example, we shall need matrix representations of linear transformations of finite dimensional vector spaces. In Appendix C, we review this notation. (This appendix does a more general case but is applicable here.) In particular, if $T : V \to W$ is a linear transformation of vector spaces over a field F with bases $\mathcal{B}$ and $\mathcal{C}$ respectively, then we let $[T]_{\mathcal{B},\mathcal{C}}$ denote the matrix representation of this linear transformation relative to these bases. If $V = V'$ and $\mathcal{B} = \mathcal{C}$, we write this as $[T]_{\mathcal{B}}$.
9. If $\sigma \in S_n$, we can write σ as
$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$
where the top row is the elements of the domain and the bottom row the corresponding values. Let $\mathcal{S}_n := \{e_1, \dots, e_n\}$ be the *standard basis* for $V = \mathbb{R}^n$. For each σ, define a linear transformation $T_\sigma : V \to V$ by $\Sigma_i \alpha_i e_i \mapsto \Sigma_i \alpha_i e_{\sigma(i)}$. Then T_σ is a vector space isomorphism with inverse $T_{\sigma^{-1}}$. Define θ by
$$\theta : S_n \to \mathrm{GL}_n(\mathbb{R}) \text{ is given by } \sigma \mapsto [T_\sigma]_{\mathcal{S}_n}. \tag{11.6}$$
It is easy to see that θ is a group monomorphism. Each $[T_\sigma]_{\mathcal{S}_n}$ is a *permutation matrix*, i.e., a matrix having exactly one entry of 1 in each row and each column and 0s elsewhere. So it is just a permutation of the rows (or columns) of the identity matrix. Let $\mathrm{Perm}_n(\mathbb{R})$ be the set of permutation matrices. It is the image of θ, so a group. Changing the target, we view $\theta : S_n \to \mathrm{Perm}_n(\mathbb{R})$, an isomorphism. If $A \in \mathrm{Perm}_n(\mathbb{R})$, then $\det A = \pm 1$, so we have a composition of maps
$$S_n \xrightarrow{\theta} \mathrm{Perm}_n(\mathbb{R}) \xrightarrow{\det} \{\pm 1\}.$$
It is a homomorphism as both θ and det are. If $n > 1$ then the kernel of this map is called the *alternating group* on *n letters* and denoted by A_n. Elements of A_n are

called *even permutations* and elements of $S_n \setminus A_n$ are called *odd permutations*. We have $A_n \lhd S_n$, and, in fact, $[S_n : A_n] = 2$ if $n \geq 2$. We shall study this group in Section 24.

Remarks 11.7. 1. We have seen that a group of prime order is a simple group, i.e., a group without any proper normal subgroups. These turn out to be the only abelian simple groups. The group A_5 is the non-abelian simple group of smallest order. This was shown by Abel. In fact,

Theorem 11.8. (Abel's Theorem) *For every $n \geq 5$, the group A_n is simple.*

The group A_2 is the trivial group and A_3 is the cyclic group of order three so simple. The group A_4 is not simple.

Abel's Theorem is important as it is the key to proving the Abel-Ruffini Theorem, which states that there is no formula for the roots of a general fifth degree polynomial with rational coefficients involving only the extraction of nth roots and $+$ and $\cdot$. We shall prove Abel's Theorem in Theorem 24.13.

2. If K and H are subgroups of G satisfying $K \subset H \subset G$ with $K \lhd H$ and $H \lhd G$, it is not necessarily true that $K \lhd G$, i.e., being normal is not transitive.
3. If K and H are subgroups of G satisfying $K \subset H \subset G$ with $K \lhd G$, then it is true that $K \lhd H$.
4. There is a useful stronger property that a subgroup of a group can satisfy than being normal. We say that a subgroup H of a group G is a *characteristic subgroup* of G if for every $\sigma \in \mathrm{Aut}(G)$, we have $\sigma|_H$ lies in $Aut(H)$. Equivalently, the restriction map $\mathrm{res} : \mathrm{Aut}(G) \to \mathrm{Aut}(H)$ given by $\theta \mapsto \theta|_H$ is well-defined, i.e., the target is correct. If H is a characteristic subgroup of G, we write $H \lhd\lhd G$. Clearly, unlike being normal, being characteristic is transitive, i.e., if $K \lhd\lhd H \lhd\lhd G$, then $K \lhd\lhd G$.

Exercises 11.9.

1. Prove Observation 11.1.
2. Find all subgroups of S_3 and determine which ones are normal.
3. Show that a subgroup $H \subset G$ is normal if and only if $gH = Hg$ for all $g \in G$. If H is not normal is it still true that for each $g \in G$ there is an $a \in G$ such that $gH = Ha$?
4. Determine all subgroups of the quaternion group of order 8 and determine which ones are normal.
5. Classify all groups of order eight up to isomorphism.
6. Let G be a group. Show
   - (i) The center $Z(G)$ is a subgroup G.
   - (ii) Any subgroup of the center of G is a normal subgroup of G.
   - (iii) G is abelian if and only if it equals its center.
7. Determine the center of the dihedral group D_n and prove your assertion.
8. Give an example to show that being a normal subgroup is not transitive, i.e., a group G with $K \lhd H \lhd G$, but $K \ntriangleleft G$.
9. Let H be a subgroup of G. In Exercises 8.5(7) and (8), we defined the core of H in G to be $\mathrm{Core}_G(H) := \bigcap xHx^{-1}$ and the *normalizer* of H in G to be $N_G(H) :\ \{x \in G \mid xHx^{-1} = H\}$, respectively. Show

(*i*) $\mathrm{Core}_G(H)$ is the unique largest normal subgroup of G contained in H.
(*ii*) $N_G(H)$ is the unique largest subgroup of G containing H as a normal subgroup.

10. Show $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$.
11. Prove the assertions in Example 11.5(6).
12. Let G be a group (not necessarily finite) such that $\mathrm{Aut}(G) = 1$. Prove that $|G| \leq 2$.
13. Let G be a cyclic group. Determine $\mathrm{Aut}(G)$ and $\mathrm{Inn}(G)$ up to isomorphism as groups that we know. Prove your result.
[Hint. Where do generators go?]
14. Let G and H be finite cyclic groups of order m and n, respectively. Show the following:
(*i*) If m and n are relatively prime, then $\mathrm{Aut}(G \times H) \cong \mathrm{Aut}(G) \times \mathrm{Aut}\, H$ and is abelian.
(*ii*) If m and n are not relatively prime, then $\mathrm{Aut}(G \times H)$ is never abelian. (Cf. Exercise 11.9(15).)
15. Compute $\mathrm{Aut}\big((\mathbf{Z}/p\mathbf{Z})^n\big)$ up to isomorphism and its order when p is a prime.
16. Let $\varphi : G \to \mathrm{Aut}(H)$, write $x \mapsto \varphi_x$, be a group homomorphism where G and H are groups. Define $H \rtimes_\varphi G$ to be the cartesian product $H \times G$ with group structure induced by the binary operation
$$(h, g) \cdot (h', g') = (h\varphi_g(h'), gg')$$
for all $g, g' \in G$ and $h, h' \in H$. Show that $H \rtimes_\varphi G$ is a group, called the *(exernal) semidirect product* of H and G, with $H \times 1$ a normal subgroup. Note that the product $G \times H$ is the case that φ is the trivial homomorphism.
17. Let $\sigma \in \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ be the automorphism defined by $x \mapsto -x$ and $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ be the group homomorphism defined by $\overline{1} \mapsto \sigma$. Show that $\mathbb{Z}/n\mathbb{Z} \rtimes_\varphi \mathbb{Z}/2\mathbb{Z} \cong D_n$.
18. Let $p > 1$ be a prime and $n > 1$. Show that $\mathbb{Z}/p^n\mathbb{Z}$ is not isomorphic to a semidirect product.
19. Find all subgroups of A_4 and determine which ones are normal. In particular, show that A_4 is not simple.
20. Let G be a group in which every subgroup is normal. Is G an abelian group? Prove or provide a counterexample.
21. A *commutator* in G is an element of the form $xyx^{-1}y^{-1}$ where $x, y \in G$. Let G' be the subgroup of G generated by all commutators, i.e., every element of G' is the product of commutators and the inverses of commutators. We call G' the *commutator* or *derived subgroup* of G. It is also denoted $[G, G]$. Show
(*i*) Every element of G' is a product of commutators.
(*ii*) $G' \lhd G$.
(*iii*) $G' \lhd\lhd G$.
22. Let $K \subset H \subset G$ be subgroups of G. Show all of the following:
(*i*) If $K \lhd\lhd H$ and $H \lhd G$ then $K \lhd G$.
(*ii*) $Z(G) \lhd\lhd G$.
(*iii*) $G' \lhd\lhd G$.

(*iv*) Inductively define $G^{(n)}$ as follows: $G^{(1)} = G'$. Having defined $G^{(n)}$ define $G^{(n+1)} := (G^{(n)})'$. Then $G^{(n+1)} \lhd\lhd G$.

12. The First Isomorphism Theorem

Given an equivalence relation, we saw that we could always form a corresponding set of equivalence classes. If G is a group and H is a subgroup, we can form the set of equivalence classes G/H. Unfortunately, in general this does not have a group structure. We also know that equivalence relations arise from surjective maps. We begin by looking at the case when $\varphi : G \to G'$ is a group epimorphism. This defines an equivalence relation $\sim$ on G by $a \sim a'$ if $\varphi(a) = \varphi(a')$ and with our usual notation, we have

$$\overline{a} = \overline{a'} \text{ if and only if } a \sim a' \text{ if and only if } \varphi(a) = \varphi(a')$$

with

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{x \in G \mid \varphi(x) = \varphi(a)\}.$$

But now, we have the additional information that φ is a homomorphism, so $\varphi(aa') = \varphi(a)\varphi(a')$. This means that

$$\begin{gathered}\varphi(a) = \varphi(a') \text{ if and only if} \\ e_{G'} = \varphi(a)^{-1}\varphi(a') = \varphi(a^{-1}a') \text{ and } e_{G'} = \varphi(a')\varphi(a)^{-1} = \varphi(a'a^{-1}) \\ \text{if and only if both } a^{-1}a' \text{ and } a'a^{-1} \text{ lie in } \ker\varphi.\end{gathered}$$

Let $K = \ker\varphi$. So we have

$$\overline{a} = \overline{a'} \text{ if and only if } \varphi(a) = \varphi(a') \text{ if and only if } a^{-1}a', a'a^{-1} \in K.$$

We know that if $a^{-1}a' \in K$, then $\overline{a'} = aK$, so

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{a' \mid a' = ak \text{ some } k \in K\} = aK,$$

the left coset of a in G relative to K. Hence $\overline{G} = G/K$. Similarly,

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{a' \mid a' = ka \text{ some } k \in K\} = Ka,$$

the right coset of a in G relative to K. So we now have

$$\begin{aligned}\overline{a} = \overline{a'} &\text{ if and only if } \varphi(a) = \varphi(a') \\ &\text{ if and only if } aK = a'K = \overline{a} = Ka = Ka'.\end{aligned}$$

In particular, we must have $aK = Ka$, or equivalently, $K = aKa^{-1}$ for all $a \in G$ as $a \sim a$, i.e., $K \lhd G$. Of course, we knew this before, but it now indicates where cosets are coming from, at least in a special case, and ties up various mathematical threads that we are developing. Note that $\overline{e_G} = \varphi^{-1}\big(\varphi(e_G)\big) = \varphi^{-1}(e_{G'}) = \ker\varphi$.

We now apply the First Isomorphism of Sets 7.1 to obtain a commutative diagram

(†)
$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[d, "-"'] & G' \\ \overline{G} \arrow[ur, "\overline{\varphi}"'] & \end{tikzcd}$$

where $\overline{\varphi} : \overline{G} \to G'$ is a well-defined bijection given by $\overline{a} \mapsto \varphi(a)$. We also know that

$$\overline{a} = \overline{a'} \text{ if and only if } a \sim a' \text{ if and only if } \overline{\varphi}(\overline{a}) = \overline{\varphi}(\overline{a'}),$$

so the fact that φ is a homomorphism implies that

$$(*) \qquad \overline{\varphi}(\overline{aa'}) = \varphi(aa') = \varphi(a)\varphi(a') = \overline{\varphi}(\overline{a})\overline{\varphi}(\overline{a'}).$$

We can use this and the fact that $\overline{\varphi}$ is a bijection to put a group structure on $\overline{G}$. Define

$$\cdot : \overline{G} \times \overline{G} \to \overline{G} \text{ by } \overline{a} \cdot \overline{a'} := \overline{aa'}.$$

It is well-defined, for if $\overline{a} = \overline{x}$ and $\overline{x'} = \overline{a'}$, then

$$\overline{\varphi}(\overline{xx'}) = \overline{\varphi}(\overline{x}) \cdot \overline{\varphi}(\overline{x'}) = \overline{\varphi}(\overline{a}) \cdot \overline{\varphi}(\overline{a'}) = \overline{\varphi}(\overline{aa'})$$

by (*), so $\overline{xx'} = \overline{aa'}$ as $\overline{\varphi}$ is a bijection. For all $a \in G$, we have $\overline{e_G} \cdot \overline{a} = \overline{a} = \overline{a} \cdot \overline{e_G}$ and $\overline{a} \cdot \overline{a^{-1}} = \overline{e_G} = \overline{a^{-1}} \cdot \overline{a}$, so $\overline{G}$ is a group with identity $\ker \varphi = \overline{e_G}$ and inverse of $\overline{a}$ given by $\overline{a}^{-1} = \overline{a^{-1}}$. Furthermore, (*) implies that

$$\overline{\varphi}(\overline{a} \cdot \overline{a'}) = \overline{\varphi}(\overline{aa'}) = \overline{\varphi}(\overline{a}) \cdot \overline{\varphi}(\overline{a'}),$$

so $\overline{\varphi}$ is a bijective group homomorphism, hence an isomorphism. Our diagram (†) becomes

$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[d, "-"'] & G' \\ G/\ker\varphi \arrow[ur, "\overline{\varphi}"'] & \end{tikzcd}$$

with $G/\ker\varphi$ a group, $^-$ a group epimorphism and $\overline{\varphi}$ an isomorphism.

We can generalize this to any group homomorphism $\psi : G \to G''$, by letting $G' = \operatorname{im}\psi$ and φ the map $\varphi : G \to G'$ given by $\varphi(a) = \psi(a)$. As the inclusion map $\mathrm{inc} : G' \to G''$ is a group homomorphism, we are in a similar situation as in Diagram 7.3, but with all maps homomorphisms. We write this, after changing notation, as:

Theorem 12.1. (First Isomorphism Theorem) *Let $\varphi : G \to G'$ be a group homomorphism. Then we have a commutative diagram of groups and group homomorphisms*

$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[d, "-"'] & G' \\ G/\ker\varphi \arrow[r, "\overline{\varphi}"'] & \operatorname{im}\varphi \arrow[u, "\mathrm{inc}"'] \end{tikzcd}$$

with $^-$ a group epimorphism, $\overline{\varphi}$ a group isomorphism between $G/\ker\varphi$ and $\operatorname{im}\varphi$ and inc *a group monomorphism.*

Example 12.2. Let G be a group. The map $\theta : G \to \mathrm{Aut}(G)$ given by $x \mapsto (\theta_x : g \mapsto xgx^{-1})$ is a group homomorphism. If $\theta_x = 1_G$, the identity map on G, then $xgx^{-1} = g$, i.e., $xg = gx$, for all $g \in G$. Therefore, $\ker\theta = Z(G)$, the center of G. By the First Isomorphism Theorem, θ induces an isomorphism $\overline{\theta} : G/Z(G) \to \mathrm{Inn}(G)$.

If $\varphi : G \to G'$ is a group homomorphism, the First Isomorphism Theorem says that G/H is a group with $H = \ker\varphi$. It remains to ask: If H is a subgroup of G, when is G/H a group such that the canonical surjection $^- : G \to G/H$ a group epimorphism? This is answered by the following:

Theorem 12.3. *Let G be a group and H a normal subgroup of G. Then G/H is a group under the binary operation $\cdot : G/H \times G/H \to G/H$ given by $(aH, bH) \mapsto abH$. If this is the case then $^- : G \to G/H$ is a group epimorphism with kernel H.*

We call G/H the *quotient* or *factor group* of G by (the normal subgroup) H.

PROOF. We first show that the map $\cdot$ is well-defined. Suppose that $aH = a'H$ and $bH = b'H$. We must show that $abH = a'b'H$. Equivalently, we must show that

$$\text{if } a'^{-1}a, b'^{-1}b \in H \text{ then } x = (a'b')^{-1}ab \in H.$$

By using Great Trick and the definition of normality, we have

$$x = b'^{-1}a'^{-1}ab = b'^{-1}a'^{-1}a(b'b'^{-1})b = (b'^{-1}a'^{-1}ab')(b'^{-1}b)$$

lies in H as needed.

Let $a, b, c \in G$. Associativity holds because $aH \cdot (bH \cdot cH) = aH \cdot bcH = abcH = abH \cdot cH = (aH \cdot bH) \cdot cH$. The unity is easily seen to be $e_{G/H} = H = e_G H$ and $(aH)^{-1} = a^{-1}H$ for all a in G. Thus G/H is a group. $\square$

The theorem shows that normal subgroups are equivalent to the kernels of group homomorphisms. Of course, it does not explain how cosets arise in a natural way if the subgroup is not a normal subgroup. We will come back to this later.

We are now in a position where we can prove some important results. Especially significant is the following:

Theorem 12.4. (General Cayley Theorem) *Let H be an arbitrary subgroup of a group G. Let $S = G/H$, the set of left cosets of H in G. For each $x \in G$, let $\lambda_x : S \to S$ be defined by $gH \mapsto xgH$. Then λ_x is a permutation and the map $\lambda : G \to \Sigma(S)$ given by $x \mapsto \lambda_x$ is a group homomorphism satisfying the following two properties:*

(1) $\ker\lambda \subset H$.

(2) $\ker\lambda$ *is the unique largest normal subgroup of G contained in H, i.e.,*

$$\textit{If } N \lhd G \textit{ and } N \subset H \textit{ then } N \subset \ker\lambda.$$

PROOF. We first must show that the map $\lambda_x : S \to S$, called *left multiplication* by x, is a permutation. But this is clear since $\lambda_{x^{-1}}$ is easily checked to be its inverse.

For all $x, y \in G$, we have $\lambda_{xy}(gH) = xygH = \lambda_x(ygH) = \lambda_x\lambda_y(gH)$ for all $g \in G$, so λ is a group homomorphism and $\ker\lambda \lhd G$. If $x \in \ker\lambda$, then $xgH = \lambda_x(gH) = gH$ for all $g \in G$. In particular, $xH = H$ so $x \in H$. Finally, suppose that $N \lhd G$ with $N \subset H$. To show $N \subset \ker\lambda$, we must show if $x \in N \subset H$, then λ_x is 1_S, the identity map on S. So let $x \in N$ and $g \in G$. As $g^{-1}xg \in N$, we have $xg \in gN \subset gH$. Therefore, $xgH = gH$, i.e., $\lambda_x(gH) = gH$ for all $g \in G$. Consequently, $\lambda_x = 1_S$ as required. $\square$

In the above, the special case where $H = 1$ yields:

Corollary 12.5. (Cayley's Theorem) *Let G be a group. Then the map*

$$\lambda : G \to \Sigma(G) \text{ given by } x \mapsto (\lambda_x : g \mapsto xg)$$

is a group monomorphism. In particular, if G is a finite group of order n then there exists a monomorphism $G \to S_n$.

Remark 12.6. Since S_m is isomorphic to a subgroup of S_n if $m \leq n$, Cayley's Theorem shows that S_n contains an isomorphic copy of every group of order at most n.

If S is a nonempty set, then a subgroup of $\Sigma(S)$ is called a *permuation group*. Cayley's Theorem shows that every abstract group is isomorphic to a permutation group. The above monomorphism is called the (*left*) *regular representation* of G. It shows that every abstract group can be realized by a "concrete" group, e.g., a permutation group.

Corollary 12.7. *Let H be a subgroup of a group G with $H < G$. If there exists no normal subgroup N of G satisfying $1 < N \subset H$ then $\lambda : G \to \Sigma(G/H)$ defined by $x \mapsto (\lambda_x : aH \mapsto xaH)$ is a monomorphism.*

PROOF. $\ker \lambda$ is the maximal such normal subgroup. □

Corollary 12.8. (Useful Counting Result) *Let G be a finite group, $1 < H$ a subgroup of G satisfying $|G| \not| \, [G : H]!$. Then there exists a normal subgroup N of G satisfying $1 < N \subset H$. In particular, G is not a simple group.*

PROOF. Exercise. □

We now give two applications of the Useful Counting Result. To do so we must assume the following is true (we shall prove it later).

Theorem 12.9. (First Sylow Theorem) *Let G be a finite group, p a prime such that $p^s \mid\mid |G|$ (i.e., $p^s \mid |G|$ but $p^{s+1} \not| \, |G|$). Then G contains a subgroup of order p^s.*

Examples 12.10. Let G be a group.

1. If $|G| = 24$ then G is not simple, as G contains a group of order 8 by the First Sylow Theorem and $|G| \not| \, 3!$.
2. Let $p < q$ be positive primes such that $|G| = pq$. Then G contains a normal group of order q, so is not simple.

Another useful corollary of the General Cayley Theorem is the following, whose proof we also leave as an exercise:

Corollary 12.11. *Let G be a finite group. Suppose that p is the smallest positive prime that divides the order of G. Suppose also that there exists a subgroup H of G of index p. Then $H \triangleleft G$.*

Exercises 12.12.

1. Let $T : V \to W$ be a surjective linear transformation of vector spaces over F. Do an analogous analysis before The First Isomorphism Theorem 12.1 for $T : V \to W$ and state and prove the analogues of The First Isomorphism Theorem 12.1 and Theorem 12.3 for it.

2. Let G be a finite group and $n > 1$ an integer such that $(xy)^n = x^n y^n$ for all $x, y \in G$. Let
$$G_n := \{z \in G \mid z^n = e\} \quad \text{and} \quad G^n := \{x^n \mid x \in G\}.$$
Show that both G_n and G^n are normal subgroups of G and satisfy $|\,G^n| = [G : G_n]$.
3. Let G be a group and H a subgroup of G. Show that $\bigcap_{x \in G} xHx^{-1}$ is the largest normal subgroup of G in H.
4. Let G be a finite group of order n, F a field (or even any ring with $1 \neq 0$). Show there exists a monomorphism $\varphi : G \to \mathrm{GL}_n(F)$, i.e., any finite group can be realized as a subgroup of matrices called a *linear group*.
5. Prove the Useful Counting Result 12.8.
6. Prove Corollary 12.11.
7. Suppose that G is a group containing a subgroup of finite index greater than one. Show that G contains a normal subgroup of finite index greater than one. In particular, show that no infinite simple group can contain a proper subgroup of finite index.
8. Let H be a subgroup of G. Define the *centralizer* of H in G to be
$$Z_G(H) := \{x \in G \mid xh = hx \ \text{ for all } h \in H\}.$$
Show that it is a normal subgroup of $N_G(H)$ and the map given by $x \mapsto (\theta_x : g \mapsto xgx^{-1})$ induces $\tilde{\theta} : N_G(H)/Z_G(H) \to \mathrm{Aut}(H)$, a monomorphism defined by $xZ_G(H) \to \theta_x|_H$.

13. The Correspondence Principle

Recall if $f : A \to B$ is a set map and D a subset of B, then the *preimage* of D in A is the set $f^{-1}(D) := \{a \in A \mid f(a) \in D\}$. We shall need the following, whose proof is left as an exercise.

Properties 13.1. Let $f : A \to B$ be a set map, $C \subset A$, and $D \subset B$ subsets. Then:

1. $C \subset f^{-1}\big(f(C)\big)$ with equality if f is one-to-one.
2. $f\big(f^{-1}(D)\big) \subset D$ with equality if f is onto.

We apply the First Isomorphism Theorem to establish the following important result:

Theorem 13.2. (Correspondence Principle) *Let $\varphi : G \to G'$ be a group epimorphism. Then*

(1) If A is a subgroup of G (respectively, a normal subgroup), then $\varphi(A)$ is a subgroup of G' (respectively, a normal subgroup). In particular, group epimorphisms preserve normality.
(2) If A is a subgroup of G containing $\ker\varphi$, then $\varphi^{-1}\big(\varphi(A)\big) = A$.
(3) If B is a subgroup of G' (respectively, a normal subgroup), then $\varphi^{-1}(B)$ is a subgroup of G (respectively, a normal subgroup) containing $\ker\varphi$ and satisfies $B = \varphi\big(\varphi^{-1}(B)\big)$.

In particular, if

$$\mathcal{G}_{\ker\varphi} := \{A \mid A \subset G \ \textit{ a subgroup with } \ker\varphi \subset A\}$$
$$\mathcal{G}' := \{B \mid B \subset G' \ \textit{ a subgroup}\},$$

then

$$\mathcal{G}_{\ker\varphi} \to \mathcal{G}' \ \textit{ given by } \ A \mapsto \varphi(A)$$

is a bijection of sets with inverse $B \mapsto \varphi^{-1}(B)$ preserving inclusions and restricting to a bijection on normal subgroups.

PROOF. Let $K = \ker\varphi$.

(1): Let $A \subset G$ be a subgroup. Then $\varphi|_A : A \to G'$ is a group homomorphism (why?), so $\varphi(A) = \operatorname{im}\varphi|_A \subset G'$ is a subgroup. Next suppose that $A \triangleleft G$ and $y \in G'$. As φ is surjective, $y = \varphi(x)$ for some $x \in G$. Hence $y\varphi(A)y^{-1} = \varphi(x)\varphi(A)\varphi(x)^{-1} = \varphi(xAx^{-1}) = \varphi(A)$. Therefore, $\varphi(A) \triangleleft G'$.

(2): By the properties of preimages, $A \subset \varphi^{-1}\big(\varphi(A)\big)$, so it suffices to show if $K \subset A$, then $\varphi^{-1}\big(\varphi(A)\big) \subset A$. Suppose that $x \in \varphi^{-1}\big(\varphi(A)\big)$. By definition, $\varphi(x) = \varphi(a)$ for some $a \in A$, hence $e_{G'} = \varphi(a)^{-1}\varphi(x) = \varphi(a^{-1}x)$. Thus $a^{-1}x \in K$, and so $x \in aK \subset A$ as needed.

(3): Let B be a subgroup of G'. If $x, y \in \varphi^{-1}(B)$, we have $\varphi(xy^{-1}) = \varphi(x)\varphi(y)^{-1} \in B$, so $\varphi^{-1}(B)$ is a subgroup of G. If $k \in K$ then $\varphi(k) = e_{G'} \in B$, so $K \subset \varphi^{-1}(B)$. As φ is onto, $B = \varphi\big(\varphi^{-1}(B)\big)$ by the properties of preimages. Finally, suppose that $B \triangleleft G'$. Let $x \in A$. Then $\varphi\big(x\varphi^{-1}(B)x^{-1}\big) = \varphi(x)\varphi\big(\varphi^{-1}(B)\big)\varphi(x)^{-1} = \varphi(x)B\varphi(x)^{-1} = B$. Hence $x\varphi^{-1}(B)x^{-1} \subset \varphi^{-1}(B)$, so $\varphi^{-1}(B) \triangleleft G$. $\square$

We leave the following alternate form of the Correspondence Principle as an exercise. (Note we did not include the analogue of the last statement in the above.)

Theorem 13.3. (Correspondence Principle, Alternate Form) *Let G be a group and K a normal subgroup. Let $^{-} : G \to G/K$ be the canonical epimorphism given by $x \mapsto xK$ and L a subgroup of G/K. Then*

(1) There exists a subgroup H of G containing K with $L = H/K$.
(2) Let H be as in (1)*. Then $H \triangleleft G$ if and only if $L \triangleleft G/K$.*
(3) Suppose that H_1, H_2 are two subgroups of G containing K. If $H_1/K = H_2/K$ then $H_1 = H_2$.
(4) If G is a finite group and H is as in (1)*, then $[G : H] = [G/K : H/K] = [G/K : L]$ and $|H| = |K| \cdot |L|$.*

Theorem 13.4. (Third Isomorphism Theorem) *Let G be a group with normal subgroups K and H satisfying $K \subset H$. Then the map*

$$\varphi : G/K \to G/H \ \textit{ defined by } \ xK \to xH$$

is a group epimorphism with kernel H/K and induces an isomorphism

$$\overline{\varphi} : (G/K)/(H/K) \to G/H.$$

Proof. As K and H are normal subgroups of G, we know that G/H and G/K are groups. If $xK = yK$ then $y^{-1}x \in K \subset H$, hence $xH = yH$. Therefore, φ is well-defined and clearly surjective. As $\varphi(xKyK) = \varphi(xyK) = xyH = xHyH = \varphi(xK)\varphi(yK)$, the map φ is a group homomorphism. If $xK \in \ker\varphi$ then $H = e_{G/H} = \varphi(xK) = xH$, so $x \in H$, i.e., $xK \in H/K$. Conversely, if $x \in H$ then $\varphi(xK) = xH = H = e_{G/H}$, so $xK \in \ker\varphi$. The result now follows by the First Isomorphism Theorem. $\square$

Let G be a group and H_1, H_2 subgroups of G. Set

$$\begin{aligned} H_1H_2 :&= \{h_1h_2 \mid h_i \in H_i,\ i = 1, 2\} \\ &= \{h_1H_2 \mid h_1 \in H_1\} = \{H_1h_2 \mid h_2 \in H_2\}. \end{aligned}$$

By Exercise 10.16(5), we have a set bijection

$$(*) \qquad f : H_1/(H_1 \cap H_2) \to H_1H_2/H_2 \ \text{ given by } \ h_1/(H_1 \cap H_2) \mapsto h_1H_2$$

for all $h_1 \in H_1$ where $H_1/(H_1 \cap H_2)$ is a subset of $G/(H_1 \cap H_2)$ and H_1H_2/H_2 is a subset of G/H_2. In particular, we have

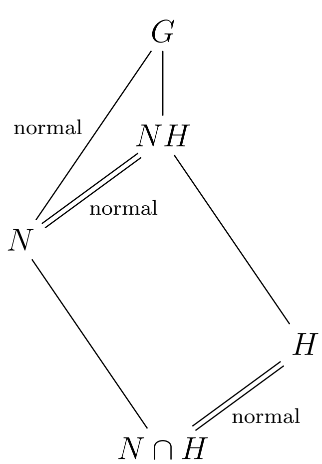
$$|H_1H_2| = |H_1||H_2|/|H_1 \cap H_2| \tag{13.5}$$

when H_1 and H_2 are finite subgroups of G, which will be useful in computation. We wish to determine a condition when H_1H_2 is a group and the set bijection f in (*) is a group isomorphism. This is the content of the next isomorphism theorem. It is very useful, but harder to visualize.

Theorem 13.6. (Second Isomorphism Theorem) *Let G be a group and H and N be subgroups with N normal in G. Then*

(1) $H \cap N \lhd H$.
(2) $HN = NH$ is a subgroup of G.
(3) $N \lhd HN$.
(4) $H/H \cap N \cong HN/N$.

We have the following picture



,

where a group below another group connected by a line is a subgroup and the quotients of the groups defined by the double lines are isomorphic.

Proof. (1): We know that $H \cap N$ is a subgroup of H and it is a normal subgroup in H, since N is normal in G and H is normal in H.

(2): As $aN = Na$ for all $a \in G$ the sets HN and NH are equal, so we need only show it is a group. Let $h_1, h_2 \in H$ and $n_1, n_2 \in N$. We must show $(h_1n_1)(h_2n_2)^{-1}$ lies in HN. But, by Great Trick,

$$(h_1n_1)(h_2n_2)^{-1} = h_1n_1{n_2}^{-1}{h_2}^{-1} = (h_1n_1{n_2}^{-1}{h_1}^{-1})(h_1{h_2}^{-1})$$

lies in $NH = HN$.

(3) is immediate.

(4): Define

$$\varphi : H \to HN/N \text{ by } x \mapsto xN.$$

As $\varphi(xy) = xyN = xNyN = \varphi(x)\varphi(y)$ for all x, y in N, the map is a homomorphism. We show $\ker\varphi = H \cap N$. If $x \in \ker\varphi \subset H$, then $N = e_{HN/N} = \varphi(x)$ so $x \in N$. Conversely, if $x \in H \cap N$, then $N = xN = \varphi(x)$ and $x \in \ker\varphi$. By definition

$$HN/N = \{hnN \mid h \in H,\, n \in N\} = \{hN \mid h \in H\},$$

so φ is surjective with kernel $H \cap N$. By the First Isomorphism Theorem φ induces an isomorphism $\overline{\varphi} : H/H \cap N \to HN/N$. □

Exercises 13.7.

1. Prove the properties of preimages in 13.1.
2. Prove the alternate form of the Correspondence Principle 13.3.
3. Let G be a group and H and K normal subgroups of G. We say that G is the *internal direct product* of H and K if $G = HK$ and $H \cap K = 1$. Show that if G is the internal direct product of H and K, then $G \cong H \times K$. Why is $H \times K$ not the internal direct product of G? For which subgroups of $H \times K$ is $H \times K$ the internal direct product?
4. Let G be be a group with normal subgroups N_i, $i = 1, \ldots, r$. Call G the the *internal direct product* of the N_i if $G = \langle\bigcup_{i=1}^{r} N_i\rangle$ and if for each j, $j = 1, \ldots, r$, $H_j \cap \langle\bigcup_{\substack{i=1\\ i\neq j}}^{r} N_i\rangle = 1$. Show that $G \cong N_1 \times \cdots \times N_r$.
5. Let G be a group with normal subgroups N_i, $i = 1, \ldots, r$ such that $G = N_1 \cdots N_r$. Show that G is the internal direct product of the N_i, $i = 1, \ldots, r$ if and only if for every $x \in G$ there exist unique $x_i \in N_i$ such that $x = x_1 \ldots x_n$.
6. Suppose that G is a group with normal subgroups N_i, $i = 1, \ldots, r$ such that $|G| = |N_1| \cdots |N_r|$ and the orders of the N_1 are pairwise relatively prime. Show that G is the (internal) direct product of the N_i, $i = 1, \ldots, r$ and if H is a subgroup of G, then H is the (internal) direct product of the $H \cap N_i$, $i = 1, \ldots, r$.
7. Let N be a normal subgroup of a group $G = H \times K$. Show that either N is abelian or N intersects one of H or K nontrivially.
8. Let N and H be two normal subgroups of G such that $G = HN$. Show that there is an isomorphism $G/(H \cap N) \cong G/H \times G/N$.

9. Let G be a finite group with normal subgroups H and K of relatively prime order. Show that the group HK is cyclic if H and K are cyclic and abelian if H and K are abelian.
10. Prove that the quaternion group is not an (external) semidirect product (cf. Exercise 11.9(16)) of of two groups neither of which is the identity group.
11. Let G be a group with subgroups H and N. We say that G is the *(internal) semidirect product* of H and N, denoted by $N \rtimes H$, if
   (i) $N \triangleleft G$
   (ii) $G = NH$
   (iii) $N \cap H = \{1\}$.
   Let $G = N \rtimes H$. Define $\varphi : H \to \operatorname{Aut}(N)$ by $\varphi_x(n) = xnx^{-1}$. Show φ is a group homomorphism and $\psi : N \rtimes H \to N \rtimes_\varphi H$ induced by $h \mapsto (e_N, h)$ and $h \mapsto (n, e_H)$ is a group isomorphism, where $N \rtimes_\varphi H$ is the (external) semidirect product of Exercise 11.9(16). [One often identifies these groups.]
12. Let G be a group. Recall the *commutator* G' of G is the subgroup generated by the *commutators* of elements of G, i.e., elements of the form $[x, y] := xyx^{-1}y^{-1}$ and the inverses of commutators. (Cf. Exercise 11.9(21).) Show
   (i) G/G' is abelian. G/G' is called the *abelianization* of G and denoted by G^{ab}.
   (ii) If $N \triangleleft G$ and G/N is abelian then $G' \subset N$. In particular, the abelianization $G^{ab} = G/G'$ of G is the maximal abelian quotient of G.
   (iii) If $G' \subset H \subset G$ then $H \triangleleft G$.
13. Let G be a finite abelian group and p a (positive) prime dividing the order of G. Show that there exists an element of order p in G.
14. In the previous exercise, prove the same result without assuming that G is abelian.

14. Finitely Generated Abelian Groups

In this section, we prove special cases of results that will be significantly generalized in subsequent sections. We do so in order to expose some of the ideas that shall be used later that apply in this much simpler situation.

Our first goal is to show that finite abelian groups are products of cyclic p-groups, unique up to order. (Cf. The Fundamental Theorem of Arithmetic.)

Lemma 14.1. *Let G be an abelian group and H_1, H_2 finite subgroups of relatively prime order. Then $H_1H_2 = H_2H_1$ is a group and if further $H_1 \cap H_2 = 1$, then $|H_1H_2| = |H_1||H_2|$. Moreover, if the groups H_1 and H_2 are both cyclic, then so is H_1H_2.*

We leave the proof of this as an exercise. The next result is a special case of Cauchy's Theorem (cf. Theorem 21.22 below) when G is abelian.

Proposition 14.2. *Let G be a finite abelian group and $p > 0$ a prime dividing the order of G. Then there exists an element of G of order p.*

Proof. We prove this by induction on $|G|$. We may assume that $G \neq 1$. As G has no nontrivial subgroups if and only if $G \cong \mathbb{Z}/p\mathbb{Z}$ for some prime p, we may assume that $|G|$ is not a prime. In particular, G has a subgroup $1 < H < G$. If $p \mid |H|$, we are done by induction on $|G|$, so we may assume $p \nmid |H|$. Therefore, there exists a prime $q \neq p$

such that $q \mid |H|$. By induction on $|G|$, there exists an element y in H of order q. As G is abelian, $\overline{G} = G/\langle y\rangle$ is a group. Let $^{-} : G \to \overline{G}$ be the canonical epimorphism. Since $|\overline{G}| < |G|$ and $p \mid |\overline{G}|$ by Lagrange's Theorem, there exists an element z in G such that $\overline{z}$ has order p in $\overline{G}$. It follows that z^p lies in the kernel of $^{-} : G \to \overline{G}$, so $z^p = y^i$ for some i, hence z^q has order p in G. □

We now use the proposition above to prove Sylow's First Theorem for the case when a group is abelian. [Cf. Theorem 22.1 below.]

Theorem 14.3. *Let G be a finite abelian group and $p > 0$ a prime dividing $|G|$, say $|G| = p^n m$ with p and m relatively prime. Then*

$$G(p) := \{x \in G \mid x^{p^r} = e \text{ some integer } r\} \triangleleft G \text{ and } |G(p)| = p^n.$$

Moreover, $G(p)$ is the unique subgroup of G of order p^n.

Remark 14.4. In the theorem, if G is finite and not abelian, there may be more than one group of order p^n. Indeed, we shall see there is only one if and only if there is a normal subgroup of order p^n.

Proof. The set $G(p)$ is a subgroup of G (why?) and normal as G is abelian. As every element in $G(p)$ has order a power of p, it follows that $|G(p)| = p^r$ some $r \geq 1$ (why?) and by Lagrange's Theorem that $r \leq n$. In particular, if x lies in $G(p)$, then $x^{p^n} = e$. We must show that $r = n$. Suppose to the contrary that we have $r < n$. Set $\overline{G} = G/G(p)$ and let $^{-} : G \to \overline{G}$ be the canonical epimorphism. By the Correspondence Principle (Alternative Form), $p \mid |\overline{G}|$, hence by the Proposition 14.2 and induction on $|G|$, there exists an element $x \in G \setminus G(p)$ such that

$$\overline{x}^p = \overline{e}, \text{ i.e., } x^p \in G(p).$$

As $x^{p^{n+1}} = (x^p)^{p^n} = e$, we have $x \in G(p)$, a contradiction. Thus $|G(p)| = p^n$ as desired.

If H is subgroup of G of order p^n, then $HG(p)$ is a subgroup of G whose order is $|H||G(p)|/|H \cap G(p)| \geq p^n$ by the Second Isomorphism Theorem. It follows that $|H| = |H \cap G(p)| = |G(p)| = p^n$, hence $H = H \cap G(p) = G(p)$. □

Corollary 14.5. *Let G be a finite abelian group of order $n = p_1^{m_1} \cdots p_r^{m_r}$ with positive primes $p_1 < \cdots < p_r$ and positive integers $m_1, \ldots, m_r$. Then $G = G(p_1) \cdots G(p_r)$. Moreover, $G \cong G(p_1) \times \cdots \times G(p_r)$.*

Proof. As $G(p_1) \cdots G(p_r)$ is a group and we know that

$$|G(p_1) \cdots G(p_r)| = |G(p_1)| \cdots |G(p_r)|$$

by Lemma 14.1, the first statement follows by the theorem. The map

$$G(p_1) \times \cdots \times G(p_r) \to G \text{ given by } (x_1, \ldots, x_r) \mapsto x_1 \cdots x_r$$

is a group homomorphism as $G(p_i)G(p_j) = G(p_j)G(p_i)$ for all i and j. Suppose that $x_1 \cdots x_r = e$ with $x_i \in G(p_i)$ for $i = 1, \ldots, r$. As $x_j^{n/p_i^{m_i}} = e$ for all $j \neq i$ and p divides the order of x_i, it follows that $x_i = e$ for all i. Thus the map is a monomorphism and by counting an isomorphism. □

It follows, to determine all finite abelian groups up to isomorphism, we have reduced to the study of groups of order a power of a prime. If $p > 1$ is a prime, a nontrivial group of order a power of p is called a *p-group*.

For notational reasons, it is easier to write abelian groups additively, i.e., use additive groups. If H and K are subgroups of an additive group, we know that $H + K$ is a subgroup. If H and K also satisfy $H \cap K = 0$, we shall write $H \oplus K$ for the group $H + K$. Of course, if H and K are finite groups of relatively prime order, we know that $H + K = H \oplus K$. The interesting case is when this is not true.

The following lemma is the key to the existence of a cyclic decomposition for p-groups.

Lemma 14.6. *Let G be a finite additive p-group and suppose that the element x in G has maximal order. Then there exists a subgroup H of G satisfying $G = \langle x \rangle \oplus H$.*

PROOF. Let p^n be the order of x. By the Well-ordering Principle, there exists a maximal subgroup H of G satisfying $H \cap \langle x \rangle = \{0\}$. Therefore, $\langle x \rangle + H = \langle x \rangle \oplus H$, and we are done if $\langle x \rangle + H = G$. So suppose not. Let $\overline{G} = G/(\langle x \rangle + H)$ and $^{-} : G \to \overline{G}$ the canonical epimorphism. By Proposition 14.2, there exists an element $y \in G$ satisfying $y \notin \langle x \rangle + H$ and $\overline{y}$ has order p in $\overline{G}$. In particular, py lies in $\langle x \rangle + H$. Write

$$(*) \qquad py = lx + h \text{ with } l \geq 0 \text{ in } \mathbb{Z} \text{ and } h \in H.$$

It follows that $p^n y = p^{n-1} lx + p^{n-1} h$. Since x has maximal order p^n among all elements in G, we have $0 = p^n y = p^{n-1} lx + p^{n-1} h$. Therefore, $p^{n-1} lx = -p^{n-1} h$ and it lies in $\langle x \rangle \cap H = 0$. It follows that $p \mid l$, say $l = l'p$. Thus (*) becomes

$$(\dagger) \qquad py = l'px + h \text{ equivalently } p(y - l'x) = h \text{ in } H.$$

As $y \notin \langle x \rangle + H$, we must have $y - l'x \notin \langle x \rangle + H$. The maximality of H now yields

$$\langle y - l'x, H \rangle \cap \langle x \rangle \neq 0.$$

Therefore, there exist integers r, s with r nonzero satisfying

$$0 \neq rx = s(y - l'x) + h' \text{ for some } h' \in H.$$

Hence $sy = rx + sl'x - h'$ lies in $\langle x \rangle + H$. We look at the integer s.

If $p \mid s$, then by (†), we have $s(y - l'x)$ lies in H. This implies that rx lies in $\langle x \rangle \cap H = 0$, a contradiction.

If $p \nmid s$, then p and s are relatively prime, so there exists an equation $1 = pa + sb$ for some integers a and b. Since both py and sy lie in $\langle x \rangle + H$, we have $y = apy + bsy$ lies in $\langle x \rangle + H$, a contradiction. The lemma now follows. □

Corollary 14.7. *Let G be a finite p-group. Then G is a product of cyclic p-groups.*

PROOF. We may assume that G is additive and not cyclic. In particular $|G| > p$. By the lemma, $G = \langle x \rangle \oplus H$ for some x in G and subgroup H of G. As G is not cyclic, $H \neq G$, and the result follows by induction on $|G|$. □

Remark 14.8. Let $p > 0$ be a prime, N a positive integer. Suppose that G is an additive group in which every element a in G satisfies $p^N a = 0$. Let x be an element of G of maximal order. Then there exists a subgroup H of G satisfying $G = \langle x \rangle \oplus H$. The proof is essentially the same as that above except that one replaces the Well-Ordering Principle by Zorn's Lemma, an axiom that we shall study later. [Cf. §28.]

Proposition 14.9. *Every finite abelian group is a product of cyclic groups.*

PROOF. By 14.5 every finite abelian group is a product of finite abelian p-groups. By Corollary 14.5, every finite abelian p-group is a product of cyclic p-groups. □

Theorem 14.10. (Fundamental Theorem of Finite Abelian Groups) *Let G be a finite additive group and for each prime $p > 0$ dividing G, let $G(p)$ be the unique p-subgroup of G of maximal order. Then*

$$G = \bigoplus_{p \mid |G|} G(p)$$

Moreover, if $p \mid |G|$, then

$$G(p) \cong \bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z}$$

with r unique and $1 \leq n_1 \leq \cdots \leq n_r$ also unique relative to this ordering. In particular, any finite abelian group is a product of cyclic p-groups for various primes p.

PROOF. Let $p \mid |G|$. By Corollary 14.5 and Proposition 14.9, it suffices to show $G(p) \cong \bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z}$ and uniquely up to isomorphism. As every abelian p-group is isomorphic to a product of cyclic p-groups by Lemma 14.6 and every cyclic p-group must be isomorphic to $\mathbb{Z}/p^a\mathbb{Z}$ for some integer a, to finish, it suffices to show that if

(*) $$\bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z} \cong \bigtimes_{j=1}^{s} \mathbb{Z}/p^{m_j}\mathbb{Z}$$

with $n_1 \geq \cdots \geq n_r$ and $m_1 \geq \cdots \geq m_s$ (we changed notation for convenience), then $r = s$ and $n_i = m_i$ for all i. We may assume that the number N of the n_i that are equal to 1 is less than the number M of the m_j that equal 1. In particular, $p^N \leq p^M$. Multiplying (*) by p, we see that

$$\bigtimes_{i=N+1}^{r} \mathbb{Z}/p^{n_i-1}\mathbb{Z} \cong \bigtimes_{j=M+1}^{s} \mathbb{Z}/p^{m_j-1}\mathbb{Z}$$

as $p(\mathbb{Z}/p^k\mathbb{Z}) \cong Z/p^{k-1}\mathbb{Z}$ for all k. By induction, $n_i - 1 = m_i - 1$ for all $i > N$. It follows that $N = M$ Hence $r = s$ and $n_i = m_i$ for all i. □

We can obtain an alternate form of the Fundamental Theorem of Finite Abelian Groups using the isomorphism $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/mnZ$ if $(m, n) = 1$, viz.,

Theorem 14.11. (Fundamental Theorem of Abelian Groups – Alternate Form) *Let G be a finite abelian group of order m. Then there exist unique positive integers $m_1, \ldots, m_r$ satisfying $m_1 \mid \cdots \mid m_r$ and $m = m_1 \cdots m_r$ for some r such that $G \cong \mathbb{Z}/m_1\mathbb{Z} \bigtimes \cdots \bigtimes \mathbb{Z}/m_r\mathbb{Z}$.*

We leave a proof of this as an exercise.

We now want to extend the Fundamental Theorem of Finite Abelian Groups to finitely generated abelian groups. In §44, we shall prove an important generalization of this extension that has significant application to the theory of matrices.

Some of our results below are true when discussing sets of infinitely many abelian groups. In this case, we must distinguish between two group constructions.

Definition 14.12. Let G_i, $i \in I$, be groups. We have defined the *direct product*

$$\bigtimes_{i\in I} G_i := \{(g_i)_I \mid g_i \in G_i,\ i \in I\}$$

with component-wise operations. (It is also denoted by $\prod_I G_i$.) If all the groups G_i, $i \in I$, are additive groups, the direct product contains a proper subgroup called the *(external) direct sum* of the G_i's defined by

$$\coprod_I G_i := \{(g_i)_I \in \bigtimes_{i\in I} G_i \mid g_i = e_{G_i} \text{ for all but finitely many } i \in I\}.$$

Of course, if I is a finite set, these two groups are the same.

We also want to look at an abelian group G containing subgroups G_i, $i \in I$, with

$$G = \sum_I G_i := \{\sum_I g_i \mid g_i \in G_i \text{ and } g_i \neq e_G \text{ for finitely many } i \in I\}$$

that is isomorphic to $\coprod_I G_i$. The condition for this to be true is that if

$$\sum_I g_i = 0 \text{ in } G,\ g_i \in G_i,\ i \in I, \text{ then } g_i = e_G \text{ for all } i \in I.$$

(Cf. this to linear independence in vector spaces.) If this is true, then we write $G = \bigoplus_I G_i$ and call it the *(internal) direct sum* of G by the G_i's. (Cf. Exercise 15.18(1).)

We begin with an analogue of vector spaces in group theory. We write our abelian groups additively.

Definition 14.13. Let G be an abelian group. We call G a *free abelian group* if there exists a *basis* $\mathcal{B}$ for G, i.e., for every element $g \in G$, there exist n_x in $\mathbb{Z}$ for $x \in \mathcal{B}$, with all but finitely many n_x nonzero, satisfying $g = \sum_{\mathcal{B}} n_x x$ and if $g = 0$, then $n_x = 0$ for all $x \in \mathcal{B}$. (Cf. bases for vector spaces.)

We show that the analogue of dimension makes sense for finitely generated abelian groups.

Lemma 14.14. *Let G be a finitely generated free abelian group on bases $\mathcal{B}$ and $\mathcal{C}$. Then $|\mathcal{B}| = |\mathcal{C}|$. In particular, any finitely generated abelian group has a finite basis.*

Proof. Let G be a free abelian group on basis $\mathcal{B} = \{x_1, \dots, x_n\}$. Let $p > 0$ be a prime and $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$, the canonical ring epimorphism. Define $V_{\mathcal{B}}$ to be the vector space over $\mathbb{Z}/p\mathbb{Z}$ on basis $\mathcal{B}$. Then $\varphi : G \to V_{\mathcal{B}}$ by $\sum_{i=1}^n n_i x_i \mapsto \sum_{i=1}^n \overline{n_i} x_i$ is a well-defined bijective map as $\mathcal{B}$ is a basis. (Note as groups, this is equivalent to the canonical group epimorphism $^- : G \to G/pG$.) In particular, it follows if $\mathcal{C} = \{y_1, \dots, y_m\}$ is another basis for G, then we must have $|\mathcal{B}| = \dim_{Z/p\mathbb{Z}} V_{\mathcal{B}} = \dim_{Z/p\mathbb{Z}} V_{\mathcal{C}} = |\mathcal{C}|$. If G is finitely generated, then the map $^- : G \to G/pG$ takes G onto the finitely generated vector space G/pG over $\mathbb{Z}/p\mathbb{Z}$. As any basis $\mathcal{B}$ of G must go to a basis of G/pG, it must be finite. $\square$

Remark 14.15. In the lemma, the common value of $|\mathcal{B}| = |\mathcal{C}|$ is called the *rank* of G. If G is a finitely generated free abelian group, we write $\operatorname{rank} G$ for the rank of G.

Corollary 14.16. *Let G be a finitely generated free abelian group of rank n. Then $G \cong \underbrace{\mathbb{Z} \times \cdots \times \mathbb{Z}}_{n} (= \mathbb{Z}^n)$ is product of n cyclic infinite groups.*

PROOF. The *standard basis* $\mathcal{S} := \{e_1, \cdots, e_n\}$ with $e_i = (0, \dots, 0, i, 0, \dots)$ for $i = 1, \dots, n$ is a basis for $\mathbb{Z}^n$. $\square$

Remark 14.17. More generally, if G is a free abelian group on basis $\mathcal{B}$ (possibly not finite), then by Exercise 14.27(12), $G \cong \coprod_{\mathcal{B}} \mathbb{Z}$ with the cardinality of $|\mathcal{B}|$ independent of basis $\mathcal{B}$ (assuming that the cardinality of a vector space is independent of bases that is proven using the Schroeder-Bernstein Theorem (cf. Remark 1.12 or Theorem Appendix A.13) and the Universal Property of Free Modules 39.3 (that includes vector spaces).

Definition 14.18. Let G be an abelian group. An element $x \in G$ is called a *torsion element* if the order of x is finite. Let

$$G_t := \{x \in G \mid x \text{ is torsion}\}.$$

We call an abelian group G a *torsion group* if $G = G_t$ and a *torsion-free group* if $G_t = 0$.

By the Binomial Theorem, $G_t \subset G$ is a subgroup and by the Correspondence Principle, G/G_t is torsion-free.

Examples 14.19. 1. Every finite abelian group if a torsion group.

2. Any infinite cyclic group is torsion-free, in fact, free abelian of rank one. In particular, any nontrivial torsion-free group is infinite.
3. If G is a free abelian group, then G is torsion-free.
4. $(\mathbb{Q}, +)$ is a torsion-free group, but it is not a free abelian group. Note that $(\mathbb{Q}, +)$ is not finitely generated.
5. If G is an abelian group, then G/G_t is a torsion-free group.

To prove our extension of the Fundamental Theorem of Finite Groups, we need the following new idea.

Definition 14.20. Let G be an additive group and $n \in \mathbb{Z}$. Set

$$nG := \{g \in G \mid g = nx \text{ for some } x \in G\}.$$

We call a subgroup $H \subset G$ *pure* if $H \cap nG = nH$ for all $n \in \mathbb{Z}^+$.

We begin with a simple observation that allows us to obtain a pure subgroup.

Lemma 14.21. *Let G be an additive group and H a subgroup of G with G/H torsion-free. Then H is a pure subgroup of G.*

PROOF. Let $^- : G \to G/H$ be the canonical epimorphism. If $g \in G$ satisfies $n\overline{g} = \overline{0}$ in $\overline{G}$, then g lies in $\ker {}^- = H$. $\square$

Next we look a a pure subgroup of an abelian group when the quotient group is a direct sum of cyclic groups.

Lemma 14.22. *Let G be an additive group and H a pure subgroup of G. Suppose that G/H is isomorphic to a direct sum of cyclic groups. Then $G \cong H \oplus K$ with K a subgroup isomorphic to G/H.*

PROOF. Let $^{-}: G \to G/H$ be the canonical epimorphism and

$$(*) \qquad \overline{G} = G/H = \oplus_I \mathbb{Z}\overline{x}_i$$

with $x_i \in G$ and the order $o(x_i)$ of x_i is n_i. If $\langle x_i \rangle \cong \mathbb{Z}$, we view n_i as ∞. Since H is a pure subgroup, there exist $h_i \in H$ satisfying $n_i x_i = n_i h_i$, so $n_i(x_i - h_i) = 0$, i.e., $o(x_i - h_i) \le n_i$ if n_i is finite. In this case, we have $\overline{x_i - h_i} = \overline{x_i}$, so $o(x_i - h_i) = o(\overline{x_i - h_i}) = o(\overline{x}_i)$. If x_i has infinite order, then $\overline{x}_i$ must also have infinite order, lest $n\overline{x}_i = 0$, some n, and $nx_i = nh$, some $h \in H$ which implies x_i has finite order. Therefore we can choose $y_i \in \overline{x_i}$ satisfying $\overline{y}_i = \overline{x}_i$ with y_i having order n_i for all i with n_i finite and $\overline{y}_i = 0$ if not. Set $K = \langle y_i \mid i \in I \rangle$. If $g \in G$, then $\overline{g} = \sum_I m_i \overline{y}_i$ for some $m_i \in \mathbb{Z}$, $i \in I$, almost all $m_i = 0$. Therefore, $g - \sum_I m_i y_i$ lies in $\ker{}^{-} = H$. It follows that $G = \langle K, H \rangle$. Suppose that $a \in H \cap K$. Relabeling, we can write

$$a = m_i y_1 + \cdots + m_r y_r$$

for some r, y_i. So

$$\overline{0} = \overline{a} = m_i \overline{y}_1 + \cdots + m_r \overline{y}_r$$

in $\overline{G}$. By (*), we have $n_i \mid m_i$ if n_i is finite and can assume $m_i = 0$ if $n_i = \infty$. It follows that $m_i = 0$ for all i. Hence $K \cap H = 0$ and $G = K \oplus H$ $\square$

Lemma 14.23. *Let G be a finitely generated free additive group and H a subgroup of G. Then H is a finitely generated free additive group. Moreover,* $\operatorname{rank} H \le \operatorname{rank} G$.

PROOF. By Lemma 14.14, we know that G has a finite basis $\mathcal{B} = \{x_1, \dots, x_n\}$ for some n. Set

$$H_j = H \cap \oplus_{i<j} \mathbb{Z}x_i.$$

Therefore, H_{j+1}/H_j, for $j < n$, is isomorphic to a subgroup of $\mathbb{Z}x_{j+1} \cong \mathbb{Z}$ (why?). Hence by the Cyclic Subgroup Theorem 9.11, $H_{j+1}/H_j \cong \mathbb{Z}$ or 0. By Lemma 14.22,

$$H_{j+1} = H_j \oplus K_j \text{ with } K_j \cong \mathbb{Z} \text{ or } 0 \text{ and } K_0 = H_1.$$

If follows that $H \cong \bigoplus_{i=0}^n H_i$, so is a free abelian group of finite rank. (Why?) $\square$

Remark 14.24. The lemma is true without the hypothesis that G be finitely generated. The proof is essentially the same if you know about ordinals noting that $H_j = \bigcup_{i<j} H_i$ if j is a limit ordinal.

Putting this together, we can classify finitely generated torsion-free abelian groups.

Proposition 14.25. *Let G be a finitely generated torsion-free additive group. Then G is a free abelian group of rank at most* $\operatorname{rank} G$.

PROOF. As G is finitely generated, it has a finite set of generators. We induct on finite sets of generators for G. If G is generated by a single element, then G is cyclic and torsion-free. Therefore, $G \cong \mathbb{Z}$. So by induction, we may assume the result for any finitely generated torsion-free additive group generated by n elements. Suppose that $G = \langle x_1, \dots, x_{n+1} \rangle$. Let $H = \langle x_1, \dots, x_n \rangle$. By induction, H is free abelian.

Let $^{-} : G \to G/H$ be the canonical epimorphism. Then G/H is a cyclic group generated by $\overline{x_{n+1}}$. Suppose that $G/H \cong \mathbb{Z}$ or 0. Then H is torsion-free. By Lemma 14.21, H is pure and by Lemma 14.22, $G \cong H \oplus K$ with $K \cong G/H$. It follows that G

is free abelian. So we may assume that $G/H \cong \mathbb{Z}/n\mathbb{Z}$, for some $n > 1$. This means that $nG \subset \ker{}^{-} = H$. Since G is torsion-free, $G \cong nG$. Therefore, G is free abelian by Lemma 14.23 We leave the last statement as an exercise. □

Putting the pieces together, we can now prove our generalization of the Fundamental Theorem of Finite Abelian Groups.

Theorem 14.26. (Fundamental Theorem of Finitely Generated Abelian Groups). *Let G be a finitely generated abelian group. Then there exists a free abelian subgroup F of G or unique rank such that $G = G_t \oplus F$. Moreover, G_t is finite and the product of cyclic p-groups unique up to isomorphism and order.*

Proof. As G is finitely generated, so is G/G_t. Since G/G_t is also torsion-free, it is a free abelian group of finite rank by Proposition 14.25. In particular, it is a finite product of infinite cyclic groups of unique rank by Lemma 14.14. By Lemma 14.21 and Lemma 14.22, $G = G_t \oplus K$ with K a finitely generated free abelian group of unique rank. Since the sum is direct, we must also have G_t finitely generated. (Why?) In particular, by the Fundamental Theorem of Finite Abelian Groups, G_t is a direct sum of cyclic p-subgroups unique up to order. The result follows. □

Exercises 14.27. 1. Prove Lemma 14.1. If we assume that G is an arbitrary group, what conditions on H_1 and H_2 will still guarantee the result. Prove this.

2. Let G be an abelian group. Show that $G(p) := \{x \in G \mid x^{p^r} = e \text{ some } r\}$ is a subgroup of G.
3. Show that $p(\mathbb{Z}/p^k\mathbb{Z}) \cong Z/p^{k-1}\mathbb{Z}$ for all primes p and positive integers k.
4. Prove the alternate form of the Fundamental Theorem of Finite Abelian Groups 14.11
5. Determine up to isomorphism all finite abelian groups up to order 360 in both forms of the Fundamental Theorem of Finite Abelian Groups.
6. Determine all $n > 1$ such that all abelian groups of order n are cyclic and prove your determination.
7. Let G be an abelian group and G_i, $i \in I$, subgroups satisfying $G = \sum_I G_i$. Prove that the following are equivalent:
   (*i*) $G = \bigoplus_I G_i$.
   (*ii*) If $\sum_I g_i = 0$ in G, then $g_i = e_G$ for all $i \in I$.
   (*iii*) For all $j \in I$, we have $G_j \cap \sum_{I \setminus \{j\}} G_i = \{e_G\}$.
   (*iv*) $\coprod_I G_i \to \sum_I G_i$ given by $(g_i)_I \mapsto \sum_I g_i$ is a group isomorphism.
8. Let G be an additive group and $x \in G$. Suppose that $\langle x \rangle$ is a pure torsion subgroup of G, Prove that $G = \langle x \rangle \oplus H$ for some subgroup $H \subset G$.
9. In the proof of Lemma 14.23, why is $H_{j+1}/H_j \cong \mathbb{Z}$?
10. If G is a finite free abelian group and H is a subgroup, prove that $\operatorname{rank} H \leq \operatorname{rank} G$. If $1 < H < G$, Can $\operatorname{rank} H = \operatorname{rank} G$?
11. Prove if G is a finitely generated group and $G = H \oplus K$ with K finitely generated, then H is also finitely generated.

12. Let G be an additive group. Suppose that G is a free abelian group on basis X. Prove all of the following:
    (*i*) If $n \in \mathbb{Z}^+$, then nG is free on basis $nX := \{nx \mid x \in X\}$.
    (*ii*) There exists an isomomorphism $\varphi : G \to \coprod_X \mathbb{Z}x \cong \coprod_{\varphi(X)} \mathbb{Z}$.
    (*iii*) Let $Y = \varphi(X)$. Then $\coprod_Y \mathbb{Z}$ is free abelian on basis Y.
    (*iv*) Let $p > 0$ be a prime and $\overline{Y} = \{y + p\coprod_Y \mathbb{Z} \mid y \in Y\}$. Then $\coprod_Y \mathbb{Z}/(p\coprod_Y \mathbb{Z}) \cong \coprod_{\overline{Y}}(\mathbb{Z}/p\mathbb{Z})$ and is a vector space over $\mathbb{Z}/p\mathbb{Z}$ on basis $\overline{Y}$.
    (*v*) We have $|Y| = |\overline{Y}|$. In particular, assuming the cardinality of a basis for a vector space is unique (which is true), any two bases for G have the same cardinality.
13. Let G be a torsion-free abelian group and H a subgroup of G. Prove all of the following:
    (*i*) H is a pure subgroup of G if and only if G/H is torsion-free.
    (*ii*) There exists a unique minimal pure subgroup H_{pu} of H containing H.
    (*iii*) if H_1 and H_2 are pure subgroups of G, then $H_1 \cup H_2$ is a pure subgroup of G

15. Addendum: Divisible Groups

In this section, we investigate another class of abelian groups consisting of non-finitely generated abelian groups (if we exclude the trivial group). These abelian groups are called divisible groups. They arise naturally from the divisibility property of the rational numbers. It is convenient to continue our use additive notation for our abelian group, and will call them additive groups as usual when we do so. In this section, we shall also need to use some results to be proven in later sections, the major one is a proof of Theorem 15.13 (which requires Zorn's Lemma) as does the fact that all vector spaces have bases. We shall also leave many details as exercises.

We extend our definition of the subgroup $G(p)$, p a prime, of a finite additive group G used in Section 14 to: If G is an additive group and $p > 0$ a prime, let

$$G(p) := \{x \in M \mid p^r x = 0 \text{ for some positive integer } r\},$$

a torsion group.

Definition 15.1. Let $\mathcal{P}$ denote the positive primes. Let G be an additive group and $p \in \mathcal{P}$. We call $G(p)$ the *p-primary* part of G. If $G = G(p)$, we call G a *p-primary group*. We also let

$$_pG := \{g \in G \mid pg = 0\},.$$

the subgroup generated by elements of order p in G (or $G(p)$).

We shall also need a suitable generalization of Fundamental Theorem Finite Abelian Groups 14.10. The general version is called the Primary Decomposition Theorem that we shall later prove in Theorem 44.20 below. The case that we need below in Theorem 15.2 can be proven by facts that you know (and Exercise 15.18(1)), so we omit it.

Theorem 15.2. *Let G be a torsion additive group. Then $G = \bigoplus_{\mathcal{P}} G(p)$.*

If G is an abelian group, it was mentioned in Section 14 [to be done later in Corollary 44.16], that $G \cong G_t \times (G/G_t)$ if G is finitely generated. It is not true in general, however, that $G \cong G_t \coprod (G/G_t)$ when G is not finitely generated. We give a counterexample.

Example 15.3. Let $G = \prod_{\mathcal{P}} \mathbb{Z}/p\mathbb{Z}$ as an additive group. We show that $G \ncong G_t \coprod (G/G_t)$. Let $x = (x_p)_{\mathcal{P}} \in G$. Suppose that $q \in \mathcal{P}$ satisfies $qy = x$ for some $y \in G$, i.e., $qy_p = x_p$ for all $p \in \mathcal{P}$. We say that x is *divisible* by q. Then $x_q = 0$. In particular, the only element $x \in G$ divisible by every $p \in \mathcal{P}$ is $x = 0$, i.e., that there cannot be a nonzero element in G that is divisible by every prime $p \in \mathcal{P}$.

Let $^{-} : G \to G/G_t$ be the natural group epimorphism. We show that G/G_t contains a nonzero element $\overline{a} \in G/G_t$, $a \in G$, divisible by all primes $p \in \mathcal{P}$. Let a_p be a generator of $\mathbb{Z}/p\mathbb{Z}$ and $a = (a_p)_{\mathcal{P}}$. Fix $q \in \mathcal{P}$. Then for every prime $p \neq q$, there exists $x_p \in \mathbb{Z}/p\mathbb{Z}$ satisfying $qx_p = a_p$. Set $x_q = 0$ and let $y \in G$ be the element with entries all 0 except for the qth coordinate that is a_q. In particular, $y \in G_t$ and $qx = a - y$. It follows that $q\overline{x} = \overline{a} - \overline{y} = \overline{a}$. Since this is true for every prime q, the element $\overline{a}$ in G/G_t is divisible by every prime. Since a cannot have finite order in G, we have $\overline{a}$ is nonzero and the assertion is verified. In particular, if $G \cong G_t \coprod (G/G_t)$, then there exists a nonzero element in G divisible by all primes, contradicting that no such element lies in G.

We shall show that that we do have such a decomposition for one type of abelian group. To do so, we will need a few facts about vector spaces. We shall assume that all vector spaces have bases. This will be proven in Proposition 28.6 below. By linear algebra, we then have two vector spaces over a field F are isomorphic if and only if they have bases of the same cardinality (in the infinite case one needs to use the Schroeder-Bernstein Theorem (cf. Remark 1.12)). For any F-vector space V, we write $\dim V$ for the cardinality of a basis (it is well-defined). We leave the proofs of the following three lemmas as exercises.

Lemma 15.4. *Let V be a vector space over a field F. Suppose that V has a basis $\mathcal{B} = \{v_i\}_I$ Then $(V, +) = \bigoplus_{\mathcal{B}} Fx$ as an additive group.*

Lemma 15.5. *Let V and W be vector spaces over F with F either $\mathbb{Q}$ or $\mathbb{Z}/p\mathbb{Z}$ with $p > 0$ a prime. Then V and W are isomorphic as vector spaces if and only if they are isomorphic as additive groups.*

Lemma 15.5 is false in general, e.g., it can be shown $\mathbb{R} \cong \mathbb{R}^2$ as abelian groups using that they have bases as vector spaces over $\mathbb{Q}$ of the same uncountable cardinality.

Lemma 15.6. *Let G be an additive group and $p > 0$ a prime. Then ${}_pG$ is a vector space over $\mathbb{Z}/p\mathbb{Z}$.*

Definition 15.7. Let G be an additive group. We say that G is a *divisible group* if for all $y \in G$, $n \in \mathbb{Z}^+$, there exists an $x \in G$ satisfying $nx = y$, i.e., $\frac{1}{n}y \in G$. Equivalently, the group homomorphism $G \to G$ given by $x \mapsto nx$ is surjective for all positive integers n.

Examples 15.8. 1. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all torsion-free divisible additive groups.

2. $\mathbb{R}^+$ is a multiplicative torsion-free divisible group

3. If G is a torsion-free divisible group, then $G \to G$ by $x \mapsto nx$ is an isomorphism for all $n \in \mathbb{Z}^+$.

4. $\mathbb{C}^\times$ is a divisible group. The circle group T in $\mathbb{C}^\times$ is the torsion subgroup of $\mathbb{C}^\times$.

Checking that the axioms of a vector space hold, we see that the following is true.

Lemma 15.9. *A torsion-free additive divisible group is a vector space over $\mathbb{Q}$.*

Corollary 15.10. *Let G be a torsion-free divisible group. Then $G \cong \coprod_I \mathbb{Q}$ for some indexing set I.*

The following observations are easily seen:

Observations 15.11. Let G be a nonzero divisible group. Then

1. G is an infinite group.
2. G_t is a divisible group.
3. If $\varphi : G \to G'$ is a group epimorphism, then G' is also a divisible group. In particular, G/G_t is a torsion-free divisible group.
4. Let G_i, $i \in I$, be additive groups. Then G_i is a divisible group for all $i \in I$ if and only if $\prod_I G_i$ is a divisible group if and only if $\coprod_I G_i$ is a divisible group.

Example 15.12. By Observation 15.11(3), the additive group $\mathbb{Q}/\mathbb{Z}$ is a divisible torsion group. It is not a finitely generated group, hence neither is $\mathbb{Q}$. Let p be a prime number. Define $\mathbb{Z}_{p^\infty} := (\mathbb{Q}/\mathbb{Z})(p)$. Then $\mathbb{Z}_{p^\infty}$ is a divisible group and is not finitely generated. (Why?) Moreover, by Theorem 15.2, we have $\mathbb{Q}/\mathbb{Z} = \bigoplus_{\mathcal{P}} \mathbb{Z}_{p^\infty}$.

We wish to classify divisible groups. To do so, we shall need a defining property of divisible groups. Unfortunately, to prove this theorem, we need to use Zorn's Lemma to be done later, so we shall postpone it proof until then [cf. Proposition 28.10]. We shall, however, assume its validity here.

Theorem 15.13. *Let G be a divisible abelian group and B an abelian group with $A \subset B$ a subgroup. Suppose that $\varphi : A \to G$ is a group homomorphism. Then there exists a group homomorphism $\psi : B \to G$ satisfying $\psi|_A = \varphi$. We say that ψ extends φ to G.*

Corollary 15.14. *Let G be an abelian group and H a subgroup of G that is a divisible group. Then $G = H \oplus X$, for some subgroup X of G.*

PROOF. The identity map $1_H : H \to H$ extends to $\psi : G \to H$, i.e., $\psi|_H = 1_H$ by the theorem. We show that $G = H \oplus \ker\psi$. If $x \in G$, then $\psi(x) \in H$, hence $\psi\big(x - \psi(x)\big) = \psi(x) - 1_H\psi(x) = 0$. Therefore, $\big(x - \psi(x)\big) + \psi(x)$ lies in $H + \ker\psi$. If $x \in H \cap \ker\psi$, then $x = \psi(x) = 0$. This shows that $G = \ker\psi \oplus H$ by Exercise 15.18(1) as needed. $\square$

Using Theorem 15.13, we prove the following lemma.

Lemma 15.15. *Suppose that G_1 and G_2 are divisible p-primary groups. Then $G_1 \cong G_2$ if and only if ${}_pG_1 = {}_pG_2$.*

PROOF. ($\Rightarrow$) follows easily from Lemma 15.6
($\Leftarrow$): Let $\varphi : {}_pG_1 \to {}_pG_2$ be a group isomorphism. We view it as a monomorphism into G_2. By Theorem 15.13, we can extend φ to $\psi : G_1 \to G_2$, i.e., $\psi|_{{}_pG_1} = \varphi$. We show that ψ is an isomorphism.

ψ is injective: We show that $x = 0$ by induction on elements of order p^r in G_1. Let x be a nonzero element in G_1 of order p^r and suppose that $\psi(x) = 0$. If $r = 1$, then $px = 0$ and

$x = 0$ as φ is an isomorphism. Let $r > 1$. Since px has order p^{r-1}, by induction $px = 0$, contradicting x has order p^r. Therefore, ψ is injective.

ψ is surjective: Let $y \in G_2$ have order p^r. If $r = 1$, since φ is an isomorphism, there exists an $x \in G_1$ such that $\psi(x) = y$. So assume that $r > 1$. Since $0 \neq p^{r-1}y \in {}_pG_2$, there exists $x \in {}_pG_1$ such that $\psi(x) = p^{r-1}y$. Since G_1 is divisible, there exist $z \in G_1$ such that $p^{r-1}z = x$. Therefore, $p^{r-1}\big(y - \psi(z)\big) = 0$. By induction on r, there exists an element $w \in G_1$ satisfying $\psi(w) = y - \psi(z)$. Therefore, $y = \psi(w + z)$. □

We can now classify divisible abelian groups.

Theorem 15.16. *Let G be a divisible additive group. Then*

$$G = \coprod_I \mathbb{Q} \coprod \Big(\coprod_{\mathcal{P}} \coprod_{I_p} \mathbb{Z}_{p^\infty}\Big)$$

for some indexing sets I, I_p, $p \in \mathcal{P}$ (possibly empty).

Proof. Since G_t is divisible by Observation 15.11(2), we have $G \cong G_t \coprod (G/G_t)$ by Corollary 15.14. Since G/G_t is torsion-free, it is a direct sum of copies of $\mathbb{Q}$ by Lemma 15.10. By Observation 15.11(4) and Theorem 15.2 each $G(p)$ is divisible. Finally, let G'_p be the direct sum of $\dim_{\mathbb{Z}/p\mathbb{Z}}({}_pG)$ copies by of $\mathbb{Z}_{p^\infty}$. By Lemma 15.15, the groups $G(p)$ and G'_p are isomorphic. □

Divisible groups have the important property that given any abelian group A, there exits a divisible abelian group G and a group monomorphism $\varphi : A \to G$. To establish this, we need some further ideas from vector space theory.

The additive group, the direct sum $\coprod_I \mathbb{Z}$ has a $\mathbb{Z}$-basis (with scalars from $\mathbb{Z}$) just as with an F-vector space over a field F (with scalars from F). In fact, $\mathcal{S} := \{e_i \mid i \in I\}$ is a basis for $\coprod_I \mathbb{Z}$ where e_i has 1 in the ith coordinate and 0 in the jth coordinate for all $j \neq i$ in I. So every every element x is a *finite* sum of some of the e_i, i.e., $x = \sum_I a_i e_i$ with $a_i \in \mathbb{Z}$, all but finitely many a_i nonzero. This means that $\mathcal{S}$ spans. In addition, the a_i are unique as $y = 0$ in $\coprod_I \mathbb{Z}$ if and only all the coordinates are zero. This means that $\mathcal{S}$ is linear independence. Just as for vector spaces, this means that if G is an abelian group and $a_i \in \mathbb{Z}$, $i \in I$, (not necessarily distinct), there exists a unique group homomorphism induced by $e_i \mapsto a_i$ for all $i \in I$. [The same proof used for vector spaces works.] For example, since $\mathcal{S} \subset \coprod_I \mathbb{Q}$ (and is a $\mathbb{Q}$-basis for it), the inclusion map $inc : \coprod_I \mathbb{Z} \to \coprod_I \mathbb{Q}$ via $e_i \to e_i$ is a group monomorphism.

Theorem 15.17. *Let A be an abelian group. Then there exist a divisible abelian group G and a group monomorphism $\varphi : A \to G$.*

Proof. Let $\{a_i \mid i \in I\}$ be a set of generators for A. Then there exists a group homomorphism $\psi : \coprod_I \mathbb{Z} \to A$ determined by $e_i \mapsto a_i$ for all $\imath \in I$. Let $K = \ker \psi$. By the First Isomorphism Theorem, ψ induces a group isomorphism $\overline{\psi} : (\coprod_I \mathbb{Z})/K \to A$. The group monomorphism $inc : \coprod_I \mathbb{Z} \to \coprod_I \mathbb{Q}$ takes $K \to \mathbb{Q}K$, where $\mathbb{Q}K$ is the $\mathbb{Q}$-vector space spanned by $\{e_i \mid i \in I\}$. Since $\mathbb{Q}K \cap \coprod_I \mathbb{Z} = K$ (why?), the map $\iota : (\coprod_I \mathbb{Z})/K \to (\coprod_I \mathbb{Q})/K$ is a group monomorphism. By Observation 4(3), the group $(\coprod_I \mathbb{Q})/K$ is a divisible group. Therefore, the composition $\varphi := \iota \circ \overline{\psi}^{-1} : A \to (\coprod_I \mathbb{Q})/K$ works. □

Exercises 15.18. 1. Let G be an abelian group and G_i, $i \in I$ subgroups satisfying $G = \sum_I G_i$. Prove that the following are equivalent:

   (i) $G = \bigoplus_I G_i$.

   (ii) If $\sum_I g_i = 0$ in G, then $g_i = e_G$ for all $i \in I$.

   (iii) For all $j \in I$, we have $G_j \cap \sum_{I \setminus \{j\}} G_i = \{e_G\}$.

   (iv) $\coprod_I G_i \to \sum_I G_i$ given by $(g_i)_I \mapsto \sum_I g_i$ is a group isomorphism.

2. Prove Lemma 15.4
3. Prove Lemma 15.5
4. Prove Lemma 15.6
5. Prove Theorem 15.2
6. Prove Lemma 15.9
7. Let G be the group in Example 15.3. Show that G/G_t is a divisible group.
8. Let p be a prime number. Show all of the following:

   (i) $\mathbb{Z}_{p^\infty}$ is a divisible group.

   (ii) The set $\{\frac{1}{p^r} + \mathbb{Z} \,|\, r$ a non-negative integer $\}$r generates $\mathbb{Z}_{p^\infty}$. In particular, $\mathbb{Z}_{p^\infty}$ is not finitely generated.

   (iii) The subgroup $\left\langle \frac{1}{p^r} + \mathbb{Z} \right\rangle$ of $\mathbb{Z}_{p^\infty}$ is isomorphic to $\mathbb{Z}/p^r\mathbb{Z}$ for all $r \in \mathbb{Z}^+$.

9. Prove Observations 15.11.
10. Let G_1 and G_2 be divisible abelian groups. Prove that $G_1 \cong G_2$ if and only if $\dim_{\mathbb{Q}} \big(G_1/(G_1)_t\big) = \dim_Q \big(G_2/(G_2)_t\big)$ and $\dim_{\mathbb{Z}/p\mathbb{Z}}({}_pG_1) = \dim_{\mathbb{Z}/p\mathbb{Z}}({}_pG_2)$ for all primes p.
11. Prove that the following groups are isomorphic: $\mathbb{R}/\mathbb{Z}$, the circle group T, $\prod_p \mathbb{Z}_{p^\infty}, \mathbb{Q}\mathbb{Z} \oplus \mathbb{R}$.
12. An abelian group G is divisible if and only if it satisfies the property in Theorem 15.13.
13. An abelian group G is divisible if and only if every nonzero quotient of G is infinite.
14. An abelian group G is divisible if and only if it has no maximal subgroups, i.e., a subgroup $M < G$ so that there exists no subgroup H in G with $M < H < G$.
15. Let A be an arbitrary abelian group and $B \subset A$ a subgroup with A. Let $a \in A \setminus B$. Show that there exists a group homomorphism $\varphi : A \to D$ with D a divisible group satisfying $\varphi|_B = 0$ and $\varphi(a) \neq 0$.

16. Addendum: Finitely Generated Groups

Most of what we do in the study of group theory in these pages involves finite groups. In this short section, we show how the General Cayley Theorem 12.4 can be used to obtain a few results about finitely generated groups, i.e., a group G containing a finite set S of elements satisfying $G = \langle S \rangle$. Of course, finite groups are finitely generated, but the results that we obtain here are trivial for finite groups, as the questions we ask are about finitely generated groups that contain a proper subgroup of finite index.

We have seen that if G is an arbitrary group and H a subgroup of finite index n in G, then the General Cayley Theorem gives a group homomorphism

$$\lambda : G \to \Sigma(G/H) \cong S_n \text{ given by } x \mapsto \lambda_x : gH \mapsto xgH,$$

with $\ker \lambda \subset H \subset G$ (and the largest normal subgroup of G in H). Hence, by the First Isomorphism Theorem, λ induces a group monomorphism

$$\overline{\lambda} : G/\ker \lambda \to \Sigma(G/H) \cong S_n \text{ given by } x \mapsto \lambda_x : g \mapsto gH.$$

Since $\Sigma(G/H)$ is a finite group, so is $G/\ker \lambda$, i.e., $\ker \lambda \subset G$ has finite index. Therefore, G contains a normal subgroup of finite index. [Cf. Exercise 12.12(7).] We want to strengthen this result if G is finitely generated.

Proposition 16.1. *Let G be a finitely generated group and n a positive integer. Then there exist finitely many subgroups (if any) of G of index n.*

PROOF. Let $G = \langle a_1, \dots, a_r \rangle$ and

$$(*) \qquad \varphi : G \to S_n$$

be a group homomorphism. Then the map φ is completely determined by

$$\varphi(a_1), \dots, \varphi(a_r)$$

in the finite set S_n. Thus there exist only finitely many possible group homomorphisms φ in (*), hence finitely many normal subgroups N of G of finite index with $N = \ker \varphi$ for such φ. Fix a φ. By the First Isomorphism Theorem, we have $G/\ker \varphi\lambda \cong \varphi(G) \subset S_n$ is a subgroup. As S_n is finite, the Correspondence Principle implies that there exist finitely many subgroups H of G satisfying $\ker \varphi \subset H \subset G$, in particular, there can be only finitely many H (if any) with $[G : H] = n$. As there are finitely many φ, the result follows. □

Our strengthening of the goal above can now be stated and established. Recall that a subgroup K of a group G is called *characteristic* if $\sigma|_K$ is an automorphism of K (equivalently, $\sigma(K) = K$) for every automorphism σ of G.

Corollary 16.2. *Let G be a finitely generated group. Suppose G contains a subgroup H of finite index. Then there exists a characteristic subgroup K of H of finite index in G.*

PROOF. By the proposition, there exist finitely many subgroups

$$H = H_1, \dots, H_m$$

of finite index $[G : H]$. It follows, given an automorphism φ of G, that $\varphi(H) = H_i$ for some i, i.e., φ permutes the H_i. Let $K := \cap_{I=1}^m H_i$. Then we have $\varphi(K) = K$, so K is a characteristic subgroup of H. By Poincaré's Lemma (cf. Exercise 10.16(7)), K is of finite index in G. □

The second result is much harder to prove.

Theorem 16.3. *Let G be a finitely generated group and H be a subgroup of finite index. Then H is a finitely generated group.*

PROOF. Let $G = \langle a_1, \dots, a_r \rangle$ and

$$y_1 = e, \dots, y_n$$

be a transversal (i.e., a system of representatives for the cosets) of H in G. Let

$$\lambda : G \to \Sigma(G/H) \text{ be given by } a \mapsto \lambda_x : aH \mapsto xaH.$$

Therefore, λ_x permutes $y_1 H, \dots, y_n H$. Fix i, $1 \leq i \leq r$. Then for each j, $1 \leq j \leq n$, there exists a $k = k(i,j)$ with $1 \leq k \leq n$ satisfying

$$a_i y_j H = y_k H.$$

Therefore, there exist h_{ij} in H, $1 \leq i \leq r$ and $1 \leq j \leq n$ satisfying

$$y_k = a_i y_j h_{ij}.$$

Let $H_0 := \langle h_{ij} \mid 1 \leq i \leq r, 1 \leq j \leq n \rangle \subset G$. Then H_0 is clearly a finitely generated subgroup of G contained in H. So it suffices to show that $H \subset H_0$. Set

$$W := \bigcup y_j H_0.$$

For each i, $1 \leq i \leq r$, we have

$$a_i W = \bigcup_{j=1}^{n} a_i y_j H_0 = \bigcup_{j=1}^{n} y_{k(i,j)} h_{ij} H_0 = \bigcup_{k=1}^{n} y_k H_0 = W,$$

since

$$\{a_i y_1 H, \dots, a_i y_n\} = \{y_1, \dots, y_n\}.$$

Consequently, $a_i W = W$ for each $i = 1, \dots, r$. As $G = \langle a_1, \dots, a_r \rangle$, we have

$$GW = W = \bigcup_{j=1}^{n} H_0.$$

Since $e \in W$, we conclude that

$$G = W = \bigcup_{j=1}^{n} y_j H_0 = \bigvee_{j=1}^{n} y_j H_0.$$

In particular,

$$H \subset G = \bigvee_{j=1}^{n} y_j H_0.$$

But $H = y_1 H$ is disjoint from $\bigcup_{j=2}^{n} y_j H_0$ as it is disjoint from $\bigcup_{j=2}^{n} y_j H$, so we conclude that $H = H_0$ as needed.

$\square$

17. Series

One way of studying algebraic objects is to break them down into simpler pieces. For example, if G is a group, it may be the direct product of groups, e.g., $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$. Unfortunately, this does not happen very often for groups. An alternative approach would be to take a given nontrivial group G and find the largest normal subgroup N_1 in G satisfying $N_1 < G$. By the Correspondence Principle, G/N_1 then would be a simple group, and we have, in some sense reduced the study of G to the groups N_1 and G/N_1. We then proceed with the same analysis on N_1, etc. Another approach would be to find a normal subgroup N_1 of G such that G/N_1 is abelian (or cyclic), then continue on N_1, etc. In any of these cases, we produce a chain of subgroups $G > N_1 > N_2 > \cdots$ (where $A > B$ means $A \supset B$, $A \neq B$). Then various possibilities open:

(1) This process never stops.
(2) There exists an i such that N_i is nice, e.g., in our cases simple, abelian, cyclic, respectively.
(3) At some point, there does not exist such an N_i.

Of course, we are interested in the second possibility.

Definition 17.1. Let G be a nontrivial group. A sequence of groups

$$N_0 \subset N_1 \subset N_2 \subset \cdots \subset N_n = G \tag{*}$$

is called a *subnormal* (respectively, *normal*) *series* if $N_i \triangleleft N_{i+1}$ (respectively, $N_i \triangleleft G$) for all i, and *proper* if $N_i < N_{i+1}$ for all i. The quotients $(G/N_{n-1} =)N_n/N_{n-1}, \dots, N_1/N_0$ are called the *factors* of the series. If (*) is a subnormal series for which $N_0 = 1$ and $N_n = G$, then it is called

(*i*) a *cyclic series* if N_{i+1}/N_i is cyclic for all i.
(*ii*) an *abelian series* if N_{i+1}/N_i is abelian for all i.
(*iii*) a *composition series* if N_{i+1}/N_i is simple for all i.

A group is called *solvable* if it has an abelian series and *polycyclic* if it has a cyclic series. (Since the trivial group is cyclic, we also say that it has a cyclic series.)

Example 17.2. Every abelian group is solvable and every cyclic group is polycyclic.

One of the major results in finite group theory is the Feit-Thompson Theorem that every finite group of odd order is solvable. It is the first major theorem in the classification of finite simple groups. Its proof takes over 250 journal pages.

Theorem 17.3. *The following are true:*

(1) A subgroup of a solvable group is solvable.
(2) The homomorphic image (i.e., the image of a group under a group homomorphism) of a solvable group is solvable.
(3) If $N \triangleleft G$ and both N and G/N are solvable then so is G.

Proof. We leave this as an exercise. It is important that you do this exercise, as it will teach you how to use the isomorphism theorems. □

Let G be a group. The *commutator* of x and y in G is $[x, y] := xyx^{-1}y^{-1}$. The group generated by all commutators is called the *derived subgroup* of G and denoted either by $[G, G]$ or by G'. By recursion, we define, the *nth derived subgroup* of G by $G^{(n)} = [G^{(n-1)}, G^{(n-1)}]$. We leave the following easy computations as an exercise.

Properties 17.4. Let G be a group. If a, b, c are elements of G and $\sigma : G \to G_1$ a group homomorphism. Then

(1) $[a, b]^{-1} = [b, a]$.
(2) $\sigma([a, b]) = [\sigma(a), \sigma(b)]$.
(3) $[a, bc] = [a, b]b[a, c]b^{-1} = [a, b][bab^{-1}, bcb^{-1}]$.

Proposition 17.5. *Let G be a group. Then $G^{(n)}$ is a characteristic subgroup of $G^{(n-1)}$ for all n. In particular, $G^{(n)}$ is characteristic in G, and*

$$G^{(n)} \subset G^{(n-1)} \subset \cdots \subset G^{(1)} \subset G^{(0)} = G$$

is a normal series (even a characteristic series — obvious definition).

PROOF. Being characteristic is transitive, so this follows immediately, using Properties 17.4 □

Theorem 17.6. *Let G be a group. Then G is solvable if and only if there exist an integer n such that $G^{(n)} = 1$.*

PROOF. By Exercise 11.9(21), each factor group $G^{(i)}/G^{(i+1)}$ is abelian. Hence if $G^{(n)} = 1$ for some n, then G is solvable. Conversely, suppose that

$$(\dagger) \qquad 1 = N_n \subset N_{n-1} \subset N_{n-2} \subset \cdots \subset N_1 \subset G$$

is an abelian series. It suffices to show that $G^{(i)} \subset N_i$ for all i. By Exercise 13.7(12), we know that $G' \subset N_1$. By induction, we may assume that $G^{(i-1)} \subset N_{i-1}$. But then we have $G^{(i)} = [G^{(i-1)}, G^{(i-1)}] \subset [N_{i-1}, N_{i-1}] \subset N_i$, again using Exercise 13.7(12). □

We turn to composition series. We need the following easy lemma.

Lemma 17.7. (Dedekind's Modular Law) *Let A, B, and C be three subgroups of G. If A is a subset of C, then $A(B \cap C) = (AB) \cap C$.*

PROOF. $\subseteq$: If $a \in A$, $x \in B \cap C$, then $ax \in AB \cap AC \subset (AB) \cap C$.
$\supseteq$: If $a \in A$, $b \in B$ satisfy $ab \in C$, then $b = a^{-1}C \subset C$. □

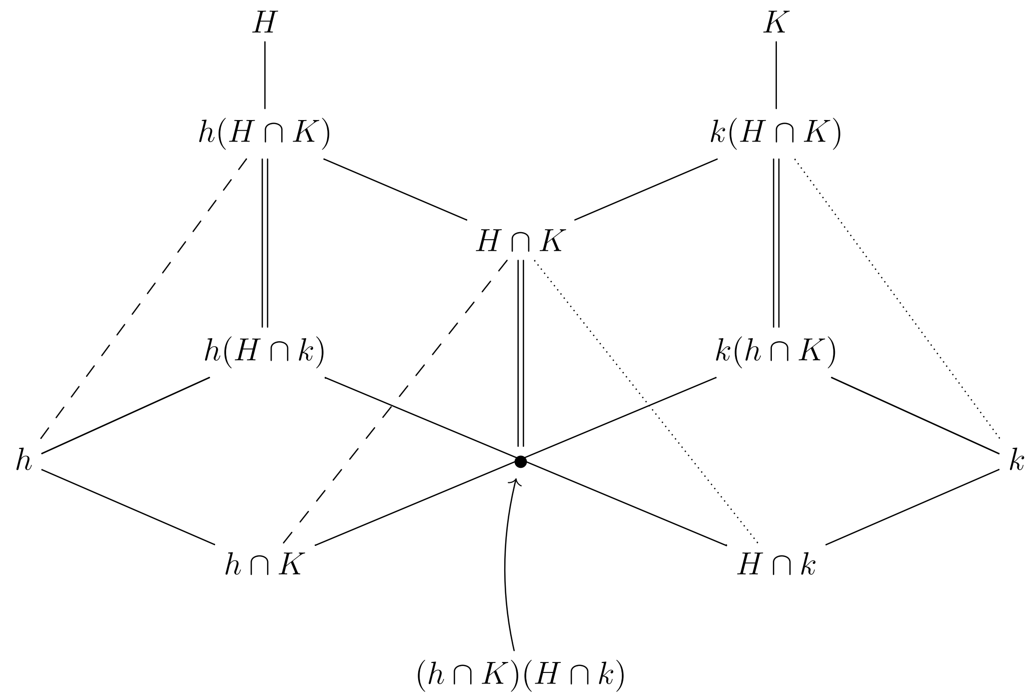
We next prove an elaborate form of the Second Isomorphism Theorem. For notational simplicity, we use right cosets.

Lemma 17.8. (Zassenhaus Butterfly Lemma) *Let G be a group containing four subgroups h, H, k, and K satisfying $h \triangleleft H$ and $k \triangleleft K$. Then we have the following:*

(1) $(H \cap k)(h \cap K) \triangleleft H \cap K$.
(2) $k(h \cap K) \triangleleft k(H \cap K)$.
(3) $h(H \cap k) \triangleleft h(H \cap K)$.
(4) There exist isomorphisms:

$$\begin{aligned} k(H \cap K)/k(h \cap K) &\cong H \cap K/(H \cap k)(h \cap K) \\ &\cong h(H \cap K)/h(H \cap k). \end{aligned}$$

We have the following picture illustrating the subgroups, where a group at the bottom of a line means that it is a subgroup of the group at the other end. We shall see the quotients (top/bottom) of like dotted, double, dashed lines, respectively, are isomorphic.



Proof. We use the Second and Third Isomorphism Theorems. As $k \triangleleft K$ in the lemma and $H \cap K \subset K$ is a subgroup, we have $k(H \cap K)$ is a group satisfying $k \triangleleft k(H \cap K)$, and $H \cap k = (H \cap K) \cap k \triangleleft H \cap K$ together with an isomorphism

$$(*) \qquad k(H \cap K)/k \xrightarrow{\sim} H \cap K/H \cap k \text{ given by } ka \mapsto (H \cap k)a.$$

Similarly, as $h \triangleleft H$ and $H \cap K \subset H$ is a subgroup, we have $h(H \cap K)$ is a group satisfying $h \triangleleft h(H \cap K)$, and $h \cap K \triangleleft H \cap K$ together with an isomorphism

$$h(H \cap K)/h \xrightarrow{\sim} H \cap K/h \cap K \text{ given by } ha \mapsto (h \cap K)a.$$

As the product of normal subgroups is normal, we also have

$$(\dagger) \qquad (H \cap k)(h \cap K) \triangleleft H \cap K.$$

By restriction, the isomorphism in (*) induces an isomorphism

$$(\star) \qquad k(H \cap k)(h \cap K)/k \xrightarrow{\sim} (H \cap k)(h \cap K)/H \cap k.$$

By the Modular Law, we have $k(H \cap k) = kH \cap k = k$. Consequently, $k(H \cap k)(h \cap K) = k(h \cap K)$. Applying the Correspondence Principle to the isomorphism in (*), we know that $(H \cap k)(h \cap K) \triangleleft H \cap K$ corresponds to $k(h \cap K) = k(H \cap k)(h \cap K) \triangleleft k(H \cap K)$.

Using the Third Isomorphism Theorem, we conclude by (*) and ($\star$) that

$$\begin{aligned} k(H\cap K)/k(h\cap K) &\cong \frac{k(H\cap K)/k}{k(h\cap K)/k} \\ &\cong \frac{H\cap K/H\cap k}{(H\cap k)(h\cap K)/H\cap k} \\ &\cong H\cap K/(H\cap k)(h\cap K). \end{aligned}$$

The other isomorphism is obtained similarly. □

Next we wish to show if a group has a composition series, then the number of terms in a minimal, proper composition series is constant with factors unique up to isomorphism.

Definition 17.9. Given a proper subnormal series of a group G, we say another subnormal series is a *refinement* of the given series, if we add in new subgroups, and it is called a *proper refinement* if the resulting subnormal series is proper. Two proper subnormal series for G are called *equivalent* if they have the same number of terms, say n, and the n factors of each are isomorphic up to a permutation.

If a group has a composition series, we want to prove that the number of factors in a composition series is unique as well as these factors are unique up to isomorphism and order. The next theorem is the key to this. We use the Butterfly Lemma to prove it.

Theorem 17.10. (Schreier Refinement Theorem) *Any two proper subnormal series for a group G have equivalent refinements.*

Proof. Suppose

$$\begin{aligned} &1 = N_0 \lhd N_1 \lhd \cdots \lhd N_r = G \quad \text{and} \\ &1 = H_0 \lhd H_1 \lhd \cdots \lhd H_s = G \end{aligned}$$

are two proper subnormal series for G. We have groups:

$$N_{i,j} = N_i(H_j \cap N_{i+1})$$

that form subnormal series

$$N_i = N_{i,0} \lhd \cdots \lhd N_{i,s} = N_{i+1} \quad \text{for all } i.$$

by the Butterfly Lemma. It follows that we have a subnormal series

$$1 = N_{0,0} \lhd \cdots \lhd N_{r,s} = G. \tag{1}$$

Similarly, we have another subnormal series

$$1 = H_{0,0} \lhd \cdots \lhd H_{s,r} = G \tag{2}$$

where

$$H_{j,i} = H_j(N_i \cap H_{j+1}).$$

By the Butterfly Lemma, we have

$$N_{i,j+1}/N_{i,j} \cong H_{j,i+1}/H_{j,i} \quad \text{for all } i \text{ and } j,$$

hence the series (1) and (2) are equivalent. In particular, if a factor isomorphic to 1 occurs in one, it occurs in the other. It follows that from these we can obtain common proper refinements. $\square$

If $1 < N_1 \triangleleft \cdots \triangleleft N_{r-1} \triangleleft G$ is a proper subnormal series, we say the series has *length* r (which equals the number of *links*). As the trivial group is not considered a simple group, any composition series for a group must be proper. Moreover, by the Correspondence Principle, no composition series can have a proper refinement.

Theorem 17.11. (Jordan-Hölder Theorem) *Let G be a group having a composition series of length r. Then every composition series for G has length r and any proper subnormal series for G can be refined to a composition series for G. Moreover, all composition series of G are equivalent.*

Proof. Let $1 = N_0 < \cdots < N_n < G$ be a proper subnormal series for G. By the Schreier Refinement Theorem, this subnormal series and a composition series for G have a common proper refinement. The result follows by the above remark. $\square$

Example 17.12. Let $m = a_1 \cdots a_r$ in $\mathbb{Z}$ with each integer $a_i > 1$. Then we have a subnormal series:

$$0 = a_1 \cdots a_r\mathbb{Z}/m\mathbb{Z} < \cdots < a_1\mathbb{Z}/m\mathbb{Z} < \mathbb{Z}/m\mathbb{Z}.$$

As $\mathbb{Z}/m\mathbb{Z}$ is a finite group, it has a composition series by Exercise 17.14(10) that is essentially unique. This is really the Fundamental Theorem of Arithmetic.

Remark 17.13. In an analogous way, one can define composition series for vector spaces. A "simple" vector space is a line [proof?]. Finite dimensional vector spaces have composition series (and infinite dimensional ones do not). Let V be a nonzero finite dimensional F-vector space. Say V has a composition series of length d, i.e, we get a a series $0 < V_1 < \cdots < V_d = V$, with simple factors V_i/V_{i-1} (after generalizing the concept of quotients to vector spaces). Thus d is the *dimension* of V, and the Jordan-Hölder Theorem gives an alternate proof of the invariance of the number of vectors in the basis for a finite dimensional vector space.

Exercises 17.14.

1. Prove Properties 17.4.
2. Prove that every finite abelian group is polycyclic. Use this to prove that every finite solvable group is polycyclic. Give an example of a solvable group that is not polycyclic.
3. Prove Theorem 17.3 using the isomorphism theorems. (Do not use Theorem 17.6.)
4. Prove the analogue of Theorem 17.3 replacing the word solvable with the word polycyclic using the isomorphism theorems.
5. Let H and K be subgroups of a group G. Define

$$[H, K] := \langle [h, k] \mid h \in H, k \in K \rangle.$$

Show both of the following:

(i) If $K \triangleleft G$ and $K \subset H$, then $[H, G] \subset K$ if and only if $H/K \subset Z(G/K)$.

(ii) If $\varphi : G \to L$ is a group epimorphism and $A \subset Z(G)$, then $\varphi(A) \subset Z(L)$.

6. Let G be group.
   (i) Set $\Gamma_1(G) = G$ and inductively define $\Gamma_{n+1}(G) := [\Gamma_n(G), G]$ for $n > 1$. Show $\Gamma_{n+1}(G) \subset \Gamma_n(G)$ and $\Gamma_n(G) \triangleleft\triangleleft G$ for all n.
   (ii) Set $Z^0(G) = 1$ and inductively define $Z^{n+1}(G)$ by
   $$Z^{n+1}(G)/Z^n(G) = Z(G/Z^n(G)),$$
   i.e., the preimage of $G/Z^n(G)$ in G under the canonical epimorphism. Show $Z^{n+1}(G) \triangleleft Z^n(G)$ and $Z^n(G) \triangleleft\triangleleft G$ for all $n > 0$.
7. Define the *descending central series* of G by
   $$G = \Gamma_1(G) \supset \Gamma_2(G) \supset \cdots \supset \Gamma_n(G) \supset \cdots$$
   and the *ascending central series* of G by
   $$1 = Z^0(G) \subset Z^1(G) \subset \cdots \subset \cdots .$$
   Prove that $Z^n(G) = G$ if and only if $\Gamma_n(G) = 1$. Moreover, if this is the case, then $\Gamma_{i+1} \subset Z^{n-i}$ for $i = 0, \ldots, n$.
8. A group G is called *nilpotent* if there exists an n such that $\Gamma_n(G) = 1$ (and n is called the class of G if n is the least such integer). Show the following
   (i) If G is nilpotent, then every subgroup of G is nilpotent.
   (ii) If G is nilpotent and $H \triangleleft G$, then G/H is nilpotent.
   (iii) A direct product of finitely many nilpotent groups is nilpotent.
   (iv) Give an example of a group G with $H \triangleleft G$ with H and G/H nilpotent, but G not nilpotent.
9. Show that a group G is nilpotent if and only if there exists a normal series
   $$G = G_0 \supset G_1 \supset \cdots \supset G_n = 1$$
   in which each $G_i \triangleleft G$ and satisfies $G_i/G_{+1} \subset Z(G/G_{i+1}$ for all i. Such a series is called a *central series* for G.
10. Prove that every finite group has a composition series. Give an example of an infinite group that does not have a composition series.
11. Let G be a group and N a normal subgroup of G. Prove that G has a composition series if and only if N and G/N have composition series. [It is not true that if G has a composition series that any subgroup does. Cf. Exercise 24.24(21).]

18. Free Groups

Let V be a vector space over F. We shall prove Proposition 28.6 below that says every vector space has a basis and any linearly independent subset in a vector space can be extended to a basis. Let V be a vector space over F with basis $\mathcal{B}$. Then $V, \mathcal{B}$ satisfies the following *universal property of vector spaces*: Given any vector space W over F and diagram of set maps

$$\begin{tikzcd} \mathcal{B} \arrow[r, hook, "inc"] \arrow[dr, "f"'] & V \\ & W \end{tikzcd}$$

with inc the inclusion of $\mathcal{B}$ into V, there exists a unique linear transformation $T : V \to W$ such that the diagram

$$\begin{tikzcd} \mathcal{B} \arrow[r, hook, "inc"] \arrow[dr, "f"'] & V \arrow[d, "T"] \\ & W \end{tikzcd}$$

commutes. This follows from the theorem in linear algebra that says a linear transformation $T : V \to W$ is completely determined by where a basis for V is mapped.

Conversely, suppose that we know that $\mathcal{B}$ is a subset of the vector space V and $V, \mathcal{B}$ satisfies the universal property above.

Let W be the subspace spanned by $\mathcal{B}$. The additive group V/W becomes a vector space by $r(v + W) := rv + W$, for $r \in F$ and $v \in V$, with $^- : V \to V/W$ the natural surjective linear transformation. By the (UP) there exists a unique transformation $T : V \to V$ satisfying

$$\begin{tikzcd} \mathcal{B} \arrow[r, hook, "inc"] \arrow[dr, hook, "0"'] & V \arrow[d, "T"] \\ & V/W. \end{tikzcd}$$

commutes. By uniqueness, T must be the zero map.

Let $v \in \mathcal{B}$ and

$$\begin{tikzcd} \mathcal{B} \arrow[r, hook, "inc"] \arrow[dr, "f_v"'] & V \\ & V \end{tikzcd}$$

with $f_v : \mathcal{B} \to V$ the map sending $v \mapsto v$ and $v' \mapsto 0$ for all $v' \in \mathcal{B} \setminus \{v\}$. Then there exists a unique linear transformation $T_v : V \to V$ such that

$$\begin{tikzcd} \mathcal{B} \arrow[r, hook, "inc"] \arrow[dr, "f_v"'] & V \arrow[d, "T_v"] \\ & V \end{tikzcd}$$

commutes. It follows that $\mathcal{B}$ is linearly independent, hence can be extended to a basis by Proposition 28.6. It follows by the universal property of vector spaces that $\mathcal{B}$ spans V, the details of which we leave as an exercise (Exercise 18.18(1)). In particular, V is completely determined by the universal property of vector spaces. Note that this says that V is completely determined by a set of generators satisfying no nontrivial relations in V.

We can, in fact, modify this universal property with the inclusion map replaced by a set injection $i_V : \mathcal{B}_0 \to V$ with $i_V(\mathcal{B}_0) = \mathcal{B}$. Indeed, we can even drop the requirement that i_V be an injection and this still shows that V is a vector space with basis $\mathcal{B}$. We leave this as an exercise. Therefore, we can reformulate the definition of the universal property of vector space as follows: Let V be a vector space over F and $i_{\mathcal{B}_0} : \mathcal{B}_0 \to V$ a set map. Then $i_{\mathcal{B}_0} : \mathcal{B}_0 \to V$ satisfies the universal property: Given any set map $i_{\mathcal{B}_0} : \mathcal{B}_0 \to W$ with W a vector space over F, there exists a unique linear transformation $T : V \to W$ satisfying

$$\begin{tikzcd} \mathcal{B}_0 \arrow[r, "i_{\mathcal{B}_0}"] \arrow[dr, "f"'] & V \arrow[d, "T"] \\ & W \end{tikzcd}$$

commutes. In this case, $\mathcal{B} = i_{\mathcal{B}_0}(\mathcal{B}_0)$ is a basis for V. Note that in this formulation, the universal property applies to the set map $i_{\mathcal{B}_0} : \mathcal{B}_0 \to V$.

We can also give an analogous definition of a free abelian group G on a basis X in a similar way as follows: Let G be an abelian group with X a nonempty set in G. We say that G is a *free abelian group* on *basis* X if it satisfies the following *universal property of free abelian groups*: Given any abelian group G' and diagram

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f"'] & G \\ & G' \end{tikzcd}$$

of set maps with *inc* the inclusion of X into G, then there exists a unique group homomorphism $\varphi : G \to G'$ such that the diagram

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f"'] & G \arrow[d, "\varphi"] \\ & G' \end{tikzcd}$$

commutes.

We now mimic the above for groups.

Definition 18.1. Let G be a group with X a nonempty set in G. We say that G is a *free group* on *basis* X if it satisfies the following *universal property of free groups*: Given any group G' and diagram

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f"'] & G \\ & G' \end{tikzcd}$$

of set maps with *inc* the inclusion of X into G, then there exists a unique group homomorphism $\varphi : G \to G'$ such that the diagram

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f"'] & G \arrow[d, "\varphi"] \\ & G' \end{tikzcd}$$

commutes.

Remark 18.2. As in the vector space case, we can replace the inclusion map of X in G with a set injection $i_{X_0} : X_0 \to G$ with $X = i_{X_0}(X_0)$; and, in fact, the map i_{X_0} need not be assumed to be an injection. Therefore, the alternative formulation of a free group is as follows: Let G be a group and $i_{X_0} : X_0 \to G$ a set map. Then $i_{X_0} : X_0 \to G$ satisfies the following universal property: Given any set map $f : X_0 \to G'$ with G' a group, there exists a unique group homomorphism $\varphi : G \to G'$ satisfying

$$\begin{tikzcd} X_0 \arrow[r, "i_{X_0}"] \arrow[dr, "f"'] & G \arrow[d, "\varphi"] \\ & G' \end{tikzcd}$$

commutes. Again in this formulation, the universal property applies to the set map $i_{X_0} : X_0 \to G$ and $X = i_{X_0}(X_0)$ is a basis for the free group G.

It shall be left as an exercise on the construction of free groups and the following lemma to show if G is a free group on basis X that there exists no nontrivial relations among the elements in X, i.e., the analogue of linear independence of vectors in a vector space holds, and X generates G.

We know if V and W are vector spaces over F with bases of the same cardinality, then they are isomorphic. The analogue for free groups holds.

Lemma 18.3. *Let X_0 and Y_0 be sets, A and B groups with set maps $i_A : X_0 \to A$ and $i_B : Y_0 \to B$ such that A is a free group on $X = i_A(X_0)$ and B is a free group of $Y = i_B(Y_0)$. Suppose that $|X_0| = |Y_0|$. Then A and B are isomorphic.*

Proof. Let $j : X_0 \to Y_0$ be a bijection. As A and B are free on bases X_0, Y_0, respectively, there exists unique group homomorphisms $\varphi : A \to B$ and $\psi : B \to A$ such that

$$\begin{tikzcd} X_0 \arrow[r, "j"] \arrow[d, "i_A"'] & Y_0 \arrow[d, "i_B"] \\ A \arrow[r, shift left, "\varphi"] & B \arrow[l, shift left, "\psi"] \end{tikzcd}$$

commutes. By the uniqueness in the universal property of free groups, $\psi \circ \varphi = 1_A$ and $\varphi \circ \psi = 1_B$. Therefore, φ and ψ are inverse isomorphisms. $\square$

We also know that if V is a finite dimensional vector space, then any two bases for V have the same cardinality. This is also true for infinite dimensional vector spaces using the Schroeder-Bernstein Theorem (cf. Remark 1.12) which we assume. We have seen that a finitely generated free abelian group group has a well-defined rank in Lemma 14.14

(and this also holds in the non-finitely generated case by an analogous argument). We, therefore, have the following proposition.

Proposition 18.4. *Let G be a free group on bases X and Y. Then $|X| = |Y|$. In particular, two free groups are isomorphic if and only if they have bases of the same cardinality.*

Proof. Let G be a free group on basis X and $^- : G \to G^{ab}$ be the canonical epimorphism where $G^{ab} = G/[G,G]$ is the abelianization of G (cf. Exercise 13.7(12)). The canonical map takes $X \mapsto \overline{X}$. Since the kernel of $^-$ is $[G,G]$ and G is free on X, we see that G^{ab} is free abelian on basis $\overline{X}$ and $|X| = |\overline{X}|$. The result follows by Remark 14.17. $\square$

We now prove the existence of free groups.

Theorem 18.5. *Let X be a nonempty set. Then there exists a free group G on X.*

Proof. Choose a set X' disjoint from X of the same cardinality, so we have a bijection $X \to X'$ which we denote by $x \mapsto x^{-1}$. Let X'' be a set disjoint from $X \cup X'$ having precisely one element. Let $X'' = \{e\}$. We call $A = X \cup X' \cup X''$ the *alphabet* and the elements of A the *letters*. If $x \in X$, we also write x^1 for x and if a is a letter, we also write $(a^{-1})^{-1}$ for a. Call a sequence $w = (a_1, \dots, a_n, \dots)$, $a_i \in A$ (not necessarily distinct), a *word* if $a_i \neq e$ for only finitely many i. Let $\widetilde{W}$ be the set of words in $\bigtimes_{i=1}^{\infty} A$. The word $(e, \dots, e, \dots)$ is called the *empty word*. A word $w = (a_1, \dots, a_n, \dots)$ in $\widetilde{W}$ is called a *reduced word* if w satisfies

(i) $a_i \neq a_{i+1}^{-1}$ if $a_i \neq e$.
(ii) If $a_i = e$, then $a_k = e$ for all $k \geq i$.

If w is a reduced word, but not the empty word, we can write it as

$$w = x_1^{n_1} \cdots x_m^{n_m},\ x_i \in X,\ \text{not necessarily distinct, } n_i \in \{\pm 1\} \text{ for all } i \text{ and some } m \text{ with}$$
$$x_i^{n_i} \neq x_{i+1}^{-n_{i+1}} \text{ for all } i.$$

We write e for the empty word. Let W be the set of reduced words.

Note. *Spelling* in W is unique by the definition of $\bigtimes_{i=1}^{\infty} A$.

Define $\cdot : W \times W \to W$ as follows: $w = e \cdot w = w \cdot e$ for any $w \in W$ and if $w = x_1^{n_1} \cdots x_r^{n_r}$ and $w' = y_1^{m_1} \cdots y_s^{m_s}$ with $x_i, y_j \in X$ (not necessarily distinct) and $n_i, m_j \in \{\pm 1\}$ for all i and j, let $w \cdot w' = x_1^{n_1} \cdots x_r^{n_r} y_1^{m_1} \cdots y_s^{m_s}$ by *juxtaposition* and then *reduce*, e.g., if $x_r^{n_r} = y_1^{-m_1}$ delete the pair unless reduced to the empty word and continue in this way to get a reduced word.

Claim 18.6. The $\cdot : W \times W \to W$ above is well-defined and makes W into a group such that the inclusion $X \subset W$ makes X into a free group on basis X, so $W = \langle X \rangle$.

The fact that $\cdot : W \times W \to W$ is a well defined map will follow from our construction which will also show that W satisfies associativity under this map, it is then clear that W is a group. To do this we use a trick of Van der Waerden. For each $x \in X$ and $n \in \{\pm 1\}$, define a map

$$|x^n| : W \to W \quad \text{by} \quad e \mapsto x^n \quad \text{and}$$
$$x_1^{n_1} \cdots x_r^{n_r} \mapsto \begin{cases} x^n x_1^{n_1} \cdots x_r^{n_r} \text{ if } \ x^n \neq x_1^{-n_1} \\ x_2^{n_2} \cdots x_r^{n_r} \text{ otherwise.} \end{cases}$$

Clearly,

$$|x^n||x^{-n}| = 1_W = |x^{-n}||x^n|,$$

so $|x^n|$ is a permutation of W, i.e., $|x^n| \in \Sigma(W)$ for all $x \in X$ and $n \in \{\pm 1\}$. Let

$$W_0 := \langle |x| \mid x \in X\rangle \subset \Sigma(W), \text{ a subgroup.}$$

As spelling is unique, we have a well-defined surjection

$$\varphi : W \to W_0 \text{ given by } e \mapsto 1_W \text{ and } x_1^{n_1} \cdots x_r^{n_r} \mapsto |x_1^{n_1}| \cdots |x_r^{n_r}|.$$

Let $z \in W_0$. Then there exist $x_i \in X$ and $n_i \in \{\pm 1\}$ such that $z = |x_1^{n_1}| \cdots |x_r^{n_r}|$. If $x_i^{n_i} = x_{i+1}^{-n_{i+1}}$, then $|x_i^{n_1}||x_{i+1}^{n_{i+1}}| = 1_W$ and we can cancel them (or reduce to 1_W), which we may assume always has been done. Since W_0 is a group, the result is always unique. But then

$$\psi : W_0 \to W \text{ given by } 1_W \mapsto e \text{ and } |x_1^{n_1}| \cdots |x_r^{n_r}| \mapsto |x_1^{n_1}| \cdots |x_r^{n_r}|(e)$$

is well-defined as $|x_1^{n_1}| \cdots |x_r^{n_r}| = |x_1^{n_1} \cdots x_r^{n_r}|$ and spelling is unique in W. In particular, $\psi \circ \varphi = e$ and $\varphi \circ \psi = 1_W$. It follows that φ is a bijection. By construction

$$\varphi(w_1 \cdot w_2) = \varphi(w_1)\varphi(w_2) \text{ for all } c_1, w_2 \in W.$$

As W_0 is a group, it follows that $\cdot : W \times W \to W$ above is well-defined and as associativity holds in W_0, associativity holds in W.

We leave it as an exercise to show that W is a free group on basis X, i.e., satisfies the universal property of free groups. □

Corollary 18.7. *Let G be a free group on the set X. Then $G = \langle X\rangle$ and there exist no nontrivial relations among the elements of X.*

We leave the proof as an exercise.

Definition 18.8. Let H be a subgroup of a group G. The *normal closure* of H in G is the smallest normal subgroup of G containing H, i.e., the group $\langle xHx^{-1} \mid x \in G\rangle$.

Definition 18.9. A group G is said be defined by *generators* $X = \{x_i \mid i \in I\}$ and *relations* $Y = \{y_j \mid j \in J\}$ if $G \cong E/N$ where E is a free group on basis X, $Y \subset E$, and N is the normal closure of $\langle Y\rangle$ in E. If this is the case, we say (X, Y) is a *presentation* and also (by abuse of notation) write $G = \langle X \mid Y\rangle$.

Corollary 18.10. *Every group has a presentation, unique up to isomorphism.*

PROOF. Let G be a group and X a subset of G that generates G (e.g., we can take G itself). Let E be the free group on basis X. By the universal property of free groups, there exists a unique group homomorphism $\varphi : E \to G$ such that

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, hook, "inc"'] & E \arrow[d, "\varphi"] \\ & G \end{tikzcd}$$

commutes. As $G =< X >$, the homomorphism φ is onto. Let Y be any subset of $\ker \varphi$ such that $\ker \varphi$ is the normal closure of $< Y >$. Then (X, Y) is a presentation of G. We leave the proof of uniqueness as an exercise. $\square$

Corollary 18.11. (Van Dyck's Theorem) *Let X be a nonempty set and Y a set of reduced words based on X (obvious definition). Let $G = \langle X \mid Y \rangle$ be a presentation of G. Suppose that H is a group satisfying $H = \langle X \rangle$ and H satisfies all the relations in Y. Then there exists a (group) epimorphism $\psi : G \to H$.*

PROOF. Let E be the free group of X and N the normal closure of $< Y >$ in E. As

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, hook, "inc"'] & E \\ & H, \end{tikzcd}$$

by the universal property of free groups, there exist unique group homomorphisms

$$\begin{tikzcd} E \arrow[r, "\theta"] \arrow[dr, "\varphi"'] & G \\ & H. \end{tikzcd}$$

We define $\psi : G \to H$ as follows: For each $g \in G$ choose x_g in the fiber $\theta^{-1}(g) := \{y \in E \mid \theta(y) = g\}$ and set $\psi(g) := \varphi(x_g)$. We must show that ψ is well-defined, i.e., independent of the choice of $x_g \in \theta^{-1}(g)$. Suppose that x, y in E satisfies $\theta(x) = g = \theta(y)$. Then $\theta(y^{-1}x) = e_G$, so $y^{-1}x$ lies in $\ker \theta = N$, the normal closure of $< Y >$ in E. If $H =< X|Y' >$, then N lies in the normal closure $\ker \varphi = N'$ of $< Y' >$ in E by hypothesis. Consequently, $\ker \theta \subset \ker \varphi$, so $\varphi(x) = \varphi(y)$ and ψ if well defined. Clearly ψ is a homomorphism and surjective as both θ and φ are. $\square$

Examples 18.12. Let $Q =< a, b \mid a^2 = b^2, a^4 = e, bab^{-1} = a^{-1} >$ be the quaternion group and G be the subgroup of $\mathbb{M}_2(\mathbb{C})$ generated by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$

We show $Q \cong G$. Using the Van Dyck's Theorem, we get a group epimorphism $Q \to G$ given by

$$a \mapsto \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix} \quad \text{and} \quad b \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Check that $|G| \geq 8$. The relations for G imply that $ba = a^3b$ and $b^3 = a^2b$, so we must have $|Q| \leq 8$. It follows that $|Q| = 8$ and $G \cong Q$.

We end this section with a further useful example of universal properties.

Definition 18.13. Let G_1 and G_2 be two groups. A *free product* of G_1 and G_2 is a group E together with group monomorphisms

(*)
$$\begin{tikzcd} G_1 \arrow[dr, "\theta_1"] & \\ & E \\ G_2 \arrow[ur, "\theta_2"'] & \end{tikzcd}$$

satisfying the following universal property: If H is a group and we group homomorphisms

$$\begin{tikzcd} G_1 \arrow[dr, "\psi_1"] & \\ & H \\ G_2 \arrow[ur, "\psi_2"'] & \end{tikzcd}$$

then there exists a unique group homomorphism $\varphi : E \to H$ such that the diagram

$$\begin{tikzcd} G_1 \arrow[dr, "\theta_1"'] \arrow[drr, "\psi_1"] & & \\ & E \arrow[r, "\varphi"] & H \\ G_2 \arrow[ur, "\theta_2"] \arrow[urr, "\psi_2"'] & & \end{tikzcd}$$

commutes.

Theorem 18.14. *If G_1 and G_2 are groups, then a free product of G_1 and G_2 exists.*

In fact if we fix the monomorphisms in (*), then the fiber product is unique up to a unique isomorphism and it is usually denoted by $G_1 * G_2$.

We leave the proof as an exercise. It is quite similar to the existence of a free group defining an alphabet in this case as follows: Assume that $G_1 \setminus \{e_{G_1}\}$ and $G_2 \setminus \{e_{G_2}\}$ are disjoint. Let $\{e\}$ be a set disjoint from $(G_1 \setminus \{e_{G_1}\}) \cup (G_2 \setminus \{e_{G_2}\})$ and set $A = (G_1 \setminus \{e_{G_1}\}) \cup (G_2 \setminus \{e_{G_2}\}) \cup \{e\}$. Then define words in the obvious way, calling a nonempty word w reduced if $w = a_1 \cdots a_r$ with $a_i \in A \setminus \{e\}$ for $i = 1, \dots, r$, some r and no adjacent letters lie in the same G_i, $i = 1, 2$. Now proceed as before.

Of course, it is clear that the coproduct of $G_1, \dots, G_n$ exists, but so does G_i, $i \in I$, a well ordered set using the Axiom of Choice and its equivalences. For example, $*_I$ is isomorphic to the free group on basis I. See the exercises.

More generally, we have

Definition 18.15. Suppose that G_1, G_2 and H are groups. A *free product* of G_1 and G_2 with *amalgamation* H is a group $G_1 *_H G_2$ together with group homomorphisms

$$\begin{tikzcd}
 & G_1 \arrow[dr, "\theta_1"] & \\
H \arrow[ur, "\varphi_1"] \arrow[dr, "\varphi_2"'] & & G_1 *_H G_2 \\
 & G_2 \arrow[ur, "\theta_2"'] &
\end{tikzcd}$$

with φ_i, $i = 1, 2$, monomorphisms
satisfying the following universal property: If G is a group and we have group homomorphisms

(*)

$$\begin{tikzcd}
 & G_1 \arrow[dr, "\rho_1"] & \\
H \arrow[ur, "\varphi_1"] \arrow[dr, "\varphi_2"'] & & G \\
 & G_2 \arrow[ur, "\rho_2"'] &
\end{tikzcd}$$

then there exists a unique group homomorphism $\mu : G_1 *_H G_2 \to G$ such that the diagram

$$\begin{tikzcd}
 & G_1 \arrow[dr, "\theta_1"'] \arrow[rrr, "\psi_1"] & & & G \\
H \arrow[ur, "\varphi_1"] \arrow[dr, "\varphi_2"'] & & G_1 *_H G_2 \arrow[urr, "\varphi"] & & \\
 & G_2 \arrow[ur, "\theta_2"] \arrow[uurrr, "\psi_2"'] & & &
\end{tikzcd}$$

commutes.

That such an free product with amalgamation exists is left as an exercise.

Free products with amalgamation have a nice application in topology. We first need some definitions.

Definition 18.16. Let X and Y be a topological space and $[0, 1]$ the unit interval in $\mathbb{R}$.

If $f, g : X \to Y$ are continuous maps, a *homotopy* $H : X \times [0, 1] \to Y$ between f and g at x is a continuous map satisfying $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$. We say that f and g are *homotopic* and write $f \approx g$. The relation $\approx$ is an equivalence relation. Moreover, if $f_1, g_1 : X \to Y$ are homotopic, and $f_2, g_2 : Y \to Z$ are homotopic, then their compositions $f_2 \circ f_1$ and $g_2 \circ g_1 : X \to Z$ are homotopic.

If $x_0 \in X$, a *loop* at x_0 is a continuous map $f : [0, 1] \to X$ satisfying $f(0) = x_0 = f(1)$, i.e., a continuous curve $f : [0, 1] \to \mathbb{R}$ with the same starting and ending point x_0. Let $\mathcal{L}(X, x_0)$ be the set of all loops at x_0. Define a *homotopy of loops* f and g for

$f, g \in \mathcal{L}(X, x_0)$ to be a continuous map $H : [0,1] \times [0,1] \to X$ satisfying

$$\begin{gathered} H(x,0) = f(x) \ \text{ and } \ H(x,1) = g(x) \ \text{ for all } \ x \in [0,1] \\ H(0,t) = x_0 = H(1,t) \ \text{ for all } \ t \in [0,1]. \end{gathered}$$

This defines an equivalence relation $\sim$ on $\mathcal{L}(X, x_0)$. The set of equivalence classes in $\mathcal{L}(X, x_0)$ can be given a group structure as follows: Define $f \star g : [0,1] \to X$ for $f, g \in \mathcal{L}(X, x_0)$ by

$$\begin{aligned} (f \star g)(t) &:= f(2t) && \text{for } \ t \in [0, 1/2] \\ (f \star g)(t) &:= f(2t-1) && \text{for } \ t \in [1/2, 1], \end{aligned}$$

and define a binary operation $*$ on $\overline{\mathcal{L}(X, x_0)} = \mathcal{L}(X, x_0)/\sim$ by

$$\overline{f} * \overline{g} := \overline{f \star g}.$$

The operation $*$ is well-defined and the quotient $\mathcal{L}(X, x_0)/\sim$ is called the *fundamental group* of X based at x_0 and denoted by $\pi_1(X, x_0)$.

The application in topology is the following (which we do not prove):

Theorem 18.17. (van Kampen) *Suppose that X is a topological space, U_1, U_2, $U_1 \cap U_2 \subset X$ all path connected subspaces, i.e., any two points can be connected by a continuous curve, with $U_1 \cap U_2 \neq \emptyset$. If $X = U_1 \cup U_2$ and $x_0 \in U_1 \cap U_2$, then*

$$\pi_1(X, x_0) = \pi_1(U_1, x_0) *_{\pi_1(U_1 \cap U_2, x_0)} \pi_1(U_2, x_0).$$

An application shows that the figure eight group defined in Example 8.4(9) is isomorphic to the free group $\mathbb{Z} * \mathbb{Z}$.

Exercises 18.18.

1. Show that if a vector space V over F satisfies the universal property of vector spaces for a subset $\mathcal{B}$, then $\mathcal{B}$ spans V and is linearly independent. In particular, 0 cannot lie in $\mathcal{B}$ unless $V = 0$
2. Let V be a vector space over F. Suppose that $\mathcal{B}_0$ is a set and $i_V : \mathcal{B}_0 \to V$ a set map. Show if for every vector space W over F and set map $f : \mathcal{B}_0 \to W$ there exists a unique linear transformation $T : V \to W$ such that

$$\begin{tikzcd} \mathcal{B}_0 \arrow[r, "i_V"] \arrow[dr, "f_v"'] & V \arrow[d, "T"] \\ & W \end{tikzcd}$$

   commutes, then i_V is an injective map and $i_V(\mathcal{B}_0)$ is a basis for V.
3. Prove the part of Claim 18.6 left undone.
4. Prove Corollary 18.7.
5. Prove that a presentation for a group is unique up to isomorphism.

6. Define a *free abelian group* A on basis X to be an abelian group satisfying the following universal property of free abelian groups: Given any abelian group B and diagram

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f_v"'] & A \\ & B \end{tikzcd}$$

of set maps with inc the inclusion of X in A, there exists a unique group homomorphism $\varphi : A \to B$ such that

$$\begin{tikzcd} X \arrow[r, hook, "inc"] \arrow[dr, "f"'] & A \arrow[d, "\varphi"] \\ & B \end{tikzcd}$$

commutes. Prove that a free abelian group on any nonempty set X exists and is unique up to isomorphism.

7. Show that $\mathbb{Z}$ is a free group on basis $\{1\}$.
8. Show that if G is a free group on basis $\mathcal{B}$ and $\mathcal{B}$ has more than two elements that G is not abelian.
9. Define the free product of a set of groups $\{G_i \mid i \in I\}$ and prove it exists. It is denoted by $*_I G_i$.
10. Show that a free group on a set X is isomorphic to $*_X \mathbb{Z}$.
11. Show that a free abelian group on a finite set X is isomorphic to $\bigtimes_X \mathbb{Z}$. [What if X is not finite?]
12. Let p be a prime and G the group $G = \langle x, y \mid x^p = y^p = (xy)^p = 1\rangle$. Prove that G is isomorphic to the Klein four group if $p = 2$ and is infinite if p is odd.
13. Define the infinite dihedral group D_∞ to be the group
$$\langle r, f \mid f^2 = e,\ f^{-1}rf = r^{-1}\rangle.$$
Prove that $D_\infty \cong \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$.
14. Let $\mathrm{PSL}_2(\mathbb{Z}) := \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$. Show the following:
   (*i*) There is a group homomorphism $\varphi : \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z} \to \mathrm{PSL}_2(\mathbb{Z})$.
   (*ii*) The homomorphism φ is an isomorphism.
15. A group is called *finitely presented* if it has a presentation $\langle X \mid R\rangle$ in which both X and R are finite. Let G be a group with $N \lhd G$. Prove that G is finitely presented if and only if N and G/N are both finitely presented.
16. Prove that any finitely generated group that has a presentation with n generators and r relations with $r < n$ is an infinite group.
17. Prove that free products with amalgamation 18.15 exist.
18. Prove the assertions in Definition 18.16.
19. Prove that the figure eight group is isomorphic to $\mathbb{Z} * \mathbb{Z}$ using van Kampen's Theorem.

CHAPTER IV

Group Actions

In general, groups arise in mathematics not abstractly, but in a very concrete way. The simplest geometric example is given an object, what are all the symmetries of that object, e.g., if the object is a circle, a sphere, a tetrahedron or cube in $\mathbb{R}^3$? Of course, we also have already seen examples in the previous chapter. The dihedral group acts on a regular n-gon, the symmetric group acts on a set. These are all examples of a given set upon which a collection of maps acts on it bijectively. This set of maps of the object, can give us much deeper knowledge about that object. As our examples above are maps, the composition of bijections satisfies associativity and the identity map acts as the identity. This is already sufficient to give us our desired abstract formulation. In this chapter, we give many examples of this, and show how powerful an idea it is. Again we concentrate on applications to finite groups, since using the natural equivalence relation that arises allows us to count, resulting in many deep theorems about the structure of such groups.

19. The Orbit Decomposition Theorem

Groups are important because they act on objects. For example,

$\Sigma(S)$ acts on S
$\mathrm{GL}_n(F)$ acts on F^n
D_n acts on the vertices of a regular n-gon,

and these actions give information about the groups and the objects on which they act. We begin by defining how a group G acts on a nonempty set S. We then see how this leads to an equivalence relation $\sim_G$ on S and form the set of equivalence classes $S/\sim_G$, followed by the application of the Mantra of Equivalence Relations. To be useful, especially when $|S|$ is finite, we must compute the cardinality of each equivalence class. Unlike the case of cosets, we cannot expect the cardinality of each equivalence class to be the same.

We begin with the definition of an action.

Definition 19.1. Let G be a group and S a nonempty set. We say that S is a *(left) G-set* under $* : G \times S \to S$, writing $g * s$ for $*(g, s)$, if $*$ satisfies the following: For all $g_1, g_2 \in G$ and for all $s \in S$,

(1) $(g_1 \cdot g_2) * s = g_1 * (g_2 * s)$.
(2) $e * s = s$.

The map $* : G \times S \to S$ is then called a *G-action* on S. Note that the first property is just a generalization of associativity, the second that the unity acts like an identity.

Note. If S is a G-set and H a subgroup of G, then we can also view S as an H-set by *restriction*, i.e., by the H-action $\star : H \times S \to S$ given by $h \star s := h * s$.

We now define the equivalence relation. Let S be a G-set under $*$. Define $\sim_G$ on S by

$$s_1 \sim_G s_2 \text{ if there exists } g \in G \text{ satisfying } s_1 = g * s_2. \tag{19.2}$$

Lemma 19.3. *Let S be a G-set under $*$. Then $\sim_G$ is an equivalence relation on S.*

PROOF. **Reflexitivity**: For all $s \in S$, we have $s = e * s$, so $s \sim_G s$.
Symmetry: If $s_1 \sim_G s_2$, then there exists a $g \in G$ such that $s_1 = g * s_2$, hence

$$g^{-1} * s_1 = g^{-1} * (g * s_2) = (g^{-1} \cdot g) * s_2 = e * s_2 = s_2.$$

Note this shows

$$s_1 = g * s_2 \text{ if and only if } g^{-1} * s_1 = s_2. \tag{19.4}$$

Transitivity: Suppose that $s_1 \sim_G s_2$ and $s_2 \sim_G s_3$. Then there exist $g, g' \in G$ satisfying $s_1 = g * s_2$ and $s_2 = g' * s_3$. Hence $s_1 = g * s_2 = g * (g' * s_3) = (g \cdot g') * s_3$, so $s_1 \sim_G s_3$. □

Next, let S be a G-set, $s \in S$. The equivalence class of s via $\sim_G$ is the set

$$\begin{aligned}\overline{s} &:= \{s' \in S \mid \text{there exists } g \in G \text{ such that } s' = g * s\} \\ &= \{g * s \mid g \in G\}.\end{aligned}$$

This equivalence class is called the *orbit* of s under $*$ and denoted by $G * s$. We always let

$$\mathcal{O} \text{ be a system of representatives for the equivalence classes under } \sim_G .$$

(So $\mathcal{O}$ consists of precisely one element from each orbit — equivalence class.) We write $G\backslash S$ for $S/\sim_G = \{G * s \mid s \in \mathcal{O}\}$, the set of orbits for the action $\sim_G$. Applying the Mantra of Equivalence Relations, we have

Mantra of G-actions. Let S be a G-set, $\mathcal{O}$ a system of representatives. Then

$$S = \bigvee_{\mathcal{O}} G * s \text{ and if } |S| \text{ is finite, then } |S| = \sum_{\mathcal{O}} |G * s|.$$

To make this useful, we must be able to compute the size of an orbit if finite. In general, orbits may have different cardinalities. We do this by attaching a group to each element of S in the following way. If S is a G-set, $s \in S$, define the *stabilizer* or *isotropy subgroup* of s by

$$G_s := \{x \in G \mid x * s = s\}.$$

Lemma 19.5. *Let S be a G-set, $s \in S$. Then G_s is a subgroup of G.*

PROOF. By definition of a G-action, $e \in G_s$, so G_s is nonempty. If $x, y \in G$, then $(xy) * s = x * (y * s) = x * s = s$ and if $y * s = s$, then $s = y^{-1} * s$, so G_s is a subgroup. □

We can now compute the cardinality of an orbit.

Proposition 19.6. *Let S be a G-set, $s \in S$. Define*

$$f_s : G/G_s \to G * s \text{ by } xG_s \mapsto x * s.$$

Then f_s is a well-defined bijection. In particular, if $[G : G_s]$ is finite, then

$$|G * s| = [G : G_s] \text{ and } |G * s| \text{ divides } |G|.$$

Proof. Let $x, y \in G$. Then

$$\begin{aligned} x * s = y * s \text{ if and only if } & y^{-1} * (x * s) = s \\ \text{if and only if } & (y^{-1}x) * s = s \\ \text{if and only if } & y^{-1}x \in G_s \\ \text{if and only if } & xG_s = yG_s. \end{aligned}$$

This shows that f_s is well-defined and one-to-one. As f_s is clearly surjective, the result follows. □

Example 19.7. Let S be the faces of a cube and G be the group of rotations of S. So G acts on S. Given any two faces s_1, s_2, there is an element of G taking s_1 to s_2. So there is one orbit under this action. We say that G acts *transitively* on S. If s is a face, then the isotropy subgroup G_s of s is the cyclic group C_4 of rotations of the face s about its center. By the proposition, we have

$$|G|/|C_4| = [G : G_s] = |G * s| = |S| = 6,$$

so $|G| = 24$. More generally, let S be the regular solid having n faces, each which has k edges (or vertices) and G the group of all rotations of S. Then the G-action is transitive and $|G| = nk$ by an analogous argument. It can be shown that there are only five such regular solids: the tetrahedron ($n = 4$, $k = 3$), the *cube*, the *octahedron* ($n = 8$, $k = 3$), the *dodecahedron* ($n = 12, k = 5$), and the *icosahedron* ($n = 20$, $k = 3$). So the corresponding rotation groups have 12, 24, 24, 60, 60 elements and are isomorphic to A_4, S_4, S_4, A_5, A_5 respectively. We give further details in Section 20.

Suppose that S is a G-set, $s \in S$. We say that $G * s$ is a *one point orbit* of S and s *is a fixed point* (under the action of G) if $G * s = \{s\}$. We set

$$F_G(S) := \{s \in S \mid |G * s| = 1\} \subset S,$$

the *set of fixed points* of S under the action of G.

If S is a G-set, a major problem is to determine if $F_G(S)$ is nonempty, and if it is, to compute $|F_G(S)|$.

The following equivalent conditions characterize the one point orbits. The proof is left as an easy exercise.

Lemma 19.8. *Let S be a G-set, $s \in S$. Then the following are equivalent:*

(1) $s \in F_G(S)$.
(2) $G_s = G$.
(3) $G * s = \{s\}$.

In particular, if $\mathcal{O}$ is a system of representative, then $F_G(S) \subset \mathcal{O}$.

If $\mathcal{O}$ is a system of representatives for a G-action on a set S, we always set

$$\mathcal{O}^* = \mathcal{O} \setminus F_G(S).$$

In particular, if $s \in \mathcal{O}^*$, we have $|G * s| = [G : G_s] > 1$, i.e., $G_s < G$. (Recall that the symbol $<$ when used for sets means a subset, but not the whole set.)

Putting this all together, we get the desired result:

Theorem 19.9. (Orbit Decomposition Theorem) *Let S be a G-set. Then*

$$S = F_G(S) \vee \bigvee_{\mathcal{O}^*} G * s.$$

In particular, if S is a finite set, then

$$|S| = \sum_{\mathcal{O}} |G * s| = |F_G(S)| + \sum_{\mathcal{O}^*} [G : G_s].$$

[Note *if $s \in \mathcal{O}^*$, then $1 < [G : G_s]$ and if G is finite then $[G : G_s] \mid |G|$.*]

PROOF. This follows from the Mantra as $|G * s| = [G : G_s]$. □

Exercises 19.10.

1. Let S be a G-set, $s_1, s_2 \in S$. Suppose that there exists an $x \in G$ such that $s_2 = x * s_1$, i.e., $s_2 \in G * s_1$. Show that $G_{s_2} = xG_{s_1}x^{-1}$.
2. Let S be a G-set. We say that the G-action is *transitive* if for all $s_1, s_2 \in S$, there exists a $g \in G$ satisfying $g * s_1 = s_2$, equivalently, $S = G * s$; and is *doubly transitive* if for all pairs of elements (s_1, s_1') and (s_2, s_2') in S with $s_1 \neq s_1'$ and $s_2 \neq s_2'$, there exists a $g \in G$ satisfying $g * s_1 = s_2$ and $g * s_1' = s_2'$. Suppose that S and G are finite. Show if the action is transitive then $|G| \geq |S|$ and if the action is doubly transitive then $|G| \geq |S|^2 - |S|$.
3. Let G be a group and H a subgroup of G. Show that the set of cosets G/H, so the quotient (set) of G mod H is a G-set by $G \times G/H \to G/H$ via $(g, kH) \mapsto gkH$.
4. Let G be a group. Prove that there exists a bijection between the following two sets:
   (i) The subgroups of G.
   (ii) The cosets (quotient sets) of the G-set G.
   This bijection establishes a correspondence between a subgroup H in G with the quotient G-set G/H, i.e., the cosets of H in G.
5. Let G be a group and X and Y G-sets. A set map $f : X \to Y$ is called a *G-set map* or a *G-equivariant* if $\varphi(gx) = g\varphi(x)$ for all $g \in G$ and $x \in X$. Show X is a disjoint union of quotients of the G-set G. Moreover, if X is finite, then this disjoint union is finite.
6. Let G be a group and H and K be subgroups. Suppose that $\varphi : G/H \to G/K$ is a G-equivariant map. Then there exists an element $x \in G$ satisfying $\varphi = \lambda_x$ (left multiplication by x), $xgK = gxK$, and $x^{-1}Hx \subset K$ for all $g \in G$.

20. Addendum: Finite Rotation Groups in $\mathbb{R}^3$.

In this addendum, we give further details to Example 19.7. Recall if R is a commutative ring, then the *orthogonal group* is $\mathrm{O}_n(R) := \{A \in \mathrm{GL}_n(R) \mid AA^t = I\}$ and the *special orthogonal group* is $\mathrm{SO}_n(R) = \mathrm{O}_n(R) \cap \mathrm{SL}_n(R)$. In this section, we shall determine the finite subgroups of $\mathrm{SO}_3(\mathbb{R})$. Elements of $\mathrm{SO}_3(\mathbb{R})$ are very special as any element is, in fact, a rotation about some axis through the origin in $\mathbb{R}^3$. This is easy to see using linear algebra. Indeed let A be a nonidentity element of $\mathrm{SO}_3(\mathbb{R})$. Viewing $\mathbb{R}^3$ as an inner product space under the dot product $\cdot$ the matrix A is an isometry, i.e., A preserves dot products as $Av \cdot Aw = v \cdot A^tAw = v \cdot w$ for all v and w in $\mathbb{R}^3$ and also in $\mathbb{C}^3$. It follows that every eigenvalue ε of A in $\mathbb{C}$ satisfies $|\varepsilon| = 1$. In particular, ε is a root of unity, so is

± 1 if real. As the characteristic polynomial f_A of A is of odd degree, it has a real root, hence an eigenvalue 1 or -1. If all the roots of f_A are real, $\det A = 1$ forces A to be similar to $\operatorname{diag}(1, -1, -1)$, a rotation of angle π around some axis. If the roots of f_A are not all real, they must be $1, e^{2\pi\sqrt{-1}/n}, e^{-2\pi\sqrt{-1}/n}$, some n, and A is similar to the rotation $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/n & \sin 2\pi/n \\ 0 & -\sin 2\pi/n & \cos 2\pi/n \end{pmatrix}$ about some axis.

We view $\mathrm{SO}_3(\mathbb{R})$ acting on the unit sphere S^2. So if A is a nonidentity element in $\mathrm{SO}_3(\mathbb{R})$, it is the rotation by some angle about some axis through the origin. This axis intersects S^2 in two points, that we call the *poles* of A. If A is of finite order, then the angle of rotation must be $2\pi/n$ for some integer n. Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n, and set $P := \{p \mid p \text{ is a pole of some element } A \text{ in } G\}$, called the *set of poles* of G.

Lemma 20.1. *Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n, and P the set of poles of G. Then P is a G-set under the natural action. In particular, if $p \in P$, $|G| = [G : G_p]|G_p|$ and $|G * p| = [G : G_p]$, where G_p is the isotropy subgroup of p and $G * p$ the orbit of p.*

PROOF. Let p be a pole of A in G. If B is a matrix in G, then $(BAB^{-1})Bp = Bp$, so Bp is a pole of BAB^{-1}. $\square$

Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n and P the set of poles. If p is a pole, let $n_p = |G_p|$. As every nonidentity element of G has two poles and each nonidentity element in G_p has p as a pole, we have

$$2(n-1) = \sum_P (n_p - 1).$$

Let $G * p_1, \dots, G * p_r$ be the (distinct) orbits of P and $n_i = n_{p_i}$, for $i = 1, \dots, r$. As the order of stabilizers of elements in a fixed orbit are all the same, we have

$$2n - 2 = \sum_{i=1}^{r} (n_i - 1)|G * p_i|,$$

so dividing by $n = |G|$ implies that

$$2 - \frac{2}{n} = \sum_{i=1}^{r} (n_i - 1)\frac{|G * p_i|}{n} = \sum_{i=1}^{r} (n_i - 1)\frac{1}{|G_{p_i}|} = \sum_{i=1}^{r} (1 - \frac{1}{n_i}).$$

As $n > 1$ and $1 - (1/n_i) \geq 1/2$, we have $1 \leq 2 - (2/n) \leq 2$. It follows that we can only have $r = 1, 2$, or 3, i.e., there are at most three orbits. We investigate each possibility separately.

Case. $r = 1$:

This cannot occur, as $1 \leq 2 = (2/n) = 1 - (1/n_1)$ is impossible.

Case. $r = 2$:

As

$$2 - \frac{2}{n} = (1 - \frac{1}{n_1}) + (1 - \frac{1}{n_2}),$$

we have

$$\frac{2}{n} = \frac{1}{n_1} + \frac{1}{n_2}.$$

It follows that $n = n_1 = n_2$ and $|G * p_1| = |G * p_2| = 1$ as $n = n_i|G * p_i|$. In particular, $|P| = 2$. Since every nonzero element in G has two poles, we must have G is a cyclic group of finite order with axis connecting the two poles.

Case. $r = 3$:

In this case, we have

$$(*) \qquad \frac{2}{n} = \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3},$$

and we may assume that $n_1 \leq n_2 \leq n_3$. We must find the groups with data (n_1, n_2, n_3). Let $[|G * p_1|, |G * p_2|, |G * p_3|] = [n/n_1, n/n_2, n/n_3]$ be the tuple of the sizes of the corresponding orbits. If $n_1 \geq 3$, then $2/n \leq 0$, which is impossible, so $n_1 = 2$, and

$$\frac{1}{2} + \frac{2}{n} = \frac{1}{n_2} + \frac{1}{n_2}.$$

This is impossible if $n_2 \geq 4$, so $n_2 = 2$ or 3.

Subcase. $r = 3, n_2 = 2$:

By (*), we must have $n_3 = n/2$. Therefore, G_{p_3} has two poles so is cyclic of order $n/2$. As P is a G-set and the nonzero element g in G_{p_1} must be a rotation of angle π, it follows that the poles of g must be perpendicular to the poles of the nonzero elements in $G * p_3$ (and reverse pole pairs). It follows that G must be the dihedral group $D_{\frac{n}{2}}$.

Therefore we may assume that $n_2 = 3$. If $n_3 \geq 6$, then $(1/2) + (1/3) + (1/n_3) \leq 0$, so $n_3 = 3, 4$, or 5.

Subcase. G has data $(2, 3, 3)$:

It follows by (*) that $n = 12$, so G also has orbit data $[6, 4, 4]$. As P is a G-group, if p lies in $G * p_3$ and q is the closest point to p in $G * p_2$, then $G_{p_3}q$ consists of at least three points equidistance from p, which form an equilateral triangle. It follows that there are four such triangles with the poles forming a regular tetrahedron. Therefore, we see that G is the rotation group of a regular tetrahedron and the orbit data $[6, 4, 4]$ is the number of edges, vertices, and faces of the tetrahedron, respectively. (Cf. Example 19.7.) This group can be shown to be isomorphic to A_4 using Exercise 24.24(9) below.

Subcase. G has data $(2, 3, 3)$:

It follows by (*) that $n = 24$, so G also has orbit data $[12, 8, 6]$. In this case, if p lies in $G * p_3$ and q is the closest point to p in $G * p_2$, then $G_{p_3}q$ consists of at least four points equidistance from p, which form a square, and we see that G is the group of rotations of a cube with orbit data $[12, 8, 6]$, the number of edges, vertices, and faces of a cube, respectively. (Cf. Example 19.7.) This group can be shown to be isomorphic to S_4 using Exercise 24.24(9) below.

Subcase. G has data $(2, 3, 5)$:

It follows by (*) that $n = 60$, so G also has orbit data $[30, 20, 12]$. In this case, if p lies in $G * p_3$ and q is the closest point to p in $G * p_2$, then $G_{p_3}q$ consists of at least five points equidistance from p, which form a pentagon, and we see that G is the group of rotations of a regular icosahedron with orbit data $[30, 20, 12]$, the number of edges, vertices, and faces of a regular icosahedron, respectively. (Cf. Example 19.7.) This group can be shown to be isomorphic to A_5 using Exercise 24.24(9) below.

21. Examples of Group Actions

In this section, to demonstrate the usefulness of this concept, we give many examples of group actions. As this is the most important concept in our study of groups, there are many exercises at the end of this section. To understand this topic, one should do many of them. In the following two sections, we shall give further specific applications of some of examples given in this section.

Example 21.1. Conjugation on Elements.

In this example, we let the G-set S be G itself. The (left) action is given by

$$* : G \times S \to S \quad \text{by} \quad g * s = gsg^{-1}$$

called *conjugation* by G. The orbit of an element $a \in S = G$ is

$$C(a) := G * a = \{xax^{-1} \mid x \in G\},$$

called the *conjugacy class* of a, and the stabilizer of a is

$$Z_G(a) := G_a = \{x \in G \mid xax^{-1} = a\} = \{x \in G \mid xa = ax\}.$$

This subgroup of G is called the *centralizer* of a. So $Z_G(a)$ is the set of elements of G commuting with a. In particular, $\langle a \rangle \subset Z_G(a)$. The set of fixed points of this action is

$$\begin{aligned} F_G(S) &= \{a \in S \mid xax^{-1} = a \text{ for all } x \in G\} \\ &= \{a \in G \mid xa = ax \text{ for all } x \in G\}, \end{aligned} \tag{21.2}$$

the *center* $Z(G)$ of G. Recall it is a normal subgroup of G. Of course, in general, a G-set is not a group, so one cannot expect the set of fixed points to have an algebraic structure.

Let $\mathcal{C}$ be a system of representatives for the conjugation action of G on G. So $\mathcal{C}^* = \mathcal{C} \setminus Z(G)$, and the Mantra of G-actions becomes

$$G = Z(G) \vee \bigvee_{\mathcal{C}^*} C(a),$$

with

$$|G| = |Z(G)| + \sum_{\mathcal{C}^*} [G : Z_G(a)], \tag{21.3}$$

if G is finite. Equation (21.3) is called the *class equation*. We give an interesting application of the class equation. If G is a finite group of order p^n, for some (positive) prime p and $n > 0$, we call G a *p-group*. So a subgroup of a p-group is either a p-group or the trivial group.

Application 21.4. If G is a p-group, then $1 < Z(G) \subset G$.
(Recall this means $1 \neq Z(G)$.) In particular, if G is a p-group of order p^n, with $n > 1$, then G is not simple:

We know if $a \in \mathcal{C}^*$, then $p \mid [G : Z_G(a)]$, so

$$0 \equiv |G| = |Z(G)| + \sum_{\mathcal{O}^*}[G : Z_G(a)] \equiv |Z(G)| \mod p.$$

Since $e \in Z(G)$, we must have $|Z(G)| \geq 1$, so $p \mid |Z(G)|$ and $|Z(G)| \geq p$. In particular, $1 < |Z(G| \subset G$. If G is not abelian, then $1 < Z(G) < G$, and G is not simple, as $Z(G) \triangleleft G$. If G is abelian, then $G = Z(G)$ and there exists an element a in G of order p by Exercise 10.16(2). As $|G| > p$, we have $1 < \langle a \rangle < G$ with $\langle a \rangle \triangleleft G$.

Useful Observation. If we are trying to prove a property about groups and G is a finite non-abelian group with $Z(G) > 1$ (which in general may not occur), then $G/Z(G)$ is a group of order less than that of G and the canonical epimorphism $^- : G \to G/Z(G)$ may allow us to apply induction.

Computation 21.5. Let $G = D_3 = \{e, r, r^2, f, fr, rf\}$, with $|G| = 6$ and satisfying $r^3 = e = f^2$ and $frf^{-1} = r^{-1} = r^2$. We have $fr = r^2 f$ and $rf = fr^2$, so

$$\begin{aligned} C(e) &= \{e\} && \text{and} \quad 1 \mid 6 \\ C(r) &= \{r, r^2\} && \text{and} \quad 2 \mid 6 \\ C(f) &= \{f, fr, rf\} && \text{and} \quad 3 \mid 6 \end{aligned}$$

and

$$\begin{aligned} Z_G(e) &= G && \text{and} \quad |C(e)| = [G : Z_G(e)] = 1 \\ Z_G(r) &= \{e, r, r^2\} && \text{and} \quad |C(r)| = [G : Z_G(r)] = 2 \\ Z_G(f) &= \{e, f\} && \text{and} \quad |C(f)| = [G : Z_G(r)] = 3 \end{aligned}$$

with fixed points

$$Z(G) = \{e\},$$

so

$$\begin{array}{ccccccc} |G| & = & |Z(G)| & + & |C(r)| & + & |C(f)| \\ 6 & = & 1 & + & 2 & + & 3. \end{array}$$

Note that the conjugacy classes (equivalence classes), $C(r) = C(r^2)$ and $C(f) = C(fr) = C(rf)$. This application leads to the following theorem.

Theorem 21.6. *Let G be a group of order p^2 with p a prime. Then G is abelian.*

Proof. Suppose that G is not abelian, then $Z(G) < G$ by Exercise 21.25(1). By Application 21.4, we have $1 < Z(G)$; so by Lagrange's Theorem, we must have $|Z(G)| = p$. Let $a \in G \setminus Z(G)$. Then we have $Z(G) < Z_G(a)$, as $a \in Z_G(a)$. But then the subgroup $Z_G(a)$ must satisfy $|Z_G(a)| = p^2 = |G|$ by Lagrange's Theorem, hence $Z_G(a) = G$. This implies that $a \in Z(G)$, a contradiction. $\square$

Remarks 21.7. Let G be a p-group.

1. If $G = p^2$, the group G is abelian, but may not be cyclic. If G is cyclic, then $G \cong \mathbb{Z}/p^2\mathbb{Z}$. However, the group $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ is not cyclic.
2. D_4 is a 2-group of order 2^3, but not abelian.

Later we shall prove a major theorem. It will imply that every finitely generated abelian group, which includes all finite abelian groups, is a direct product of cyclic groups. There is also a uniqueness statement. At present, we only can show that abelian groups of order pq, with p, q (not necessarily distinct) primes are of this form.

We now jazz up Example 21.1 by letting G act on a subgroup H by conjugation. For this to work, however, we must have $xHx^{-1} \subset H$ for all $x \in G$, i.e, we must have $H \triangleleft G$. So suppose that $H \triangleleft G$. Let G act on H by conjugation, i.e.,

$$* : G \times H \to H \text{ is given by } g * h = ghg^{-1}.$$

We still have $C(a) = G * a$ for all $a \in H$, so if H is a finite group, we have

$$|H| = |F_G(H)| + \sum_{\mathcal{O}^*}[G : Z_G(a)],$$

as $Z_G(a) = G_a := \{x \in G \mid xax^{-1} = a\}$, with $a \in \mathcal{O}$, where $\mathcal{O}$ is a system of representatives for the G-action on H. An application of this is:

Proposition 21.8. *Let G be a p-group and N a nontrivial normal subgroup of G. Then there exists an non-identity element $x \in N$ such that $xy = yx$ for all $y \in G$, i.e., $1 < N \cap Z(G)$.*
[Cf. the case $N = G$ with Application 21.4.]

Proof. Let G act on N by conjugation. Every subgroup of G is either 1 or a p-group, hence

$$0 \equiv |N| = |F_G(N)| + \sum_{\mathcal{O}^*}[G : Z_G(a)] \equiv |F_G(N)| \mod p$$

by the Orbit Decomposition Theorem. As $xex^{-1} = e$ for all $x \in G$, the unity e is a fixed point, so $p \mid |F_G(N)| \geq 1$. Therefore, $|F_G(N)| \geq p$. The result now follows from

$$\begin{aligned} F_G(N) &= \{z \in N \mid xzx^{-1} = z \text{ for all } x \in G\} \\ &= \{z \in N \mid xz = zx \text{ for all } x \in G\} = N \cap Z(G). \end{aligned}$$ □

We can even generalize this conjugation type action a bit further as follows: Let H be a subgroup of G. Define the *normalizer* of H in G by

$$N_G(H) := \{x \in G \mid xHx^{-1} = H\}.$$

The normalizer has the following properties:

Properties 21.9. Let H be a subgroup of G. Then the following are true:
1. $N_G(H)$ is a subgroup of G.
2. $H \triangleleft N_G(H)$.
3. If $H \triangleleft K$ with K a subgroup of G, then $K \subset N_G(H)$.
4. $N_G(H)$ is the unique largest subgroup of G containing H as a normal subgroup.

Replacing G by $N_G(H)$, we can let $N_G(H)$ act on H by conjugation, by the above, i.e., we have an $N_G(H)$-action

$$* : N_G(H) \times H \to H \text{ given by } x * h = xhx^{-1}.$$

In fact, we can go one step further. Let K be a subgroup of G and set $A = K \cap N_G(H)$. Then the action

$$* : A \times H \to H \text{ given by } x * h = xhx^{-1}$$

makes H into an A-set. Instead of pursuing this, we generalize conjugation in another direction.

Example 21.10. Conjugation on Sets.

Let $S = \mathcal{P}(G) := \{A \mid A \subset G, \text{ a subset}\}$, the *power set* of G. Then S becomes a G-set by

$$* : G \times S \to S \text{ given by } x * A = xAx^{-1}$$

where $xAx^{-1} = \{xax^{-1} \mid a \in A\}$. Note that $|A| = |xAx^{-1}|$ for all $x \in G$. We also call this action *conjugation*. The orbit of $A \in \mathcal{P}(G)$ is the *conjugacy class*

$$C(A) = G * A = \{xAx^{-1} \mid x \in G\},$$

and the stabilizer of A is

$$N_G(A) := G_A = \{x \in G \mid xAx^{-1} = A\},$$

called the *normalizer* of A in G. The Orbit Decomposition Theorem yields

$$S = \mathcal{P}(G) = \bigvee_{\mathcal{O}} C(A) = F_G(S) \vee \bigvee_{\mathcal{O}^*} C(A).$$

If G is finite so is $\mathcal{P}(G)$, and we would then have

$$2^{|G|} = |S| = |F_G(S)| + \sum_{\mathcal{O}^*} [G : N_G(A)].$$

This is not so useful, so we try to cut S down. We use the following:

Observation 21.11. Let S be a nonempty set and $\sim$ an equivalence relation of S. Suppose a subset $\emptyset \neq T \subset S$ satisfies the following:

$$\text{For all } a \in T, \text{ we have } [a]_\sim \subset T.$$

Then $\sim|_T$ (actually $\sim|_{T\times T}$) defines an equivalence relation on T and

$$T/\sim_T = \{[a]_\sim \mid a \in T\}.$$

[We are just saying: If $\mathcal{C}$ partitions S and $\emptyset \neq \mathcal{D} \subset \mathcal{C}$, then $\mathcal{D}$ partitions $\bigcup_{\mathcal{D}} \overline{a}$.]

There are many interesting subsets of $\mathcal{P}(G)$. Recall that if H is a subgroup of G, then so is xHx^{-1} for all $x \in G$.

Specific Examples 21.12. Let G be a group. Then the following subsets of $\mathcal{P}(G)$ are G-sets by conjugation (i.e., restricting the action above) if nonempty:

(1) $\mathcal{P}(G)$.
(2) $\mathcal{G} := \{H \mid H \subset G \text{ a subgroup }\}$.
(3) $\mathcal{T}_n := \{A \mid A \subset G \text{ with } |A| = n\}$.
(4) $\mathcal{G} \cap \mathcal{T}_n$.

(5) If G is a finite group, p a (positive) prime satisfying $p^r || |G|$, with $r > 0$, so $G = p^r m$ with p and m relatively prime, then

$$\mathrm{Syl}_p(G) := \mathcal{G} \cap \mathcal{T}_{p^r} = \{H \mid H \text{ a subgroup of } G \text{ of order } p^r\}.$$

(6) If $H \in \mathcal{G}$ then $C(H)$.
(7) If $A \in \mathcal{P}(G)$ then $C(A)$.

We can also push this further by restricting the action of G on one of these sets to the action by a subgroup. We shall later study this action in greater detail, especially (5) and (6).

Example 21.13. Translation Action.

Let G be a group, then the power set $\mathcal{P}(G)$ is a G-set via

$$* : G \times \mathcal{P}(G) \to \mathcal{P}(G) \text{ defined by } g * T = gT := \{gx \mid x \in T\}$$

called *translation*. As before, this is more interesting when we replace $\mathcal{P}(G)$ by appropriate subsets. For example, using the notation of (21.12), G acts on $\mathcal{T}_n$ by translation.

Warning 21.14. G does not act on $\mathcal{G}$ by translation as $H \in \mathcal{G}$ does not mean that $x * H \in \mathcal{G}$. In fact, $x * H$ would lie in $\mathcal{G}$ if and only if $x \in H$.

One also can translate other sets by G. For example, if H is a subgroup of G, we can let G act on the cosets G/H by translation, i.e., let

$$* : G \times G/H \to G/H \text{ be given by } x * aH = xaH.$$

We have used this action before, viz., in the proof of the General Cayley Theorem. The stabilizer of aH in G/H under translation by G is given by:

$$\begin{aligned} G_{aH} &= \{x \in G \mid xaH = aH\} = \{x \in G \mid a^{-1}xaH = H\} \\ &\{x \in G \mid a^{-1}xa \in H\} = \{x \in G \mid x \in aHa^{-1}\} \\ &= aHa^{-1}. \end{aligned}$$

Note that if $H < G$ in the above then $F_G(G/H) = \emptyset$ as $G = G_{aH}$ would imply that $G = aHa^{-1}$.

The following proposition is a useful computing devise. We leave its proof as an exercise.

Proposition 21.15. *Let G be a finite group, H a subgroup of G. Suppose that H is a p-group and $p \mid [G : H]$. Then $p \mid [N_G(H) : H]$. In particular, $H < N_G(H)$. Consequently, if G is a finite p-group and $H < G$ a subgroup, then $H < N_G(H)$.*

We finish the discussion of translation actions to answer our outstanding question on how cosets arise naturally. Indeed, if H is a subgroup of G, and G is a *right* H-set by the *right H-action* by translation on G, i.e.

$$* : G \times H \to G \text{ given by } g * h = gh,$$

then the orbit $g * H$ is the left coset gH and the Orbit Decomposition Theorem in this case is just Lagrange's Theorem. Note the map $*$ is surjective. What is the fiber of $*$ at $g \in G$?

Example 21.16. Automorphic Action.

Let V be a vector space over a field F. A linear transformation $T : V \to V$ is called an *endomorphism* of V. The set

$$\mathrm{End}_F(V) := \{T \mid T : V \to V \text{ a linear transformation}\}$$

is a ring under the usual addition and composition of functions. [From linear algebra, you should know the relationship between $\mathrm{End}_F(V)$ and $\mathbb{M}_n(F)$ if V is n-dimensional.] Let $G = \mathrm{End}_F(V)$. Then we have a map

$$* : G \times V \to V \text{ given by } T * v = T(v)$$

called *evaluation*. If $V = F^n$, we can replace $\mathrm{End}_F(V)$ by $\mathbb{M}_n(F)$. Unfortunately, this does not give a G-action on V as the zero endomorphism under addition is not the identity on V nor is it a group under composition, so although interesting, it is not an example of what we are studying. However, if we let $G = \mathrm{Aut}_F(V) = \left(\mathrm{End}_F\left(V\right)\right)^{\times}$. Then V is a G-set via the *evaluation*

$$* : G \times V \to V \text{ given by } T * v = T(v).$$

Example 21.17. Addendum to Example 21.1.

Let R be any ring, $S = \mathbb{M}_n(R)$, and $G = \mathrm{GL}_n(R)$. Then S is a G-set under conjugation, i.e.,

$$* : G \times S \to S \text{ is given by } A * B = ABA^{-1}.$$

This is an important action and leads to the following:

Problem 21.18. Let F be a field and $*$ the action above. Find a nice system of representatives for this action.

If every non-constant polynomial with coefficients in the field F has a root in F, we say that F is *algebraically closed*, e.g., the complex numbers are algebraically closed by the *Fundamental Theorem of Algebra* (to be proven later). The above sought after system of representatives is called the set of *Jordan canonical forms*. For a general field F, the sought after set is the set of *rational canonical forms*. Proving this will be a major goal later.
[We have another action by

$$* : G \times S \to S \text{ given by } A * B = ABA^t.$$

and ask the same questions. This is especially interesting if $S = \mathbb{M}_n(F)$ is replaced by the subset of symmetric matrices or the subset of skew symmetric matrices.]

Example 21.19. Evaluation.

We have previously seen another example of an action given by evaluation, viz., if S is a nonempty set, then S is a $\Sigma(S)$-set by the evaluation map

$$* : \Sigma(S) \times S \to S \text{ given by } \gamma * s = \gamma(s).$$

If H is a subgroup, we can restrict the action to H. For example, let $\gamma \in \Sigma(S)$ and $\Gamma = \langle\gamma\rangle$. Then S becomes a Γ-set by the action $* : \Gamma \times S \to S$ given by the *evaluation*

$$\gamma^i * s = \gamma^i(s) \text{ for all integers } i.$$

We shall use this action later.

It should be mentioned that evaluation actions are used all the time. If you have an object, then its *automorphism group* [definition?] acts on it by evaluation. For example, such objects can be algebraic as above, topological, or geometric.

Example 21.20. Pullback Action.

As mentioned before a G-action on a set S can be restricted to the action of any subgroup of G. In the examples above, we have used this. We now generalize this. Recall that a subgroup of a group is just a group in which the inclusion map is a monomorphism. More generally, suppose that we have a group homomorphism $\theta : G \to G'$ and a G'-set S defined by $* : G' \times S \to S$. Then S becomes a G-set by the *pullback action* defined by

$$\star : G \times S \to S \text{ is given by } g \star s = \theta(g) * s.$$

Example 21.21. Shift Action.

Our last example, is an action that leads to a nice result called Cauchy's Theorem, which is the first step toward showing that finite groups have subgroups of the order of the power of a prime for any power of that prime dividing the order of the group. Let G be a finite group and p a (positive) prime dividing the order of G. Set

$$S = \{(g_1, \dots, g_p) \in G^p \mid g_1 \cdots g_p = e \text{ in } G\} \subset G^p,$$

where $G^p = G \times \cdots \times G$ (p times). As $(e, \dots, e) \in S$, the set S is nonempty. Moreover, it follows immediately that

$$(g_1, \dots, g_p) \in S \qquad \Longrightarrow \qquad g_p = (g_1 \cdots g_{p-1})^{-1} \text{ in } G.$$

Therefore, for all ordered $(p-1)$-tuples $g_1, \dots, g_{p-1}$, with $g_i \in G$, we have a unique g_p such that $(g_1, \dots, g_p)$ lies in S. Therefore, $|S| = |G^{p-1}|$, so $p \mid |S|$.

We want a group to act on S. Let $^{-} : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism. If $y \in \mathbb{Z}$, let $\tilde{y} \in [y]_p = \overline{y}$ denote the representative that is the smallest positive integer in $\overline{y}$. So $\tilde{y} \in \{1, \dots, p\}$. With this notation, if $x \in \mathbb{Z}$, the ordered tuple

$$\widetilde{1+x}, \dots, \widetilde{p+x}$$

will be a (cyclic) permutation of $1, \dots, p$. Let $(\mathbb{Z}/p\mathbb{Z}, +)$ act on S by

$$(*) \qquad \overline{x} * (g_1, \dots, g_p) = (g_{\widetilde{1+x}}, \dots, g_{\widetilde{p+x}})$$

i.e., if $1 \le i, x \le p$, then the action of $\overline{x}$ shifts i by x slots to the number $x + i$ until it hits p which it then maps to 1, etc. For example, if $p = 5$, then $\overline{3} * (g_1, g_2, g_3, g_4, g_5) = (g_4, g_5, g_1, g_2, g_3)$. The right hand side of (*) still lies in S, since $g_1 \cdots g_p = e$ implies $g_{i+1} \cdots g_p g_1 \cdot g_i = e$ by conjugating successively by $g_1^{-1}, \dots, g_i^{-1}$ $(1 \le i < p)$. It follows easily that S is a $\mathbb{Z}/p\mathbb{Z}$-set under this action. With this action, we can now prove:

Theorem 21.22. (Cauchy's Theorem) *Let p be a (positive) prime dividing the order of the finite group G. Then there exists an element of G of order p.*

Proof. Let S be the $\mathbb{Z}/p\mathbb{Z}$-group under the shift action of Example 21.21. Apply the Orbit Decomposition Theorem 19.9 to get

$$|G^{p-1}| = |S| = |F_{\mathbb{Z}/p\mathbb{Z}}(S)| + \sum_{\mathcal{O}^*}[\mathbb{Z}/p\mathbb{Z} : (\mathbb{Z}/p\mathbb{Z})_x].$$

If $x \in \mathcal{O}^*$, then $(\mathbb{Z}/p\mathbb{Z})_x = 1$ as $\mathbb{Z}/p\mathbb{Z}$ has only the trivial subgroups. So

$$0 \equiv |S| \equiv |F_{\mathbb{Z}/p\mathbb{Z}}(S)| \mod p.$$

As $(e, \dots, e) \in F_{\mathbb{Z}/p\mathbb{Z}}(S)$, we have $|F_{\mathbb{Z}/p\mathbb{Z}}(S)| \geq p > 1$. Thus there exists an element $(e, \dots, e) \neq (g_1, \dots, g_p) \in F_{\mathbb{Z}/p\mathbb{Z}}(S)$. But this means that

$$(g_1, \dots, g_p) = (g_2, \dots, g_p, g_1) = \cdots = (g_{i+1}, \dots, g_p, \dots, g_i) = \cdots = (g_p, g_1, \dots, g_{p-1})$$

lies in G^p. It follows that $e \neq g_1 = \cdots = g_p$. Consequently, if $g = g_1$, then $(g, \dots, g) \in S$, which means that $g^p = e$ as needed. □

Corollary 21.23. *Let p be a prime and G a nontrivial finite group. If every non-identity element in G has order a power of p, then G is a p-group.*

Corollary 21.24. *Let p and q be primes and G of order pq. Then G is not simple.*

Proof. If $p = q$, then G is abelian hence not simple as G is not of prime order. If $p < q$, then G contains a normal group of order q by Useful Counting 12.8. □

Exercises 21.25.

1. Let G be a group. Show all of the following:
   (a) $Z(G) = \bigcap_{a \in G} Z_G(a)$.
   (b) $a \in Z(G)$ if and only if $C(a) = \{a\}$ if and only if $|C(a)| = 1$.
   (c) $a \in Z(G)$ if and only if $G = Z_G(a)$.
   (d) If G is finite, then $a \in Z(G)$ if and only if $|Z_G(a)| = |G|$.
2. Let G be a group and $k(G)$ the number of conjugacy classes in G. Suppose that G is finite. Show that $k(G) = 3$ if and only if G is isomorphic to the cyclic group of order three or the symmetric group on three letters.
3. Let G be a group of order p^2 with p a prime. Show that either $G \cong \mathbb{Z}/p^2\mathbb{Z}$ or $G \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.
4. Show that Properties 21.9 are valid.
5. Compute all the conjugacy classes and isotropy subgroups in A_4.
6. Let G be a group of order p^n, p a prime. Suppose the center of G has order at least p^{n-1}. Show that G is abelian.
7. Let G be a p-group. Using Exercise 17.14(8), show that G is a nilpotent group. In particular a finite product of p-groups for various p is nilpotent.
8. Let G be a nilpotent group (cf. 17.14(8) and H a proper subgroup of G. Show that $H \neq N_G(H)$.

9. Suppose that H is a proper subgroup of a finite group G. Show that $G \neq \bigcup_{g \in G} gHg^{-1}$. [This may not be true if G is infinite.]
10. Let H be a subgroup of G. Let H act on G/H by translation. Compute the orbits, stabilizers, and fixed points of this action.
11. Prove Proposition 21.15. [Hint use the previous exercise.]
12. Let G be a group and S a nonempty set. Show
    (a) If $* : G \times S \to S$ is a G-action, then $\varphi : G \to \Sigma(S)$ given by $\varphi(x)(s) = x * s$ (i.e., if we let $\varphi_x = \varphi(x)$ then $\varphi_x(s) = x * s$) is a group homomorphism (called the *permutation representation*).
    (b) If $\varphi : G \to \Sigma(S)$ is a group homomorphism then $* : G \times S \to S$ given by $x * s = \varphi_x(s)$ where $\varphi_x = \varphi(x)$ is a G-action.
13. Let Q be the quaternion group. Show that there exist no group monomorphism $\varphi : Q \to S_7$.
14. Let G be a group of order pq with $q \leq p$ primes. Show that G is not simple and has a normal subgroup of order p.
15. Let G be a finite p-group. Show if $p^n \mid |G|$ then G has a normal subgroup of order p^n.
16. Let p be an odd prime. Classify all groups G of order $2p$ up to isomorphism.
17. Let S be a G-set. Suppose that both S and G are finite. If $x \in G$ define the *fixed point set* of x in G by
$$F_S(x) := \{s \in S \mid x \cdot s = s\}.$$
Show the number of orbits N of this action satisfies
$$N = \frac{1}{|G|} \sum_G |F_S(x)|.$$
[So N is the average of the size of the fixed point sets of elements of G.]
18. Let G be a finite group acting transitively on a finite set S with $|S| > 1$. Prove that there exists an element $g \in G$ fixing no element in S.

22. Sylow Theorems

Let G be a finite group of order of order $p^r m$ with p a prime relatively prime to m and $r > 0$. A subgroup H of G is called a *Sylow p-subgroup* of G if H has order p^r. Let
$$\mathrm{Syl}_p(G) := \{H \mid H \text{ a Sylow } p\text{-subgroup of } G\}.$$

The natural question is answered by the next result. Its proof uses much of what we have done.

Theorem 22.1. (First Sylow Theorem) *Let p be a (positive) prime dividing the order of a finite group G. Then $\mathrm{Syl}_p(G)$ is nonempty, i.e., there exists a Sylow p-subgroup.*

We prove the stronger

Theorem 22.2. (Generalized First Sylow Theorem) *Let p be a (positive) prime such that p^s, with $s \geq 0$, divides the order of a finite group G. Then there exists a subgroup of G of order p^s.*

Proof. We induct on the order of G. If $|G| \leq p$, the result is trivial. We make the following:

Induction Hypothesis. If T is any finite group satisfying $|T| < |G|$ and $p^s \mid |T|$, $s \geq 1$, then T contains a subgroup of order p^s.

We must show that G contains a subgroup of order p^s. If $H < G$ is a subgroup of G satisfying $p \not\mid [G : H]$, then $p^s \mid |H|$. By the Induction Hypothesis, H, hence G, contains a subgroup of order p^s. Therefore, we are done unless we make the following:

Assumption. If $1 < H < G$ is a subgroup, then $p \mid [G : H]$.

For example, if $a \in G$ satisfies $Z_G(a) < G$, then $p \mid [G : Z_G(a)]$. In particular, this is true for any $a \notin Z(G)$, i.e., $a \in \mathcal{C}^* = \mathcal{C} \setminus Z(G)$, where $\mathcal{C}$ is a system of representatives for conjugate action of G on G in Example 21.1. Applying the Class Equation (21.3), we have

$$0 \equiv |G| = |Z(G)| + \sum_{\mathcal{C}^*} [G : Z_G(a)] \equiv |Z(G)| \mod p,$$

so $p \mid |Z(G)|$. By Cauchy's Theorem 21.22, there exists an element $a \in Z(G)$ of order p. Let $H = \langle a \rangle$, a subgroup of G of order p. Since $a \in Z(G)$, we have $xa^ix^{-1} = a^i$ for all $x \in G$ and $i \in \mathbb{Z}$. In particular, $H \triangleleft G$ (cf. Example 11.5(3)) and G/H is a group of order $[G : H] = |G|/|H| = |G|/p < |G|$. Consequently, $p^{s-1} \mid |G/H|$; and, by the Induction Hypothesis, there exists a subgroup $T \subset G/H$ of order p^{s-1}. Let $^- : G \to G/H$ be the canonical epimorphism. Then H is the kernel and by the Correspondence Principle, there exists a subgroup $\widetilde{T}$ of G containing H and satisfying $T = \widetilde{T}/H$. Hence $|\widetilde{T}| = |T||H| = p^s$, and $\widetilde{T}$ works. $\square$

To investigate the set of Sylow p-subgroups, $\mathrm{Syl}_p(G)$ of G, we need three lemmas.

Lemma 22.3. *Let H and K be subgroups of a finite group G. Then $|HK| = |H||K|/|H \cap K|$.*

Proof. This is equation (13.5). (Cf. Exercise 10.16(5).) $\square$

The main mathematical reason for the validity of the further Sylow theorems below is the following:

Lemma 22.4. *Let G be a finite group, P a Sylow p-subgroup, and H a subgroup of G that is also a p-group. If H is a subgroup of $N_G(P)$, i.e., $hPh^{-1} = P$ for all $h \in H$, then H is a subgroup of P. In particular, if, in addition, H is also a Sylow p-subgroup of G, then $H = P$.*

Proof. The key to the proof of this lemma is the Second Isomorphism Theorem together with the counting version of the Second Isomorphism Theorem of Lemma 22.3. By definition, $P \triangleleft N_G(P)$ and by hypothesis $H \subset N_G(P)$. Hence, by the Second Isomorphism Theorem, we have HP is a subgroup of $N_G(P)$, with $P \triangleleft HP$ and $H \cap P \triangleleft H$, and satisfying $HP/P \cong H/H \cap P$. In particular, by Lemma 22.3, we have

$$(*) \qquad |HP| = |P| \frac{|H|}{|H \cap P|}.$$

As P, H are p-groups and $H \cap P \subset H$ a subgroup, (*) implies that HP is a p-group satisfying $P \subset HP \subset G$. Thus HP is a Sylow p-subgroup of G, i.e., $|P| = |HP|$, so $P = HP$, hence $H \subset P$. If, in addition, $|H| = |P|$ then $H = P$. □

We convert the lemma above into the language of group actions with the action given by conjugation on sets. Explicitly, we use Specific Examples 21.12(6) after restricting the action of G to a subgroup.

Lemma 22.5. *Let P be a Sylow p-subgroup of a finite group G and H a subgroup of G that is also a p-group. Then the following hold:*

(1) $C(P)$ consists of Sylow p-subgroups of G.

(2) The conjugacy class $C(P)$ is an H-set by conjugation, i.e., via the action $* : H \times C(P) \to C(P)$ by conjugation.

(3) Suppose that T is a fixed point under the H-action in (2), i.e.,

$$T \in F_H\big(C(P)\big) = \{W \in C(P) \mid xWx^{-1} = W \text{ for all } x \in H\}.$$

Then $H \subset T$.

(4) If H in (3) is a Sylow p-subgroup, then, under the H-action in (2), H is the only possible fixed point, i.e., $F_H\big(C(P)\big) \subset \{H\}$.

Proof. (1): As $|P| = |xPx^{-1}|$ for all x in G, this is clear.

(2) is immediate.

(3): As T is an H-fixed point, the orbit $H * T = \{T\}$, so $xTx^{-1} = T$ for all $x \in H$. By definition, this means that $H \subset N_G(T)$. By Lemma 22.4 and (1), we conclude that $H \subset T$.

(4): If $T \in F_H\big(C(P)\big)$, then $H \subset T$ by (3). If further, H is a Sylow p-subgroup, then by (3), we have $H = T$. □

Note. Under the hypothesis of Lemma 22.5, we have the isotropy subgroup of P is given by $H_P = N_G(P) \cap H$ for any Sylow p-subgroup of G. But P is then also a Sylow p-subgroup of $N_G(P)$, so by restricting the H-action to H_P the proof of Lemma 22.5 shows that $N_G(P) \cap H \subset P$.

Theorem 22.6. (Sylow Theorems). *Let G be a finite group, p a (positive) prime dividing the order of G. Then the following are true:*

First Sylow Theorem. *There exists a Sylow p-subgroup of G.*

Second Sylow Theorem. *All Sylow p-subgroups are conjugate, i.e., if P is a Sylow p-subgroup of G, then $C(P) = \mathrm{Syl}_p(G)$.*

Third Sylow Theorem. *Let P be a Sylow p-subgroup of G. Then the following are true:*

(i) $|\mathrm{Syl}_p(G)| = [G : N_G(P)]$.

(ii) $|\mathrm{Syl}_p(G)| \mid |G|$.

(iii) $|\mathrm{Syl}_p(G)| \equiv 1 \mod p$.

(iv) $|\mathrm{Syl}_p(G)| \mid [G : P]$, i.e., if $|G| = p^n m$ with $p \nmid m$, then $|\mathrm{Syl}_p(G)| \mid m$.

Fourth Theorem. *Let H be a subgroup of G that is a p-group. Then H lies in some Sylow p-subgroup of G.*

Proof. Let H be a subgroup of G that is a p-group. We have already proven the First Sylow Theorem, so let P be a Sylow p-subgroup of G. By Lemma 22.5, we have $C(P) \subset \mathrm{Syl}_p(G)$ is an H-set under conjugation. By the Orbit Decomposition Theorem, we have

$$|C(P)| = |F_H\big(C(P)\big)| + \sum_{\mathcal{O}^*} [H : H_T].$$

If $T \in \mathcal{O}^*$, then $H_T < H$, so $p \mid [H : H_T]$, hence

$$(\dagger) \qquad |C(P)| \equiv |F_H\big(C(P)\big)| \mod p.$$

Note. Equation (†) is interesting, since H does not appear at all on the left hand side, i.e., (†) is true for all p-subgroups of H. To evaluate a constant function (of such p-subgroups of H) given by (†), it suffices to evaluate at any such H.

Case 1. Let $H = P$:

By Lemma 22.5, we conclude that $F_P\big(C(P)\big) \subset \{P\}$. As $xPx^{-1} = P$ for all $x \in P$, we must have $F_P\big(C(P)\big) = \{P\}$.

We use Case 1, to evaluate the constant in (†) as

$$(*) \qquad |C(P)| \equiv |F_P\big(C(P)\big)| = 1 \mod p.$$

Case 2. Let H be a Sylow p-subgroup (not necessarily P):

Applying (†) and (*), we have

$$1 \equiv |C(P)| \equiv |F_H\big(C(P)\big)| \mod p.$$

Hence $F_H\big(C(P)\big)$ is nonempty so $F_H\big(C(P)\big) = \{H\}$ by Lemma 22.5. In particular, $H \in F_H\big(C(P)\big) \subset C(P)$ proving that $\mathrm{Syl}_p(G) = C(P)$, which is the Second Sylow Theorem. Plugging this information into (*), we have

$$|\,\mathrm{Syl}_p(G)| = |C(P)| \equiv 1 \mod p.$$

Since $|C(P)| = [G : N_G(P)] = |G|/|N_G(P)|$ and $P \subset N_G(P)$ with $p \nmid [G : N_G(P)]$, we also have proven the Third Sylow Theorem.

Case 3. Let H be an arbitrary p-subgroup of G:

By (*), we have

$$|F_H\big(C(P)\big)| \equiv |C(P)| \equiv 1 \mod p.$$

Thus there exists a $T \in F_H\big(C(P)\big)$, hence $H \subset T \in \mathrm{Syl}_p(G)$ by Lemma 22.5. This proves the Fourth Theorem.

□

Recall that a subgroup H of a group G is called a *characteristic subgroup* of G if the restriction map $res : \mathrm{Aut}(G) \to \mathrm{Aut}(H)$ given by $\varphi \mapsto \varphi|_H$ is well-defined, i.e., the target is correct. We write $H \lhd\lhd G$. We know that characteristic subgroups are normal, and being characteristic is a transitive property.

Proposition 22.7. *Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Then the following are equivalent:*

(1) $\mathrm{Syl}_p(G) = \{P\}$.
(2) $C(P) = \{P\}$.

(3) $P \triangleleft G$.
(4) $P \triangleleft \triangleleft G$.

Proof. The equivalence of (1), (2), and (3) follows from the Second Sylow Theorem. Clearly, (1) implies (4) as all isomorphic images of a group have the same cardinality. Finally (4) implies (3) by the comment above.

□

Corollary 22.8. *Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Then $N_G(P) = N_G\big(N_G(P)\big)$.*

Proof. As P is a Sylow p-subgroup of $N_G(P)$ and normal in it, P is a characteristic subgroup of $N_G(P)$ by the proposition. As $N_G(P)$ is a normal subgroup of $N_G\big(N_G(P)\big)$, we conclude by Exercise 11.9(22) that P is normal in $N_G\big(N_G(P)\big)$, i.e., $N_G\big(N_G(P)\big) \subset N_G(P)$. So we must have equality. □

A stronger result is left as Exercise 22.15(1) below.

Examples 22.9. Let G be a finite group, p a prime dividing the order of G. In the following examples, we shall use the following notation:

P_p will denote an arbitrary Sylow p-subgroup of G.
n_p will denote the number of Sylow p-subgroups.

So there exists an integer k (depending on p) with

$n_p = |\operatorname{Syl}_p(G)| = [G : N_G(P)] = 1 + pk$
P_p is normal if and only if $k = 0$.

1. Every group of order 15 is cyclic (hence isomorphic to $\mathbb{Z}/15\mathbb{Z}$) and every group of order 45 is abelian (hence isomorphic to $\mathbb{Z}/45\mathbb{Z}$ or $\mathbb{Z}/15\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$):

   In both cases, the Third Sylow Theorem implies the Sylow p-subgroups are normal for $p = 3, 5$. As groups of order p are cyclic and of order p^2 are abelian, the result follows by Exercise 13.7(9). [We leave the statement about isomorphism as an exercise.]
2. If G is a group of order $p^r q$ with $p > q$ primes, then P_p is normal by the Third Sylow Theorem.
3. If G is a group of order $315 = 3^2 \cdot 5 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. We have
   $n_3 = 1 + 3k \mid 5 \cdot 7$ some k, so $[G : N_G(P_3)] = 7$.
   $n_5 = 1 + 5k \mid 3^2 \cdot 7$ some k, so $[G : N_G(P_5)] = 21$.
   By Lagrange's Theorem,
   $$N_G(P_3) \text{ is of order } 45 \text{ and } N_G(P_5) \text{ is of order } 15,$$
   so both are abelian, and $\operatorname{Syl}_5\big(N_G(P_3)\big) \subset \operatorname{Syl}_5(G)$. In particular, if $T \in \operatorname{Syl}_5\big(N_G(P_3)\big)$, we have $T \triangleleft N_G(P_3)$ (or $1+5k \mid 3^2$ implies $k = 0$). Thus $N_G(P_3) \subset N_G(T)$. We conclude that
   $$45 = |N_G\big(P_3)\big)| \mid |N_G(T)| = 15,$$
   which is impossible.

4. If G is a group of order $525 = 3 \cdot 5^2 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. Let P and Q be distinct Sylow 5-subgroups and set $I = P \cap Q$. We must have $|I| = 1$ or 5. Although PQ is not a group, we know by Exercise 10.16(5) that
$$|PQ| = |P|\,|Q|/|P \cap Q| = |P|\,|Q|/|I|.$$
   Suppose that $I = 1$. Then $|PQ| = 5^4 > |G|$, so this is impossible. Hence $|I| = 5$. Both P and Q are abelian as of order 5^2, so $P, Q \subset N_G(I)$. Therefore,
$$|N_G(I)| \geq |PQ| = 5^3.$$
   As $|N_G(I)| \mid |G|$, it follows that
$$|N_G(I)| = 3 \cdot 5^2 \cdot 7 \text{ or } 5^2 \cdot 7.$$
   In the first case, $I \triangleleft N_G(I) = G$; and in the second case $[G : N_G(I)] = 3$ and $|G| \not\mid 3!$, so $N_G(I)$ contains a nontrivial normal subgroup of G by Useful Counting 12.8. In fact, $N_G(I) \triangleleft G$, as 3 is the smallest prime dividing $|G|$ (cf. Corollary 12.11).

5. If G is a group of order $945 = 3^3 \cdot 5 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. As $n_3 = 1 + 3k \mid 5 \cdot 7$, we have $[G : N_G(P_3)] = n_3 = 7$. Since $|G| \not\mid 7!$, the subgroup $N_G(P_3)$ contains a nontrivial normal subgroup of G by Useful Counting 12.8.

6. If G is a group of order $1785 = 3 \cdot 5 \cdot 7 \cdot 17$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. We have
$$n_{17} = 1 + 17k \mid 3 \cdot 5 \cdot 7, \quad \text{so } n_{17} = 5 \cdot 7.$$
$$n_3 = 1 + 3k \mid 5 \cdot 7 \cdot 17, \quad \text{so} \quad n_3 = 7,\ 5 \cdot 17, \quad \text{or} \quad 5 \cdot 7 \cdot 17.$$
   As $|G| = 1785 \not\mid 7!$, we can eliminate $n_3 = 7$ by Useful Counting 12.8. Thus
$$|N_G(P_{17})| = 3 \cdot 17 \quad \text{and} \quad |N_G(P_3)| = 3 \cdot 7 \quad \text{or} \quad 3.$$
   As $3 \mid |N_G(P_{17})|$ and $3 \mid |G|$, we may assume that $P_3 \subset N_G(P_{17})$. But $1 + 3k \mid 17$ implies $k = 0$, so $P_3 \triangleleft N_G(P_{17})$. Consequently, $N_G(P_{17}) \subset N_G(P_3)$, implying that
$$3 \cdot 17 = |N_G(P_{17})| \leq |N_G(P_3)| = 3 \cdot 7 \text{ or } 3,$$
   which is impossible.

7. Here is a fact useful in counting. Suppose that $|G| = pm$ with $p \not\mid m$, p a prime, then
   G contains $n_p(p-1)$ elements of order p, so $|G| \geq n_p(p-1) + 1$,

   adding in the identity. Similarly, if $|G| = pqm$ with $p \not\mid m$, $q \not\mid m$, and p and q distinct primes, then
$$|G| \geq n_p(p-1) + n_q(q-1) + 1,$$
   adding in the identity. This last remark does not work if $p = q$, because two Sylow p-groups can intersect nontrivially, and it is difficult to estimate the number of elements in the union of all the Sylow p-subgroups.

For some other results, see the exercises below.

The proof of the following theorem illustrates many of the ideas that we have used in the examples.

Theorem 22.10. *Let p and q be primes, G a group of order $p^s q$ with $s \geq 1$. Then G is not simple.*

Proof. We may assume that $p \neq q$ and there exist no normal Sylow subgroups. In particular, if $n_p = |\operatorname{Syl}_p(G)|$, then $1 < n_p = 1 + kp \mid q$ with $k \geq 1$, an integer. In particular, $n_p \leq q$. Among all the Sylow p-subgroups of G, choose two distinct ones, P_1 and P_2, satisfying $I = P_1 \cap P_2$ has maximal order. If $I = 1$, then all Sylow p-subgroups of G intersect pairwise in the identity, so the number of non-identity elements in the Sylow p-subgroups is $(p^s - 1)q$. As $|G| = p^s q$, we conclude that the Sylow q-group is normal, a contradiction. So we may assume that $1 < I$. As I is a p-group but not a Sylow p-subgroup, we have $1 < I < N_{P_i}(I)$ for $i = 1, 2$ by Proposition 21.15. Let $H = \langle N_{P_1}(I), N_{P_2}(I)\rangle$, the group generated by $N_{P_1}(I)$ and $N_{P_2}(I)$. Then we must have $I \triangleleft H$. In particular, $H \subset N_G(I)$ by the definition of $N_G(H)$.

Claim. $q \mid |H|$.

Suppose not. Then H is a p-group, so by the Fourth Theorem, there exists a Sylow p-group P_3 of G such that $H \subset P_3$. Hence for $i = 1, 2$, we have

$$P_1 \cap P_2 = I < N_{P_i}(I) \subset P_i \cap H \subset P_i \cap P_3.$$

This contradicts the maximality of $|I|$, lest $P_1 = P_1 \cap P_3 = P_2 \cap P_3 = P_2$. This proves the claim, i.e., $q \mid |H|$. Let Q be a Sylow q-group of H, hence also of G as $q \,||\, |G|$. Consequently, by Exercise 10.16(5), $|P_1 Q| = |P_1|\,|Q|/|P_1 \cap Q| = |P_1|\,|Q| = |G|$ as p and q are relatively prime. This means that

$$(*) \qquad G = P_1 Q := \{xy \mid x \in P_1,\, y \in Q\}.$$

If $g \in G$, set $I^g := gIg^{-1}$ and $N = \langle I^x \mid x \in G\rangle$. By Great Trick 11.3. we have $I \subset N \triangleleft G$. Let $g \in G$. By (*), we can write $g = xy$, for some $x \in P_1$ and $y \in Q$. In particular, as $y \in Q \subset H \subset N_G(I)$, we have

$$I^g = I^{xy} = xyIy^{-1}x^{-1} = xIx^{-1} = I^x \subset P_1,$$

as $I \subset P_1$. Thus $N \subset P_1 < G$, so $1 < N \triangleleft G$, and the theorem follows. □

Definition 22.11. Let G be a group. We say that a subgroup M of G is a *maximal subgroup* if $M < G$ and if $M < H \subset G$ is a subgroup, then $H = G$. We say that G is called an *(internal) direct product* of normal subgroups $N_1, \cdots, N_r$ if $G = N_1 \cdots N_r$ and $N_i \cap N_1 \cdots N_{i-1} N_{i+1} \cdots N_r = 1$, for $i = 1, \dots, r$. (Cf. Exercises 13.7(3)(4)(5).)

An interesting class of groups in the finite case is determined by the next result theorem.

Theorem 22.12. *Let G be a finite group. Then the following are equivalent*

(1) *Every Sylow subgroup of G is normal in G.*
(2) *G is isomorphic to an (internal) direct product of its Sylow subgroups.*
(3) *Every maximal subgroup of G is normal in G.*

Proof. Let $|G| = p_1^{e_1} \cdots p_r^{e_r}$ be a factorization of $|G|$ with distinct primes $p_1, \dots, p_r$ and $e_i > 0$, $i = 1, \dots, r$. Let $P_i \in \operatorname{Syl}_{p_i}(G)$ for $i = 1, \dots, r$. So P_i is normal in G if and only if $\operatorname{Syl}_{p_i}(G) = \{P_i\}$.

(1) $\Rightarrow$ (2): Since $P_1 \cdots P_n$ is a subgroup of G of the same order as G they are equal. This is an internal direct product as p_i and $p_1 \cdots p_{i-1}p_{i+1} \cdots p_r$ are relatively prime and all the P_i are normal subgroups of G.

(2) $\Rightarrow$ (3): Let M be a maximal subgroup of G. By Exercise 13.7 (6), M is a direct product of the $M \cap P_i$, $i = 1, \dots, r$. Since M is maximal, it must be $P_1 \cdots P_{i-1}P_{i+1} \cdots P_r$ for some i.

(3) $\Rightarrow$ (1): If P_i is not normal in G, then $P < N_G(P_i) < G$ by Corollary 22.8. As G is a finite group, there exists a maximal subgroup M of G containing $N_G(P_i)$ and hence also $P_i \in \mathrm{Syl}_{p_i}(M)$. As $P_i \subset N_G(P_I) \subset M$ and $M \triangleleft G$, by the Frattini Argument (Exercise 22.15(2) below), we have $G = N_G(P_i)M$. But $N_G(P_i)M \subset M$, a contradiction. □

Definition 22.13. A finite group G is called *nilpotent* if it satisfies the equivalent conditions of Theorem 22.12 are satisfied. This agrees with the definition of a nilpotent group defined in Exercise 17.14(8) for finite groups.

Corollary 22.14. *Every finite nilpotent group is solvable.*

Exercises 22.15.

1. Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Suppose that H is a subgroup of G satisfying $N_G(P) \subset H \subset G$, then $H = N_G(H)$.
2. (Frattini Argument) Let G be a finite group and $K \triangleleft G$. If P is a p-Sylow subgroup of K, show that $G = KN_G(P)$.
3. Classify groups up to isomorphism of order pq with p and q primes.
4. Let G be a finite group of order p^2q with p and q primes. Show that G is not simple without using Theorem 22.10.
5. Let G be a finite abelian group of order p^2q with p and q distinct primes. Show that G is isomorphic to either $\mathbb{Z}/p^2q\mathbb{Z}$ or $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/pq\mathbb{Z}$ and the second group is isomorphic to $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$.
6. Let G be a finite group of order p^2q^2 with p and q primes. Show that G is not simple.
7. Let G be a group of order 56. Using Example 22.9(7) but not Theorem 22.10, show that G is not simple.
8. Let G be a finite group of order pqr with p, q, and r primes. Show that G is not simple.
9. Show that groups of order $231 = 3 \cdot 7 \cdot 11$ are semi-direct products and show that there are exactly two such groups up to isomorphism.
10. Let G be a finite group of order pqr with $p < q < r$ primes. Show that the Sylow r-subgroup of G is normal.
11. It can be shown that any finite group that is the product of distinct primes (each to degree one) is solvable. Prove that the Sylow p-subgroup of any such group G with p the largest prime dividing the order of G is normal in B.
12. Show that no group of order 112 is simple.
13. Show that no group of order 120 is simple.
14. Show that no group of order 144 is simple.

15. Show that no group of order 2000 is simple.
16. Show that no group of order 4000 is simple.
17. Show that no group of order 8000 is simple.
18. Let G be a finite group, p the smallest prime dividing the order of G. Suppose that G contains a cyclic Sylow p-subgroup. Show that $N_G(P) = Z_G(P)$. (Cf. Exercise 12.12(8)).
19. Let p be a prime dividing the order of a finite group G. Let $P \in \mathrm{Syl}_p(G)$. If $N \triangleleft G$ show that $N \cap P \in \mathrm{Syl}_p(N)$.
20. Let G be a finite group with p a prime dividing $|G|$ but not a p-group. Suppose that G contains two Sylow p-subgroups whose intersection I has maximal cardinality. Show the $N_G(I)$ is not a p-group.
21. Let G be a finite group with p a prime dividing G but not a p-group. Suppose that G contains two Sylow p-subgroups whose intersection is not trivial. Show that there exists a subgroup $H < G$ such that either $G = N_G(H)$ or H is not a p-subgroup but contains contains a Sylow p-subgroup of G.
22. Using Exercise 17.14(8) and Exercise 21.25(7) show that a finite group is nilpotent if and only if it is a product of its Sylow subgroups.
23. A normal subgroup $N \neq 1$ of G is called a *minimal normal* subgroup if it contains no normal subgroups of G other than itself and 1. Show if G is a nontrivial finite solvable group, then any minimal normal subgroup of G is an *elementary p-group*, i.e., a group isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$ for some prime p and positive integer n.
24. If G is a finite nilpotent group, then the center of G is not trivial. In particular, S_3 is not nilpotent.
25. If G is a finite nilpotent group, then every subgroup and quotient group of G is nilpotent. Show the converse is false in general.
26. Let G be a finite group. If every proper subgroup of G is nilpotent, then G is solvable.

23. Addendum: Finite Solvable Groups

In this section, we wish to extend the Sylow theorems in the case of a finite solvable group. We use this to give a new characterization of finite solvable groups. The proofs give a nice application of results that we have proven.

We will be interested in the following subgroups.

Definition 23.1. Let G be a finite group of order mn with m and n relatively prime. A subgroup of G of order n is called an *Hall n-subgroup* of G.

Of course, every Sylow p-subgroup is a Hall subgroup. We are interested when Hall n-subgroups exist and if any two such are conjugate. In general, this is false. We shall prove that this is true in the case that G is a solvable group.

Crucial to our study will be the nontrivial normal subgroups of a group, if they exist.

Definition 23.2. Let G be a nontrivial group. A subgroup $1 < N \subset G$ is called a *minimal normal subgroup* of G if there exists no normal subgroup H of G with $1 < H < N$.

Note that if G is a nontrivial finite group, then minimal normal subgroups of G exist. Of course, if G is simple, then G is the minimal normal subgroup of G.

In the following proposition, we use use properties about the internal direct product of finitely many groups. [Cf. Exercise 13.7(3)-(6).]

Proposition 23.3. *Let G be a finite group and $1 < N \subset G$ be a minimal normal subgroup. Among all the subgroups of G isomorphic to N, choose those $N = N_1, \ldots, N_r$, $r \geq 1$ that satisfy*

$$H = N_1 \cdots N_r \cong N_1 \times \cdots \times N_r \text{ and with } |H| \text{ maximal.}$$

Then H is a characteristic subgroup of G.

Proof. We know that $H \triangleleft G$. Let $\sigma : G \to G$ be a group automorphism. We have to show that $\sigma(N_i) \subset H$ for $i = 1, \ldots, r$. We know that $\sigma(N_i)$ is a normal subgroup of G isomorphic to N, so if $\sigma(N_i)$ is not a subgroup of H, then $\sigma(N_i) \cap H < \sigma(N_i)$, hence is of lesser order. Since $\sigma(N_i) \cap H$ is a normal subgroup of G, the minimality of $|H|$ implies that $\sigma(N_i) \cap H = 1$. But this implies that the subgroup generated by H and $\sigma(N_i)$ is isomorphic to $H \times \sigma(N_i)$ (why?), contradicting the maximality of $|H|$. Therefore, $\sigma(N_i) \subset H$, and H is a characteristic subgroup of G. □

Corollary 23.4. *Let G be a nontrivial finite group having no nontrivial characteristic subgroups. Then G is either a simple group or a direct product of isomorphic simple groups.*

Proof. We use the notation of Proposition 23.3. If there exists a normal subgroup $1 < H \triangleleft N$, then $H \triangleleft N_1 \cdots N_r$, since $N_1 N_2 \cdots N_r \cong N_1 \times \cdots \times N_r$. By Proposition 23.3 and the hypothesis, we must have $N_1 N_2 \cdots N_r = G$. As N is a minimal normal subgroup of G, we must have $N = H$, so N is simple. The result follows. □

Corollary 23.5. *Let N be a minimal normal subgroup of finite group G. Then N is either simple group or a direct product of isomorphic simple groups.*

Proof. If $1 < H$ is a characteristic subgroup of N, then N is a normal subgroup of G by Exercise 11.9(22). The result follows by Corollary 23.4. □

Corollary 23.6. *Let N be a minimal normal subgroup of a finite solvable group G. Then there exists a prime $p > 0$ such that N is cyclic of order p or the direct product of cyclic groups of order p. In particular, N is abelian.*

Proof. This follows immediately as groups of prime order constitute the abelian simple groups. □

We need a simple lemma whose proof we leave as an exercise.

Lemma 23.7. *Let G be a group and $H \subset G$ a subgroup. Then for all $x \in G$, we have $N_G(xHx^{-1}) = xN_G(H)x^{-1}$.*

We state and prove the first key result in this section. It generalizes the Sylow theorems in the case of a finite solvable group.

Theorem 23.8. (Hall) *Let G be a finite solvable group of order mn with m and n relatively prime. Then*

(1) *There exists a Hall m-subgroup of G.*
(2) *Every subgroup of G whose order divides m lies in a Hall m-subgroup of G.*
(3) *Every pair of Hall m-subgroups of G are conjugate.*
(4) *Suppose that $|G| = p_1^{e_1} \dots p_r^{e_r}$ is a factorization into distinct primes with $e_i > 0, i = 1, \dots, n$. Then the number h_m of Hall m-subgroups of G is a product of factors each of which is congruent to 1 modulo some prime factor of m and a power of a prime dividing some $p_i^{e_i}$, $i = 1, \dots, r$.*

Proof. As subgroups and homomorphic images of solvable groups are solvable by Theorem 17.3, this allows us to induct on the order of a solvable finite group. If $G = 1$, the result is trivial, so we may also assume that $|G| > 1$. We may also assume that $m > 1$ and $n > 1$.

Case 1. G contains a normal subgroup $N \triangleleft G$ of order $m_1 n_1$ with $m_1 \mid m$ and $n_1 \mid n$ satisfying $n_1 < n$:

(1): As N is normal, G/N is solvable and $|G/N| = \frac{m}{m_1}\frac{n}{n_1}$ with $\frac{m}{m_1}$ and $\frac{n}{n_1}$ relatively prime. Since $n_1 < n$, by the Correspondence Principle, there exists a subgroup $N \subset H \subset G$ with the group H/N a Hall $\frac{m}{m_1}$-subgroup of G/N by induction. As $|H| = \frac{m}{m_1}|N| = mn_1 < mn$ and H is also solvable, H (hence G) contains a Hall m-subgroup.

(2): Suppose that H' is another Hall m-subgroup. We must show that H and H' are conjugate. The subgroup HN of G is of order $\frac{|N||H|}{|N\cap H|} = \frac{|N||H|}{|N|} = mn_1$. Similarly, the subgroup $H'N$ of G is of order mn_1. The subgroups HN/N and $H'N/N$ of G/N are of order $\frac{mn_1}{m_1 n_1} = \frac{m}{m_1}$, so Hall $\frac{m}{m_1}$-subgroups of solvable G/N. By induction on $|G|$, they are conjugate. Let $^{-} : G \to G/N$ be the canonical group epimorphism. Then there exists an x in G satisfying $\overline{x}(HN/N)\overline{x}^{-1} = H'N/N$. It follows that $xHNx^{-1} = H'N$. Since H' and xHx^{-1} are Hall m-subgroups of solvable $H'N$, they are conjugate by induction.

(3): Let $H_0 \subset G$ be a subgroup of order m_0 with $m_0 \mid m$. Then $|H_0N/N| \mid \frac{m}{m_1}$, so H_0N/N lies in a Hall $(\frac{m}{m_1})$-subgroup of H_0N/N by induction. Therefore, H_0 is a Hall m-subgroup of some subgroup K of G of order mn_1. By induction on $\frac{n}{n_1}$, the group K, hence H_0, lies in a Hall m-subgroup of G.

(4): By the proof above, h_m is a product of the number of Hall $(\frac{m}{m_1})$-subgroups of G/N and the number of conjugates of H in HN. The maximal power of the primes factors $p_i^{f_i}$ of $|HN|$ divide $p_i^{e_i}$ and the maximal power of the primes factors of $|G/N|$ are $p_i^{e_i}$ for some of the $p_i, 1 = 1, \dots .r$. By induction, h_m is a product of two factors that satisfy condition (4). Therefore condition (4) is established for G.
This establishes Case 1.

Reduction. By Case 1, if there exists a normal subgroup $1 < N < G$ with $n \not| \, |N|$, we are done. So we may assume that $n \mid |N|$ for every normal subgroup $1 < N < G$. In particular, this applies to each minimal normal subgroup H of G. By Corollary 23.6, H is an abelian group of order p^r for some prime p and $r \in \mathbb{Z}^+$. Fix such an H. As $n \mid |H|$, we must have $n = p^r$. Therefore, H is a normal Sylow p-subgroup of G, hence the unique

Sylow p-subgroup of G. In particular, if N is any proper normal subgroup of G, then $H \subset N$. It follows that H is the unique minimal normal subgroup of G.

Case 2. Case 1 does not hold. In particular, the group H in the Reduction is the unique minimal normal subgroup of G:

(1): The nontrivial finite group G/H contains a minimal normal subgroup. By the Correspondence Principle, it follows that there exists a normal subgroup $1 < K < G$ satisfying $K/H < G/H$ is a minimal normal subgroup. By Corollary 23.6, $|K/H| = q^s$ for some prime q and $s \in \mathbb{Z}^+$ with $q \neq p$. Therefore, $|K| = p^r q^s$. Let $Q \in \mathrm{Syl}_q(G)$. Then we have $K = HQ$, since they are groups of the same order, and $Q \cap H = 1$, since $|Q|$ and $|H|$ are relatively prime.

Claim. $N_G(Q)$ is a Hall m-subgroup of G:

We have $N_K(Q) = K \cap N_G(Q)$. As $H \triangleleft K$ and $K \triangleleft G$, by the Frattini Argument (Exercise 22.15(2)), we know that

$$K = HN_K(Q) \text{ and } G = KN_G(Q).$$

As

$$G/H = KN_G(Q)/K \cong N_G(Q)/\big(K \cap N_G(Q)\big) = N_G(Q)/N_K(Q),$$

we have $|N_G(Q)| = |G||N_K(Q)|/|K|$. As $|K| = |HN_K(Q)| = \frac{|H||N_K(Q)|}{|H \cap N_K(Q)|}$, we have

$$|N_G(Q)| = \frac{|G|}{|K|}|N_K(Q)| = \frac{|G|}{|H|}|H \cap N_K(Q)| = m|H \cap N_K(Q)|,$$

since $n = p^r$. To show the claim, we must show that $|H \cap N_K(Q)| = 1$. To do this we first show that $H \cap N_K(Q) \subset Z(K)$; and then we shall show that $Z(K) = 1$. Let $z \in H \cap N_K(Q)$. As $K = HQ$, there exist $h \in H$, $y \in Q$ such that $z = hy$. As H is abelian, $zh = hz$. Since $z \in N_K(Q)$ and H is normal in G, we have $(z^{-1}y^{-1}z)y = z^{-1}(y^{-1}zy) \in Q \cap H = 1$. Therefore, $z \in Z(K)$. Now we show that $Z(K) = 1$. Since $Z(K)$ is a characteristic subgroup of K and K is normal in G, $Z(K)$ is normal in G by Exercise 11.9(22). If $1 < Z(K)$, then $Z(K)$ contains a minimal normal subgroup of G. But then $H \subset Z(K)$ by assumption. As $K = HQ$ and $H \subset Z(K)$, it follows that $Q \triangleleft K$. This then implies that $H \subset Q$. As $q \neq p$, this is a contradiction. Therefore, we have $H \cap N_K(Q) = 1$ and the claim is proven.

(2) and (4): We must show that if A is another Hall m-subgroup of G, then A and $N_G(Q)$ are conjugate. Let K be as above, then $m \mid |AK|$ and $p^r q^s = |K| \mid |AK|$, so we have $|AK| = |G| = mn$. It follows that the subgroup $AK = G$. Therefore, $G/K = AK/K \cong A/(A \cap K)$ and $|A \cap K| = q^s$. So $A \cap K$ is a Sylow q-subgroup of K. Hence $A \cap K$ and Q are conjugate in K by the Second Sylow Theorem. By Lemma 23.7, $N_G(Q)$ and $N_G(A \cap K)$ are conjugate in G. By the claim, $m = |N_G(Q)| = |N_G(A \cap K)|$. Since $A \cap K \triangleleft A$, we have $A \subset N_G(A \cap K)$. It follows that $A = N_G(A \cap K)$, as both are of order m. Consequently, $N_G(Q)$ and A are conjugate subgroups of G. This proves (2). Since the p^r conjugate subgroups of order m constitute all subgroups of order m, (4) follows.

(3): Let A_0 be a subgroup of G of order m' with $m' \mid m$ and A be a Hall m-subgroup of G. We have $A \cap A_0H$ is a subset of A_0H and $G = AH$. By Dedekind's Modular Law

17.7, we have
$$A_0H = A_0H \cap AH = (A_0H \cap A)H,$$
as $H \subset A_0H$. Therefore, $|A_0H \cap AH| = |(A_0H \cap A)H|$. It follows that. $|A_0H \cap H| = |A_0| = m'$. By property (2) for A, we have $A_0H \cap H$ is conjugate to A_0, so a Hall m-subgroup. □

Remark 23.9. The conclusions of the theorem for a finite group are usually false. For example, A_5 is simple [by Abel's Theorem – to be proven in Theorem 24.13]. It can be shown that A_5 has no subgroup of order 15 violating (1) and contains a subgroup of order 6 not contained in a subgroup of order 12, violating (3). Also $|\operatorname{Syl}_5(A_5)| = 6 = 3 \cdot 2$ which violates (4). Condition (2) is also violated in general. Let $H = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. It can be shown that $G = \operatorname{Aut}(H)$ is a simple group of order 168 permuting the seven groups of order two in H transitively and also the seven subgroups of order four in H transitively. Therefore, G has two distinct conjugate sets of subgroups of index 7 and order 24, so violates property (2).

An interesting special case of this theorem needs the following definition.

Definition 23.10. Let G be a finite group of order $p^r n$ with p a prime not dividing n. A Hall n-subgroup of G, if it exists, is called a *p-complement* of G. We also call such a p-complement (if it exists) a p'-subgroup of G. So, by definition, a p'-group if it exists, is a Hall group of order $\frac{|G|}{p^r}$.

Corollary 23.11. *If G is a finite solvable group, p a prime dividing the order of G, then G has a p-complement.*

We now turn to another characterization of solvable groups. This will follow by proving the converse to Corollary 23.11. This is also due to Philip Hall and is the second key result in this section. To prove the theorem, we need Burnside's Theorem 114.6 that any group of order $p^a q^b$, p, q primes, is solvable. The proof of this theorem needs the representation theory of groups that we shall do later. We shall, however, assume its validity here.

Theorem 23.12. *Let G be a nontrivial finite group. Suppose for every prime p dividing $|G|$ that G has a p-complement. Then G is solvable.*

Proof. Suppose the result is false. We may assume that G has been chosen so that G is a counterexample of minimal order. We first show that such a G must be a simple group. Suppose that G is not a simple group, then there exists a normal subgroup N of G with $1 < N < G$. Let p be a prime dividing $|G|$ and H a p'-subgroup. Then $H \cap N$ and HN/N are p'-subgroups of N and G/N, respectively. By the minimality of $|G|$, the groups N and G/N are solvable. Hence G is solvable by Theorem 17.3, a contradiction. Therefore, we can assume that G is simple. We show that this leads to a contradiction.

Let $|G| = p_1^{e_1} \cdots p_r^{e_r}$ with $p_1, \dots, p_r$ distinct primes and $e_i > 0$ for $i = 1, \dots, r$. By Burnside's Theorem 114.6, we have $r > 2$. By assumption, there exist p_i'-subgroups G_i of G for $i = 1, \dots, r$. Let $H = G_3 \cap \cdots \cap G_r$. Then we have $[G : G_i] = p_i^{e_i}$, hence $[G : H] = p_3^{e_3} \cdots p_r^{e_r}$ and $H = p_1^{e_1} p_2^{e_2}$ by Exercise 6(6). In particular, H is solvable by Burnside's Theorem 114.6. Let M be a minimal normal subgroup of H, so either a p_1- or p_2-subgroup, say a p_1-subgroup. Since $[G : H \cap G_2] = p_2^{e_2} \cdots p_r^{e_r}$, we have $|H \cap G_2| = p_1^{e_1}$, i.e., $H \cap G_2$ is a Sylow p_1-subgroup of H. It follows that $M \subset H \cap G_2 \subset G_2$. The same counting argument shows that $|H \cap G_1| = p_2^{e_2}$. Therefore, we have $G = (H \cap G_1)G_2$ as sets. Since M is normal in H, we have

$$\begin{aligned} N_G(M) =& \{g^{-1}xg \mid x \in M,\, g \in G\} \\ =& \{(hg_2)^{-1}xhg_2) \mid x \in M,\, h \in H \cap G_1, g_2 \in G_2\} \\ =& \{(g_2)^{-1}yg_2) \mid y \in M,\, g_2 \in G_2\} \subset G_2. \end{aligned}$$

Therefore, $N_G(M)$ is a proper nontrivial normal subgroup of G. This contradicts G being simple. □

Putting all this together, we have:

Theorem 23.13. *Let G be a finite solvable group. Then the following are equivalent:*

(1) G is solvable.
(2) Let n be a positive integer satisfying $n \mid |G|$ with n relatively prime to $|G|/n$. Then G contains a Hall n-subgroup.
(3) G has a p-complement for every prime p dividing the order of G.

Recall from Exercise 13.7(3), that a group G is called an (internal) semidirect product of a normal subgroup N of G and a subgroup H of G if $G = NH$ and $N \cap H = 1$. [This is isomorphic to the (external) semidirect product of Exercise 11.9(16).] This means that the group homomorphism $\theta : H \to \mathrm{Aut}(N)$ by $h \mapsto \theta_h$ (conjugation by h on N), satisfies $nhn'h' = n\theta_h(n')hh'$.

Therefore, we have

Corollary 23.14. *Let G be a finite solvable group. Suppose that there exists a normal Hall n-subgroup of G with n coprime to $\frac{|G|}{n}$. Then G is the semidirect product of N and a Hall $\frac{|G|}{n}$-subgroup of G.*

One can ask whether there are other interesting results concerning the existence of Hall subgroups.

Remark 23.15. Let G be a finite group of order mn with m and n relatively prime. The Schur-Zassenhaus Theorem is a fundamental result in group theory. It says if there exists a normal Hall m-subgroup K of G, then there exists a Hall n-subgroup H of G. In particular, G is a semidirect product of K and H. The Schur-Zassenhaus Theorem uses group cohomology in its proof, so will not be done here. Using the Feit-Thompson Theorem that finite groups of odd order are solvable, one then also proves that any two Hall n-subgroups in the Schur-Zassenhaus Theorem are conjugate and any subgroup of H of G of order dividing n is a subgroup of some Hall n-subgroup of G.

Exercises 23.16. 1. Prove Lemma 23.7

2. (Wielandt) Suppose that G is a finite group having three solvable groups H_1, H_2, and H_3 satisfying $[G : H_1]$, $[G : H_2]$, $[G : H_3]$ are pairwise relatively prime. Prove that G is solvable.
3. Let G be a finite group. Suppose every nontrivial subgroup H of G has a subgroup of index p for every prime dividing G. Show that G is solvable.

4. Let G be a group of order 12. Suppose that G is not isomorphic to A_4. Show that G contains a normal Sylow 2-subgroup or a normal Sylow 3-subgroup. Classify all groups of order 12 up to isomorphism.

24. The Symmetric and Alternating Groups

Recall if $\sigma \in S_n$, we write σ as

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

where the top row is the elements of the domain and the bottom row the corresponding values. For example, if

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$$

then

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \quad \text{and} \quad \sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

We want to write permutations in other ways.

Definition 24.1. A permutation $\alpha \in S_n$ is called a *cycle* of *length* r, or simply an r*-cycle*, if there exist distinct elements $a_1, \dots, a_r$ in $\{1, \dots, n\}$ (so $r \leq n$) satisfying

(1) $\alpha(a_i) = a_{i+1}$ for $i = 1, \dots, r-1$.
(2) $\alpha(a_r) = a_1$.
(3) $\alpha(a) = a$ for all $a \in \{1, \dots, n\} \setminus \{a_1, \dots, a_r\}$.

We denote the above r-cycle by $(a_1 \cdots a_r)$. If $r > 1$, we say that α is *nontrivial*.

Examples 24.2. (*i*) $(a_1) = 1$ the identity on $\{1, \dots, n\}$.

(*ii*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} = (23)$.

(*iii*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (34)(12) = (12)(34)$.

(*iv*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} = (1342)$.

(*v*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} = (1243)$.

The basic properties of cycles are given by:

Properties 24.3. Let $\alpha = (a_1 \cdots a_r)$ in S_n with $n \geq 2$. Then

(1) $\alpha = (a_i a_{i+1} \cdots a_r a_1 \cdots a_{i-1})$ if $1 \leq i \leq r$.
(2) α has order r.
(3) $\alpha^{-1} = (a_r \cdots a_1)$.
(4) If $\sigma \in S_n$, then $\sigma\alpha\sigma^{-1} = \big(\sigma(a_1) \cdots \sigma(a_r)\big)$.

PROOF. We leave the proof of (1) and (3) as an exercise.

(2): $\alpha^i(a_1) = a_{i+1} \neq a_1$ for $1 \leq i < r$, so the order of α is at least r. It is easy to check that $\alpha^r(a_i) = a_i$ for $1 \leq i \leq r$, so the order of $\alpha = r$.

(4): we have

$$\sigma\alpha\sigma^{-1}(j) = \begin{cases} \sigma(a_{i+1}) & \text{if } \sigma^{-1}(j) = a_i \text{, i.e., } j = \sigma(a_i) \text{ for } i < r. \\ \sigma(a_1) & \text{if } \sigma^{-1}(j) = a_r \text{, i.e., } j = \sigma(a_r). \\ j \ \big(= \sigma\sigma^{-1}(j)\big) & \text{otherwise.} \end{cases}$$

□

We say that two cycles $\alpha = (a_1 \cdots a_r)$ and $\beta = (b_1 \cdots b_l)$ are *disjoint* if $\{a_1, \ldots, a_r\} \cap \{b_1, \ldots, b_l\} = \emptyset$.

Lemma 24.4. *If α and β are disjoint cycles, then α and β commute, i.e., $\alpha\beta = \beta\alpha$.*

PROOF. $\beta(a_i) = a_i$ for all i and $\alpha(b_j) = b_j$ for all j. □

We now use the evaluation action in Example 21.19, which we recall.

Construction 24.5. Suppose that $S = \{1, \ldots, n\}$ and $\gamma \in S_n$ with $n > 1$. Let $\Gamma = \langle\gamma\rangle \subset S_n$. Then S is a Γ-set via the evaluation

$$* : \Gamma \times S \to S \text{ defined by } \gamma^j * a = \gamma^j(a).$$

The orbit of $a \in S$ is $\Gamma * a = \{\gamma^j(a) \mid j \in \mathbb{Z}\}$. Since $|\Gamma| \leq |S_n| = n! < \infty$, just as in the proof of the Classification of Cyclic Groups 9.9, there exists a least positive integer $m = m(a)$ satisfying $\gamma^m(a) = a$. Therefore, the orbit of a is $\Gamma * a = \{a, \gamma(a), \ldots, \gamma^{m-1}(a)\} \subset S$. Associate to this orbit, the m-cycle $\gamma_a = \big(a\gamma(a) \cdots \gamma^{m-1}(a)\big)$ in S_n. Let $\mathcal{O}$ be a system of representatives for the equivalence relation $\sim_\Gamma$ arising from the Γ-action on S and let $b \in \{1, \ldots, n\}$. There exists an $a \in \mathcal{O}$ satisfying $b \in \Gamma * a$ and $\gamma(b) = \gamma_a(b)$. By the Mantra of G-actions, $S = \bigvee_{\mathcal{O}} \Gamma * a$. Since disjoint cycles commute, we have $\gamma = \prod_{\mathcal{O}} \gamma_a$, i.e., γ is a product of disjoint cycles with the product unique up to order. We call this the (*full*) *cycle decomposition* of γ.

For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 4 & 1 & 6 \end{pmatrix} = (135)(2)(4)(6)$$

and

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 5 & 3 & 4 & 6 \end{pmatrix} = (12)(354)(6).$$

Note (again) that a 1-cycle (a) is the identity 1_S. Therefore, the *fixed points* of the Γ-action (i.e., elements a in S satisfying $\gamma(a) = a$) are precisely the elements of S occurring in the 1-cycles (if any) in a cycle decomposition of γ. We usually delete 1-cycles in the cycle decomposition of a permutation, e.g., $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 4 & 1 & 6 \end{pmatrix}$ is written (135). We shall call the cycle decomposition *full* if we do not delete the 1-cycles.

By Properties 24.3(4), we have the following useful observation:

Remark 24.6. If $\prod_{\mathcal{O}} \gamma_a$ is a cycle decomposition for γ in S_n, then $\prod_{\mathcal{O}} \sigma\gamma_a\sigma^{-1}$ is a cycle decomposition for $\sigma\gamma\sigma^{-1}$. (Cf. Exercise 24.24(5).)

We turn to another representation of permutations. We call a 2-cycle a *transposition*. Note if $i \neq j$ then $(ij)^{-1} = (ij)$ has order 2 as needed. We first show that transpositions generate S_n (for $n > 1$).

Proposition 24.7. *Let $n > 1$, then every element in S_n is a product of the transpositions $(12), (13), \ldots, (1n)$.*

PROOF. We first show that S_n is generated by transpositions. If $\sigma = (a_1 \ldots a_r)$, then

$$\sigma = (a_{r-1}a_r) \cdots (a_2a_r)(a_1a_r) \tag{24.8}$$

is a product of $r - 1$ transpositions.
To show that S_n is generated by the specific transpositions in the proposition, we use Properties 24.3(4) and (5), respectively, to see that

$$\begin{aligned} &(1i) = (i1) = (i1)^{-1} && \text{if } i \neq 1 \\ &(ij) = (1i)(1j)(1i)^{-1} = (1i)(1j)(1i) && \text{if } 1,\ i, \text{ and } j \text{ are distinct.} \end{aligned}$$

The result follows using the cycle decomposition of permutations. □

Remarks 24.9. If $\sigma \in S_n$, with $n > 1$, then σ is not a product of a unique number of transpositions. However, by equation (11.6), we have group homomorphisms

$$S_n \xrightarrow{\theta} \mathrm{Perm}_n(\mathbb{R}) \xrightarrow{\det} \{\pm 1\},$$

given by $\sigma \mapsto [T_\sigma]_{\mathcal{S}_n} \mapsto \det[T_\sigma]_{\mathcal{S}_n}$, where $\mathcal{S}_n$ is the standard basis for $\mathbb{R}^n$. The first map is an isomorphism, the second an epimorphism. So we have the alternating group $A_n := \ker \det \circ\, \theta \triangleleft S_n$ and $S_n/A_n \cong \{\pm 1\} \cong \mathbb{Z}/2\mathbb{Z}$ by the First Isomorphism Theorem. In particular, we have $[S_n : A_n] = 2$. If σ is a transposition, $\det[T_\sigma]_{\mathcal{S}_n} = -1$ (as it corresponds to interchanging two columns of the identity matrix). It follows that if $\tau \in S_n$ is a product of r transpositions and a product of s transpositions, then $(-1)^r = (-1)^s$, i.e., $r \equiv s \mod 2$. It follows by equation (24.8), if σ is an r-cycle with r odd, then $\sigma \in A_n$ and if r is even, then $\sigma \in S_n \setminus A_n$. In particular, if $n \geq 3$, then every 3-cycle lies in A_n. Using the cycle decomposition of any permutation, we conclude:

$$A_n = \{\sigma \in S_n \mid \sigma \text{ is a product of an even number of transpositions}\}$$

and

$$S_n \setminus A_n = \{\sigma \in S_n \mid \sigma \text{ is a product of an odd number of transpositions}\}.$$

Elements in A_n are called *even permutation* and elements in $S_n \setminus A_n$ are called *odd permutations*. So if τ is an odd permutation, we have $S_n = A_n \vee \tau A_n$.

Actually, one must be a bit careful in the above so that we do not get a circular argument. This would depend on the proof of the determinant, whose existence we have not proven. One can prove the properties of the determinant so that one does not have a circular argument, the problem being that interchanging two rows of a square matrix changes the sign of its determinant. [Cf. Section 121 for a sophisticated proof of the determinant and its properties.]

The possible place for the circular reasoning arises from the definition of the *signum* map. For completeness, we shall give a proof for this. The standard proof uses polynomials, and although simple seems out of place, so we give a group theoretic proof. Let σ be a permutation in S_n and suppose that $\sigma = \gamma_1 \cdots \gamma_r$ is its full cycle decomposition. Define the *signum* of σ by $\operatorname{sgn}(\sigma) = (-1)^{n-r}$. As a cycle decomposition is unique, this defines a map $\operatorname{sgn} : S_n \to \{\pm 1\}$.

Note if σ is an s-cycle, then in its full cycle decomposition, the number of 1-cycles is $n - s$, so $\operatorname{sgn} \sigma = (-1)^{s-1}$. In particular, if σ is a transposition, then $\operatorname{sgn}(\sigma) = -1$. The result that we need is that sgn is a group homomorphism. Clearly, the sgn of disjoint cycles is the product of the sgn of those cycles. As every permutation is a product of transpositions, the result now follows easily from the following:

Let $a, b, c_1, \dots, c_r, d_1, \dots, d_s$ be distinct elements in $\{1, \dots, n\}$ and $\tau = (ab)$. If $\sigma = (ac_1 \cdots c_r)$, then $\tau\sigma = (c_1 \cdots c_r ba)$, hence

$$\operatorname{sgn}(\tau\sigma) = (-1)^{r+1} = -\operatorname{sgn}(\sigma);$$

and if $\sigma = (ac_1 \cdots c_r bd_1 \cdots d_s)$, then $\tau\sigma = (ac_1 \cdots c_r)(bd_1 \cdots d_s)$, hence

$$\operatorname{sgn}(\tau\sigma) = (-1)^{r+s} = -(-1)^{r+s+1} = -\operatorname{sgn}(\sigma).$$

In particular, the parity of the number of transpositions in a decomposition of a permutation is invariant.

We turn to the study the alternating group. We know that 3-cycles are even. We start by showing that 3-cycles generate A_n. As A_n, $n = 1, 2$ is the trivial group, it is generated by the empty set of 3-cycles, so we may assume that $n \geq 3$.

Proposition 24.10. *The alternating group A_n is a normal subgroup of index two in S_n, and, if $n \geq 3$, then A_n is generated by the 3-cycles:*

$$(123),\ (124), \dots, (12n).$$

Proof. We need only show the last statement. We first show that A_n is generated by 3-cycles. We have shown that every element in A_n is a product of an even number of transpositions, so it suffices to show a product of two distinct transpositions is a product of 3-cycles. Suppose that i, j, k, l are distinct elements in $\{1, \dots, n\}$. Then this follows by the equations

$$\begin{aligned}(ij)(ik) &= (ikj)\\ (ij)(kl) &= (kil)(ijk).\end{aligned}$$

To finish, we show that the alleged 3-cycles generate. Let i, j, k be distinct elements in $\{1, \dots, n\}$, and different from 1 and 2. Then, using Property 24.3, this follows from the equations:

$$\begin{aligned}(ijk) &= (12i)(2jk)(12i)^{-1}\\ (2jk) &= (12j)(12k)(12j)^{-1}\\ (1jk) &= (12k)^{-1}(12j)(12k).\end{aligned}$$

□

Remark 24.11. We make the following remark to simplify notation below. Let $n \geq 3$ and $K \subset A_n$ a subgroup. If $\sigma \in S_n$, then $\sigma K \sigma^{-1} \subset \sigma A_n \sigma^{-1} = A_n \triangleleft S_n$. In particular, K contains a 3-cycle if and only if $\sigma K \sigma^{-1}$ does.

The key to Abel's Theorem on the simplicity of the alternating group A_n for $n \geq 5$, is the following:

Lemma 24.12. *Let K be a normal subgroup of A_n. If K contains a 3-cycle, then $K = A_n$.*

Proof. As K contains a 3-cycle, we have $n \geq 3$. By the proposition, it suffices to show that $(12k)$ lies in K for $k = 3, \ldots, n$. By Remark 24.11, we may assume that (123) lies in K, hence $(213) = (123)^{-1}$ lies in K. Let $\sigma = (12)(3k)$ with $k > 3$, a product of an even number of transpositions, so an element of A_n. As $K \triangleleft A_n$, we have

$$(12k) = \sigma(213)\sigma^{-1} \text{ lies in } K, \text{ for all } k \geq 3$$

as needed. □

We now prove Abel's Theorem. This gives our first examples of non-abelian simple groups.

Theorem 24.13. (Abel's Theorem) *Let $n \neq 4$, then the alternating group A_n is simple.*

Proof. We know that A_2 and A_3 are simple, so assume that $n \geq 5$, and $1 < K \triangleleft A_n$. We must show $K = A_n$. By the lemma, it suffices to show that K contains a (single) 3-cycle.
Let $e \neq \alpha \in K$ be chosen such that

$$\alpha \text{ moves precisely } m \text{ elements in } \{1, \ldots, n\}$$

and

$$\begin{aligned}&\text{no other non-identity element in } K\\&\text{moves a fewer number of elements,}\end{aligned}$$

where we say α *moves* i, if $\alpha(i) \neq i$, i.e.,

$$\begin{aligned}m \ &\text{ is the minimal number of elements moved}\\&\text{by any non-identity element in } K.\end{aligned}$$

Let

$$\alpha = \gamma_1 \cdots \gamma_r$$

be the decomposition of α into a product of disjoint cycles. We may assume that γ_1 is a k-cycle with k maximal among the lengths of $\gamma_1, \ldots, \gamma_r$.
As α is not the identity, we must have $m \geq 2$. If $m = 2$, then we must have $\alpha = \gamma_1$ and be a transposition, hence it is an odd permutation, so not an element of A_n, a contradiction. If $m = 3$, then we must have $\alpha = \gamma_1$ and be a 3-cycle, and we are done by the lemma. So we may assume that $m \geq 4$. For clarity, it is convenient to changing notation if necessary. So, using Remark 24.11, we may assume that

$$\begin{aligned}\alpha \ &\text{ moves } 1, 2, 3, 4.\\\alpha \ &\text{ moves } 5 \text{ if } m \geq 5.\\\gamma_1 \ &= (12 \cdots k).\end{aligned}$$

We have two cases:

Case 1. $k \geq 3$: If this is the case, then $m \geq 5$:

For if not, then we have $m = 4$, and we must have $\alpha = \gamma_1 = (1234)$ and be a 4-cycle which is not in A_n, a contradiction.

Case 2. $k = 2$: Then each γ_i is a transposition or a 1-cycle, and we may assume, using Remark 24.11, that $\alpha = (12)(34)\cdots$:

Set $\beta = (345)$ in A_n. As $K \triangleleft A_n$, we have $\alpha_1 = \beta\alpha\beta^{-1}$ lies in K and

$$\alpha_1 = \beta\alpha\beta^{-1} = \begin{cases} (124\cdots)\cdots & \text{in Case 1.} \\ (12)(45)\cdots & \text{in Case 2.} \end{cases}$$

In either case, we check that $\alpha(3) \neq \alpha_1(3)$ (e.g., in Case 1, if $k = 3$, then $\alpha_1(3) \neq 1$ and if $k \geq 4$, then $\alpha_1(3) \neq 4$), so $e \neq \alpha^{-1}\alpha_1$ lies in K. Since β fixes every $i > 5$, we have $\alpha^{-1}\alpha_1$ fixes every $i > 5$ that α fixes. If $\alpha(5) \neq 5$, then we have $m \geq 5$ and $\alpha(1) = \alpha_1(1)$. It follows that $\alpha^{-1}\alpha_1$ moves at most $m - 1$ elements, contradicting the minimality of m. Therefore, we may assume that $\alpha(5) = 5$, i.e., $m = 4$. In particular, we are in Case 2 and have by computation that $\alpha^{-1}(345)\alpha_1(345)^{-1} = (354)$. This contradicting the maximality of k. The result follows. $\square$

Remark 24.14. If G is a group containing a non-abelian simple group, then G is not solvable by Theorem 17.3. Therefore, S_n is not solvable for any $n \geq 5$.

We wish to show that A_5 is the only simple group of order 60 up to isomorphism (non-abelian as its order is not a prime). It is the first non-abelian simple group by computation, using the Sylow Theorems. We make the general observation:

Lemma 24.15. *Let $n \geq 5$, then A_n is the only subgroup of S_n of index two.*

Proof. Suppose that K is a normal subgroup of G of index two and $K \neq A_n$. By the Second Isomorphism Theorem, we conclude A_nK is a group with $K < A_nK$ normal and $A_nK/K \cong A_n/(A_n \cap K)$. As $A_n < A_nK \subset S_n$, we must have $S_n = A_nK$, so $2 = [A_n : A_n \cap K]$. In particular, $1 < A_n \cap K < A_n$. It follows that $A_n \cap K$ is a proper normal subgroup of the simple group A_n, a contradiction. $\square$

Theorem 24.16. *Up to isomorphism, the alternating group A_5 is the only simple group of order 60.*

Proof. Let G be a simple group of order 60. We prove this in three steps.

Step 1. If G contains a subgroup H of index 5, then $G \cong A_5$:

Let $\lambda : G \to \Sigma(G/H)$ given by $x \mapsto (\lambda_x : gH \mapsto xgH)$ be the homomorphism defined in the General Cayley Theorem 12.4. Since $H < G$, we have $\ker(\lambda) \subset H < G$. Consequently, λ must be monic (since not the trivial map), as G is simple. Therefore, G is isomorphic to $\lambda(G) \subset \Sigma(G/H)$, a subgroup of index two. By the lemma $G \cong A_5$.

Step 2. G does not contain a subgroup of index three:

Suppose that H is a subgroup of index three in G. Then $|G| \nmid [G : H]!$, which implies that G is not simple by Useful Counting 12.8.

Step 3. Finish:

By the first two steps, we may assume that G has no subgroup of index 3 or index 5. Let P be a Sylow 2-subgroup of G. Then P is not a normal subgroup of G by assumption, hence by the Third Sylow Theorem, we have $[G : N_G(P)] = 1+2k \mid 3\cdot 5$, for some integer $k \geq 1$. By the first two steps, we must have $[G : N_G(P)] = 15$, so $|N_G(P)| = 4$. (Note that this means that $P = N_G(P)$.) Let $Q \neq P$ be another Sylow 2-subgroup of G, and set $I = P \cap Q$. We may also assume that Q and P were chosen such that $I = P \cap Q$ has maximal cardinality.

Case 1. $1 < I$, so $|I| = 2$:

In this case, we have, by Exercise 10.16(5) (cf. equation (13.5)),

$$|PQ| = |P|\,|Q|/|P \cap Q| = |P|\,|Q|/|I| = 2^3.$$

As both P and Q are groups of order 2^2, they are both abelian, so normalize I. Consequently, $P, Q \subset N_G(I)$. It follows that

$$2^3 = |PQ| \leq |N_G(I)| \mid 2^2\cdot 3\cdot 5 \quad \text{and} \quad 2^2 \mid |N_G(I)|.$$

Therefore, $|N_G(I)| = 2^2\cdot 3, 2^2\cdot 5$, or $2^2\cdot 3\cdot 5$. By the first two steps, we conclude that $|N_G(I)| = 2^2\cdot 3\cdot 5 = |G|$. Hence $I \triangleleft N_G(I) = G$, contradicting the simplicity of G.

Case 2. $I = 1$:

As Q and P were chosen with $I = P \cap Q$ of maximal cardinality, and each Sylow 2-subgroup contains $(4-1) = 3$ non-identity elements, the Sylow 2-subgroups of G contain $45 = 15(4-1)$ distinct non-identity elements. By the simplicity of G, the number of the Sylow 5-subgroups, which divides $2^2\cdot 3$, must be at least six. Each Sylow 5-subgroup of G is cyclic of order 5, so G must contain at least $6\cdot(5-1) = 24$ distinct elements of order 5, which implies that $|G| \geq 45 + 24 > 60$, a contradiction. □

Remarks 24.17. Let F be a field, then the group

$$\mathrm{PSL}_n(F) := \mathrm{SL}_n(F)/Z\big(\mathrm{SL}_n(\mathrm{F})\big)$$

is called a *projective special linear group* of F. It can be shown that $\mathrm{PSL}_n(F)$ is simple for all $n \geq 3$.

We shall see later that for all primes p and $q = p^n$, with $n \in \mathbb{Z}^+$, there exists a field $\mathbb{F}_q$ with q elements, and unique up to isomorphism of fields [definition?]. Computation shows that

$$|\mathrm{PSL}_2(\mathbb{F}_q)| = \begin{cases} (q+1)(q^2-q) & \text{if } q = 2^n \\ \frac{1}{2}(q+1)(q^2-q) & \text{if } q = p^n \text{ with } p \text{ and odd prime} \end{cases}$$

(cf. Exercise 9.12(1)). We will show that $\mathrm{PSL}_2(\mathbb{F}_q)$ is simple if and only if $q > 3$ in Section 25. The computation shows that $\mathrm{PSL}_2(\mathbb{F}_4)$ and $\mathrm{PSL}_2(\mathbb{F}_5)$ both have order 60, so are isomorphic to A_5. The simple group $\mathrm{PSL}_2(\mathbb{F}_7)$ of order 168 is the next non-abelian simple group up to isomorphism. Computation shows that the groups A_8 and $\mathrm{PSL}_3(\mathbb{F}_4)$ both have $8!/2$ elements. It can also be shown that they are not isomorphic, i.e., there exist non-isomorphic simple groups of the same order.

Our results and exercises have shown that groups of orders $p^s q$ ($s \geq 1$), p^2q^2, and pqr are not simple when p, q, and r are primes. It follows that these groups are therefore solvable, by induction and reduction to simpler cases. Burnside showed that groups of order $p^a q^b$ are solvable for any primes p and q for over a hundred years ago. The original proof did not use abstract group theory (but group representation theory). A group theoretic proof was finally found, but is much more difficult than this original proof. More spectacularly, Feit-Thompson proved that any group of odd order is solvable. The proof is very difficult, and was the first major step in classifying finite simple groups. Groups of even order are, therefore, much harder to understand. We give two examples when we can say something more.

The first example, is left as an exercise (cf. the proof of Lemma 24.15):

Proposition 24.18. *Let $n \geq 5$. If H is a normal subgroup of S_n, then either $H = 1$, $H = A_n$, or $H = S_n$.*

The order of A_5 is divisible by 4. We show groups of even order larger than two but not divisible by 4 are never simple, i.e., if G is a group of order $2n$, with $n > 1$ an odd integer, then G is not simple.

Review 24.19. Let G be a nontrivial finite group. We have seen that Cayley's Theorem gives a group monomorphism

$$\lambda : G \to \Sigma(G) \text{ defined by } x \mapsto (\lambda_x : a \mapsto xa)$$

called the (*left*) *regular representation* of G. If $xa = \lambda_x(a) = a$, then $x = e$, i.e., the permutation λ_a has a fixed point if and only if $a = e$. Note also that the regular representation corresponds to the translation action of G on $S = G$ which is transitive, i.e., S is the only orbit under this action. In particular, this action has no fixed points (one-point orbits).

We want to prove that every finite group of order $2n$ with $n > 1$ odd is not simple. We need two lemmas which themselves are interesting.

Lemma 24.20. *Let G be a group of order mn and $\lambda : G \to \Sigma(G)$ the regular representation with $m > 1$. If the element a in G has order m, then*

(1) λ_a is a product of n disjoint m-cycles.
(2) λ_a is an odd permutation if and only if m is even and n is odd.

[Note. This characterizes elements in $\lambda(G)$, but not in $\Sigma(G)$.]

Proof. (1): By the review, we know if $1 \leq r < m$, then the permutation $\lambda_{a^r} = (\lambda_a)^r$ has no fixed points. Let $\lambda_a = \gamma_1 \cdots \gamma_s$ be a full cycle decomposition of λ_a. As λ_{a^r} has no fixed points for $1 \leq r < m$, each γ_i must be a cycle of order at least m. But $\lambda_{a^m} = 1_G$, so each γ_i must be a cycle of length m. [Note: γ_i^r may no longer be a cycle, but the γ_i^r are still disjoint and each is a product of disjoint cycles.] Since each element of G occurs precisely once in the cycle decomposition for λ_a, we must have $s = n$.

By (1), we have a full cycle decomposition of λ_a given by $\lambda_a = \gamma_1 \cdots \gamma_n$, with each γ_i an m-cycle. Therefore, λ_a is an odd permutation if and only if $n(m-1)$ is odd, i.e., if and only if n is odd and m is even. $\square$

Lemma 24.21. *Let G be a finite group and $\lambda : G \to \Sigma(G)$ the regular representation. If $\lambda(G)$ contains an odd permutation, then there exists a normal subgroup of G of index two.*

Proof. Let $|G| = n$ and identify $\Sigma(G)$ and S_n. As $\lambda(G)$ contains an odd permutation, $H := A_n \cap \lambda(G) < \lambda(G)$ and $\lambda(G)A_n = \Sigma(G)$. The regular representation is monic, so, $G \cong \lambda(G)$. Using the Correspondence Principle and the Second Isomorphism Theorem, we must have

$$\begin{aligned}[G : \lambda^{-1}(H)] &= [\lambda(G) : H] = [\lambda(G) : A_n \cap \lambda(G)] \\ &= [A_n\lambda(G) : A_n] = [S_n : A_n] = 2.\end{aligned}$$

Therefore, $\lambda^{-1}(H) < G$ is a subgroup of index two hence is also normal. □

Application 24.22. Let G be a finite group of order $2n$ with n odd. Then G contains a normal subgroup of index two. In particular, if $n > 1$, G is not simple.

Proof. By Cauchy's Theorem, there exists an element a in G of order two. Let $\lambda : G \to \Sigma(G)$ be the regular representation. By Lemma 24.20, the permutation λ_a is a product of n disjoint transpositions, hence is an odd permutation. The result now follows by Lemma 24.21. □

This application can be pushed a bit further. Since all Sylow p-subgroups of a group are conjugate, if one is abelian (respectively, cyclic) all are.

Theorem 24.23. *Let G be a finite group of order $2^r m$ with m odd. If G contains a cyclic Sylow 2-subgroup then there exists a normal subgroup of G of index 2^r. In particular, if $m > 1$ or $r > 1$, then G is not simple.*

Proof. We may assume that $m > 1$ and $r \geq 1$. Let $\lambda : G \to \Sigma(G)$ be the regular representation. By Lemma 24.20, the group $\lambda(G)$ contains an odd permutation, so by Lemma 24.21, there exists a normal subgroup N of G of index two. Hence $|N| = 2^{r-1}m$, and we may assume that $r > 1$. Let $P_0 \in \mathrm{Syl}_2(N)$. By the Fourth Theorem, there exists $P \in \mathrm{Syl}_2(G)$ containing P_0. By the remark above, P is cyclic, say $P = \langle x \rangle$, so $P_0 = \langle x^2 \rangle$ by the Cyclic Subgroup Theorem. By induction on r, there exists a normal subgroup H in N of index 2^{r-1}. In particular, $|H| = m$. By the Second Isomorphism Theorem, $P_0H \subset N$ is a subgroup, and P_0 and H have relatively prime orders, so $P_0 \cap H = 1$ (why?). It follows by counting that $N = P_0H$. Let $H^x := xHx^{-1}$. We have $G = N \vee xN$, since $x \notin N$. Therefore, $H \triangleleft G$ if and only if $H^x = H$, as $H \triangleleft N$. If this is the case, we are done, so we may assume that $H \neq H^x$. The group N being normal in G implies that $N^x = xNx^{-1} = N$. Since $N = P_0H$ and $P_0^x = xP_0x^{-1} = P_0$ in cyclic P, we have $N = P_0H = N^x = (P_0H^x) = P_0^xH^x$, so $H^x \subset N$. As $|H| = |H^x|$, we know that $2 \not| \, |H^x|$. By the Second Isomorphism Theorem, as $H \triangleleft N$ and $H^x \subset N$, we have HH^x is a subgroup of N. In particular, $|H| \mid |HH^x|$. But we also have $2 \not| \, |H||H^x|/|H \cap H^x| \mid |HH^x|$, so we must have $|HH^x| \leq |N|/2^{r-1} = |H|$, hence $|HH^x| = |H| = |H^x|$. We conclude that $H = H \cap H^x = H^x$, a contradiction. Therefore, $H \triangleleft G$. □

Exercises 24.24.

1. Let α be a product of disjoint cycles $\gamma_1, \dots, \gamma_m$. Show that the order of α is the least common multiple of the orders of the γ_i, $1 \leq i \leq m$. In particular, show that this means that the order of each γ_i divides the order of α and further, that if $p > 1$ is a prime, then every power of a p-cycle is either a p-cycle of the identity.

2. Let $n \geq 3$. Show that the center $Z(S_n) = 1$. We call a group *centerless* if its center is trivial.
3. Let p be a prime and G a subgroup of the symmetric group S_p containing a p-cycle and a transposition. Show that $G = S_p$.
4. Show that S_4 and $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ are not isomorphic.
5. We say that an element $\alpha \in S_n$ has *cycle type* $(r_1, \dots, r_m)$ if α is a product of disjoint cycles $\gamma_1, \dots, \gamma_m$ with γ_i an r_i-cycle for $1 \leq i \leq m$ and $r_1 \leq \cdots \leq r_m$. Prove that two elements of S_n are conjugate in S_n if and only if they have the same cycle type.
6. A permutation in S_n is called *regular* if it is the identity or has no fixed points and is the product of disjoint cycles of the same length. Show that a permutation is regular if and only if it is a power of an n-cycle.
7. Let α be an r-cycle. If $k > 0$ is an integer and (r, k) the gcd of r and k, show that α^k is a product of (n, k) disjoint cycles, each of length $r/(r, k)$.
8. Let $\varphi : S_n \to S_n$ be a group automorphism that takes every transposition to a transposition. Prove that φ is an inner automorphism.
9. Show the following:
   (a) A_4 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^3 = 1$.
   (b) S_4 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^4 = 1$.
   (c) A_5 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^5 = 1$.
10. Show that A_4 is the only subgroup of index two in S_4.
11. Show That any subgroup of order 12 not isomorphic to A_4 contains a cyclic subgroup of order 6.
12. Show that A_5 is generated by 5-cycles. Generalize.
13. Show that S_n has no normal subgroups of order two if $n \geq 5$.
14. Let $n \geq 5$. Show if H is a normal subgroup of S_n, then either $H = 1$, $H = A_n$, or $H = S_n$.
15. Let $n > 2$ be an integer. Show that every subgroup of index n in A_n is isomorphic to A_{n-1}.
16. Let p be a prime and H the cyclic subgroup generated by a p-cycle in S_p. Determine the order of the normalizer $N_{S_p}(H)$. Prove your determination is correct. Generalize this to the cyclic subgroup generated by an n-cycle in S_n.
17. Let G be a finite group and N a normal subgroup of G of prime index p. Prove that $[G, G] \subset N$ and use it to prove that S_n, $n \geq 5$ is not polycyclic, hence not solvable without using Abel's Theorem.
18. Let $n \geq 5$. Prove that the only proper subgroup of S_n of index at most $n - 1$ is A_n.
19. Let G be a free group on basis $\mathcal{B}$ with $\mathcal{B}$ having at least two elements. Show that G is not solvable.
20. Show that the free product $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/5\mathbb{Z}$ (cf. Definition 18.13 and Theorem 18.14) is not solvable.

21. Let $A_\infty := \bigcup_{i=1}^{\infty} A_n$ and $S_\infty := \bigcup_{i=1}^{\infty} S_n$ be groups defined in the obvious way (with A_n and S_n respective subgroups for all n). Show that A_∞ is a normal subgroup of index two in S_∞ and $1 < A_\infty < S_\infty$ is a composition series. Also show the infinite abelian subgroup of S_∞ generated by the transpositions $(2k-1\,2k)$, $k \geq 1$, does not have a composition series.(Cf. Exercise 17.14(11).)
22. Classify all groups of order $2pq$ up to isomorphism if $p < q$ are odd primes with $p \not| \, q-1$.
23. Classify all groups of order $2pq$ up to isomorphism if $p < q$ are odd primes.

25. Addendum: The Projective Special Linear Group

In this section, we shall produce another infinite collection of finite simple groups, viz. $\mathrm{SL}_2(F)/Z(\mathrm{SL}_2(F))$, with F a finite field and $|F| > 3$. Do so, we shall need some results from other sections in the book that we shall assume as facts in this section.

We need the following results about fields.

Facts 25.1. (Lagrange) *Let F be a field. Then a nonzero polynomial $f \in F[t]$ of degree n has at most n roots in F.* (Cf. Corollary 34.8)

Most of the following Facts will be exercises arising as examples of our results in the study of polynomials over a field.

Fact 25.2. *Let F be a finite field.*

1. *There exist a prime $p > 0$ and $q = p^n$, some n, such that $|F| = q$. In particular, F is an n-dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$.*
2. *Let $|F| = q$. Then $F^\times$ is a cyclic group of order $q-1$.*(Cf. Theorem 34.15.)
3. *If $|F| = q$, then every element of F is a root of the polynomial $t^q - t \in F[t]$ and every element of $F^\times$ is a root of $t^q - 1$.* (Cf. Fermat's Little Theorem.)
4. *For every prime p, there exists a field F having p^n elements.*
5. *Given $q = p^n$, $p > 0$ a prime, all fields of order q are isomorphic as fields.* (What is the definition of a field?) *We shall denote a choice for such a finite field of q elements by $\mathbb{F}_q$. We may assume also that $\mathbb{F}_q$ contains $\mathbb{Z}/p\mathbb{Z}$, e.g., $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.*
(Warning: $\mathbb{Z}/p^n\mathbb{Z}$ is not a field if $n > 1$).

We shall also need a special case of the following from linear algebra that classifies square matrices over an arbitrary field up to similarity: Let F be a field. If h is the monic polynomial $h = t^n + a_{n-1}t^{n-1} + \cdots + a_0 \in F[t]$, let C_h denote the matrix

$$C_h := \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & \cdots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}$$

in $\mathbb{M}_d(F)$. It is called the *companion matrix* of h.

Facts 25.3. *If $A \in \mathbb{M}_n(F)$, there exist unique monic polynomials $q_1 \mid q_2 \mid \cdots \mid q_r$ in $F[t]$* (where polynomial $f \mid g$ in $F[t]$ if $g = fh$ in $F[t]$, $g \in F[t]$), *satisfying $B = PAP^{-1}$ with*

$P \in GL_n(F)$ and

$$B = \begin{pmatrix} C_{q_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{q_r} \end{pmatrix} \text{ (in block form).}$$

The matrix C is unique and called the rational canonical form of A. (Cf. Section 45).

We shall only need this fact in the special case that $n = 2$ and $A \in GL_2(F)$. The two possibilities for the rational canonical form C that can occur for such an A is C is an invertible matrix with C either a diagonal matrix (2 blocks) or of the form $\begin{pmatrix} 0 & -a_0 \\ 1 & -a_1 \end{pmatrix}$.

Definition 25.4. If R is a commutative ring, a matrix $E = (e_{ij}) \in \mathrm{GL}_n(R)$ is called an *elementary matrix (of Type I)* or a *transvection* if if there exists $0 \neq \lambda \in R$ and $l \neq k$ such that

$$e_{ij} = \begin{cases} 1 & \text{if } i = j \\ \lambda & \text{if } (i,j) = (k,l) \\ 0 & \text{otherwise.} \end{cases}$$

We denote E by $E_{ij}(\lambda)$. Transvections lie in $\mathrm{SL}_n(R) := \{A \in \mathrm{GL}_n(R) \mid \det A = 1\}$, the special linear group of R of degree n. They will play the role that 3-cycles played in A_n, $n \geq 5$.

Note. If $B \in R^{m\times n}$ (respectively $B \in R^{n\times m}$), then multiplying B on the left (respectively right) by a transvection E is just adding a multiple of a row (respectively column) of B to another row (respectively column) of B. (Cf. Gaussian elimination.)

Proposition 25.5. *Let R be a commutative ring. Then the centralizer of $\mathrm{SL}_n(R)$ in $\mathrm{GL}_n(R)$ is the subgroup of invertible scalar matrices $\{aI \mid a \in R^\times\}$.*

PROOF. Let $A = (a_{ij}) \in \mathrm{SL}_n(R)$. As $\{aI \mid a \in R^\times\}$ centralizes $\mathrm{GL}_n(R)$, we need only show the converse. Suppose that $i \neq j$. Then $AE_{ij}(1) = E_{ij}(1)A$. Hence $A(E_{ij}(1) - I) = (E_{ij}(1) - I)A$ (where I is the identity matrix as usual). As the kjth entry of $A(E_{ij}(1) - I)$ is a_{ki} and the kjth entry of $(E_{ij}(1) - I)A$ is zero if $k \neq i$ and a_{jj} otherwise, it follows that $a_{ki} = 0$ if $k \neq i$ and $a_{ii} = a_{jj}$. The result follows. $\square$

Corollary 25.6. *Let R be a commutative ring. The center of $\mathrm{GL}_n(R)$ is $\{aI \mid a \in R^\times\}$ and the center of $\mathrm{SL}_n(R)$ is $\{aI \mid a^n = 1,\, a \in R^\times\}$.*

Using row and column operations on a matrix over a field (or over a polynomial ring over a field using the division algorithm (cf. Appendix D, Theorem D.2), we deduce the following:

Corollary 25.7. *Let F be a field. Then any element $A \in \mathrm{GL}_n(R)$ can be written as $A = UD(\mu)$ with $U \in \mathrm{SL}_n(R)$ a product of transvections and $D(\mu) = (d_{ij})$ a diagonal matrix with $d_{nn} = \mu \in \mathbb{R}^\times$ and all the other diagonal entries one. In particular, $\mathrm{SL}_n(F)$ is generated by transvections.*

If R a commutative ring, then the subgroup $\mathrm{SL}_n(R)$ is normal in $\mathrm{GL}_n(R)$. However, conjugates of a normal subgroup N of $\mathrm{SL}_n(R)$ by elements of $\mathrm{GL}_n(R)$ may not lie in N. Of course, it may happen that some conjugates of N by an element in $GL_n(R)$ may still lie in N. We give an illustration of this that will be useful.

Construction 25.8. Let F be a field and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathrm{SL}_2(F)$.

Suppose that there exists an $S \in \mathrm{GL}_2(F)$ satisfying $SAS^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & r \end{pmatrix}$ with $\det S = \mu$. Set $T = D(\mu)^{-1}S$. We have $T \in \mathrm{SL}_2(F)$ and

$$\begin{pmatrix} 0 & -1 \\ 1 & r \end{pmatrix} = D(\mu)T \begin{pmatrix} a & b \\ c & d \end{pmatrix} T^{-1}D(\mu)^{-1}.$$

or

$$(*) \qquad D(\mu)^{-1} \begin{pmatrix} 0 & -1 \\ 1 & r \end{pmatrix} D(\mu) = T \begin{pmatrix} a & b \\ c & d \end{pmatrix} T^{-1}.$$

The left hand side of (*)

$$\begin{pmatrix} 1 & 0 \\ 0 & \mu^{-1} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mu \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ \mu^{-1} & \mu^{-1}r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mu \end{pmatrix} = \begin{pmatrix} 0 & -\mu \\ \mu^{-1} & r \end{pmatrix}.$$

As $T = D(\mu)^{-1}S \in \mathrm{SL}_2(F)$, we have A is conjugate to $\begin{pmatrix} 0 & -\mu \\ \mu^{-1} & r \end{pmatrix}$ in $\mathrm{SL}_2(F)$ by an element in $\mathrm{SL}_2(F)$. In particular, if N is a normal subgroup of $\mathrm{SL}_2(F)$ and $A \in N$, then the constructed matrix $\begin{pmatrix} 0 & -\mu \\ \mu^{-1} & r \end{pmatrix}$ lies in N.

Definition 25.9. Let R be a commutative ring. Define the *projective general linear group* by $\mathrm{PGL}_n(R) := \mathrm{GL}_n(R)/Z(\mathrm{GL}_n(R))$ and the *projective special linear group* by $\mathrm{PSL}_n(R) := \mathrm{SL}_n(R)/Z(\mathrm{SL}_n(R))$.

So we have shown the following:

Corollary 25.10. *Let R be a commutative ring. Then $Z(\mathrm{SL}_n(R)) = \mathrm{SL}_n(R) \cap Z(\mathrm{GL}_n(R))$.*

Notation 25.11. Let $q = p^n$, $p > 0$ a prime, and $n > 0$. We shall denote the groups $\mathrm{GL}_n(\mathbb{F}_q)$, $\mathrm{SL}_n(\mathbb{F}_q)$, $\mathrm{PSL}_n(\mathbb{F}_q)$ by $\mathrm{GL}(n,q)$, $\mathrm{SL}(n,q)$, $\mathrm{PSL}(n,q)$, respectively.

Over a finite field, we know the cardinality of these groups.

Lemma 25.12. *Let $p > 0$ be a prime and $q = p^n$ with $n > 1$. Then we have*

(1) $|\mathrm{GL}(n,q)| = (q^n - 1)(q^n - q)\cdots(q^n - q^{n-1})$.
(2) $|\mathrm{SL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1) = |\mathrm{PGL}(n,q)|$.
(3) $|\mathrm{PSL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1)(n, q-1)$.

Proof. We leave a proof of (1) and (2) to the reader. (Cf. Exercise 9.12(1).) As for (3), we use the fact that if F is a finite field of cardinality q, then $F^\times$ is a cyclic group by Fact 25.2(2). For such an F, every element x in $F^\times$ satisfies $x^{q-1} = 1$ and the number of x with $x^n = 1$ is the gcd $(n, q-1)$. The result follows. $\square$

Our goal is to show that $\mathrm{PSL}(2,q)$ is a simple group if $q > 3$. To do so, it suffices to show that every proper noncentral normal subgroup of $\mathrm{SL}(2,q)$ is simple whenever $q > 3$. We first show that transvections play the role analogous to that played by 3-cycles in the proof of the simplicity of A_n, $n \geq 5$.

Lemma 25.13. *Let $N \triangleleft \mathrm{SL}(2,q)$. If N contains a transvection, then $N = \mathrm{SL}(n,q)$.*

PROOF. By Corollary 25.7, it suffices to show that N contains all the tranvections in $\mathrm{SL}(n,q)$. Suppose that N contains the transvection $E_{12}(\mu)$ (so $\mu \neq 0$). Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an element of $\mathrm{SL}(2,q)$. Since $E_{12}(\mu) \in N \triangleleft \mathrm{SL}(2,q)$, we have

$$(*) \qquad AE_{21}(\mu)A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1-\mu ac & \mu a^2 \\ -\mu c^2 & 1+\mu ac \end{pmatrix}$$

also lies in N. In particular if $c = 0$, we have $E_{21}(-\mu a^2)$ lies in N and if $a = 0$, we have $E_{12}(-\mu a^2)$ lies in N.

Now consider the map of multiplicative groups $\mathbb{F}_q^\times \to \mathbb{F}_q^\times$ by $x \mapsto x^2$. It is a group homomorphism with kernel $\{x \mid x^2 = 1\}$. In particular if $p = 2$, this kernel is trivial and if $p > 2$, it is $\{\pm 1\}$. It follows that in either case more than half the elements in $\mathbb{F}_q$ are squares. Consequently, the additive subgroup $H = \{x \in \mathbb{F}_q \mid E_{ij}(x) \in N\} \cup \{0\}$ of $\mathbb{F}_q$ must be $\mathbb{F}_q$ as its index in $\mathbb{F}_q$ is less than two. Therefore, by the special case of (*) with $c = 0$ and $a = 0$ we have $E_{12}(\lambda)$ lies in N for all $\lambda \in \mathbb{F}_q^\times$. As $U = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ lies in $\mathrm{SL}(2,q)$, we also have

$$E_{21}(\lambda) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -\lambda \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = UE_{12}U^{-1}$$

lies in N for all nonzero λ. The result follows. □

Theorem 25.14. (Jordan-Moore) *The groups $\mathrm{PSL}(2,q)$ are simple if and only if $q > 3$.*

PROOF. By Lemma 25.12, we have

$$\mathrm{PSL}(2,q) = \begin{cases} (q+1)(q^2-q) & \text{if } q = 2^n. \\ \frac{1}{2}(q+1))(q^2-q) & \text{if } q = p^n,\ p \text{ an odd prime.} \end{cases}$$

As $|\,\mathrm{PSL}(2,2)| = 6$ and $|\,\mathrm{PSL}(2,3)| = 12$, we know that $\mathrm{PSL}(2,2)$ and $\mathrm{PSL}(2,3)$ are not simple. So it suffices to show that $\mathrm{PSL}(2,q)$ is simple if $q > 3$.

Let $N \triangleleft \mathrm{SL}(2,q)$ be a noncentral subgroup and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in N$ not in the center of $\mathrm{SL}(2,q)$. We first show that N contains a transvection in the special case when $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ satisfies $b = 0$, and $a \neq \pm 1$.

Let $S = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Then N contains the commutator

$$(*) \qquad SAS^{-1}A^{-1} = \begin{pmatrix} 1 & 0 \\ 1-d^2 & 1 \end{pmatrix}.$$

As $\det A = ad = 1$, if $a \neq \pm 1$, we would have $E_{21}(1-a^2)$ lies in N and we would be done for the case that $b = 0$, and $a \neq \pm 1$.

From the general form of A, we shall construct a matrix in N of the form $\begin{pmatrix} \alpha & 0 \\ \beta & \alpha^{-1} \end{pmatrix} \in N$ not in the center of $\mathrm{SL}(2,q)$ with $\alpha \neq \pm 1$, $\beta \in \mathbb{F}_q$. This would finish the proof.

By Fact 25.3, there exists a matrix $P \in \mathrm{GL}(2,q)$ such that $C := PAP^{-1}$ is in rational canonical form. As $\mathrm{SL}(2,q) \triangleleft \mathrm{GL}(2,q)$, the matrix C lies in $\mathrm{SL}(2,q)$. In the case at hand, this rational canonical form has one or two blocks.

Case 1. The matrix C is diagonal, i.e., there are two blocks.

If $C = \begin{pmatrix} y & 0 \\ 0 & x \end{pmatrix}$, then $y = x^{-1}$ as $C \in \mathrm{SL}(2,q)$. As A was assumed not to be in the center, neither is C. So $y \neq \pm 1$. Replacing A by C in (*), we see that N contains a transvection.

Case 2. The matrix C has one block, i.e., it is the companion matrix of a monic polynomial in $\mathbb{F}_q[t]$ of degree two.

Suppose that $C := \begin{pmatrix} 0 & y \\ 1 & x \end{pmatrix}$.

Since $C \in \mathrm{SL}(2,q)$, we must have $y = -1$. As $P \in \mathrm{GL}(2,q)$, by Construction 25.8, there exists a matrix $D = \begin{pmatrix} 0 & -\mu^{-1} \\ \mu & x \end{pmatrix}$ that is a conjugate of C by an element in $\mathrm{SL}(2,q)$ for some nonzero element μ. Hence the matrix D lies in $N \triangleleft \mathrm{SL}(2,q)$.

Let $\alpha \in \mathbb{F}_q^{\times}$ and $T = \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \alpha \end{pmatrix}$ in $\mathrm{SL}(2,q)$. Then N contains the commutator

$$(\dagger) \qquad U = TDT^{-1}D^{-1} = \begin{pmatrix} \alpha^{-2} & 0 \\ \mu x(\alpha^2 - 1) & \alpha^2 \end{pmatrix}$$

in $\mathrm{SL}(2,q)$. Therefore, we are done if $\alpha^{-2} \neq \pm 1$. i.e., $\alpha^4 \neq 1$. The polynomial $t^4 - 1 \in \mathbb{F}_q[t]$ has at most four roots in $\mathbb{F}_q$ by Fact 25.1 and every root of $\mathbb{F}_q$ is a root of $t^5 - t \in \mathbb{F}_q[t]$ by Fact 3. It follows that $\mathbb{F}_q = \mathbb{Z}/5\mathbb{Z}$ is the only case left. Note in this case, As 1 and -1 are the only distinct nonzero squares in $\mathbb{Z}/5\mathbb{Z}$, $\alpha^2 = \pm 1$, so we must look more closely. There are two possible subcases.

Subcase 1. The element $x \neq 0$.

In $\mathbb{Z}/5\mathbb{Z}$, there exists an element $\alpha^2 - 1 \neq 0$, e.g., $\alpha = 2$ in $\mathbb{Z}/5\mathbb{Z}$. Therefore, $\mu x(\alpha^2 - 1) \neq 0$ in ($\dagger$) and $U^2 = E_{21}(-2\mu x(\alpha^{-2}))$ lies in N.

Subcase 2. The element $x = 0$.

We know that $D = \begin{pmatrix} 0 & -\mu^{-1} \\ \mu & x \end{pmatrix}$ lies in N. Let $z \in \mathbb{F}_q^{\times}$. Then

$$\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \mu^{-1} \\ \mu & x \end{pmatrix} \begin{pmatrix} 1 & -z \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} z\mu & -z^2\mu - \mu^{-1} \\ \mu & -\mu z \end{pmatrix}$$

lies in N. Setting $z = 2\mu^{-1}$, we see that $\begin{pmatrix} 2 & 0 \\ \mu & 2 \end{pmatrix}$

lies in N. $\square$

It is, in fact, true that $\mathrm{PSL}_n(F)$ with $n \geq 3$ is simple for any field R is simple. We shall not prove this. A key lemma that is used to prove this is the following:

Lemma 25.15. *Let $n \geq 3$. Then any two transvections in $\mathrm{SL}_n(F)$ are conjugate.*

We leave this as an exercise. It can be proved using conjugation of transvections by good choices of elements in $\mathrm{SL}_n(F)$.

Exercises 25.16. 1. Prove Corollary 25.7.

2.

3. Let F be a field. Show that $\mathrm{GL}_n(F)$ is a semidirect product of $\mathrm{SL}_n(f)$ be $F^{\times}$.

4. Prove Lemma 25.15

Part 3

Ring Theory

CHAPTER V

General Properties of Rings

In this chapter, we begin our formal study of rings, especially that of commutative rings. We find the analogues in ring theory for the notions and basic theorems we learned in group theory. The analogue in ring theory of a normal subgroup is called an ideal. As in group theory, this turns out to be those subsets of a ring that are kernels of ring homomorphisms. This leads to analogous isomorphism theorems and an analogous correspondence principle.

When studying the integers, we looked at integers themselves. In this chapter, we switch our perspective from studying elements to studying ideals. We discover the type of ideal corresponding to a prime integer in commutative rings. This type of ideal, called a *prime ideal*, is the major object of study in commutative ring theory whose definition is based upon Euclid's Lemma. (For elements the analogue of a prime integer is called an *irreducible element*.) We also introduce an important generalization of the ring of integers called a principal ideal domain or PID. Although this ring is a very special type of domain (a commutative ring in which $ab = 0$ implies $a = 0$ or $b = 0$), it will be the main type of domain that we shall study. However, we shall show that even for a general domain, there exists a *field of fractions* or *quotient field*, thus constructing, in particular, the field of rational numbers. We shall also find the ring analogue of the Chinese Remainder Theorem that we have previously proved for the integers. Finally, we shall define the set theoretic extension of finite induction used in algebra, called Zorn's Lemma. This axiom, equivalent to the Axiom of Choice that you may have seen, is a powerful tool. [A proof of the equivalence of these two axioms can be found in Appendix A.] We shall then give some simple applications of it that are basic to ring theory and linear algebra.

26. Definitions and Examples

We begin by recalling the definition of a ring, together with specific types of rings.

Definition 26.1. Let R be a set with two binary operations

$$+ : R \times R \to R \qquad \text{and} \qquad \cdot : R \times R \to R.$$

We call R a *ring* (under $+$ and $\cdot$) if:

(1) $(R, +)$ is an additive group (with identity written 0 or 0_R).

(2) $(R, \cdot)$ is a monoid (with identity written 1 or 1_R).

(3) R satisfies the *distributive laws*, i.e., for all $a, b, c \in R$, we have:
   (a) $a \cdot (b + c) = a \cdot b + a \cdot c$.
   (b) $(b + c) \cdot a = b \cdot a + c \cdot a$.

If (2) is replaced by

(2′) $(R, \cdot)$ satisfies associativity (i.e., possibly no 1),

then (following Jacobson), we call R (under $+$ and $\cdot$) a *rng*. [Take out the condition of having an identity, take out the i from ring.]
A ring satisfying

(4) $(R \setminus \{0\}, \cdot)$ is a group

is called a *division ring*.
As usual if $a, b \in R$, we usually write ab for $a \cdot b$.

Also recall that if R is a ring with $1 \neq 0$, then

$$\begin{aligned} R^\times &:= \{x \in R \mid \text{There exists } x^{-1} \in R \text{ such that } xx^{-1} = 1 = x^{-1}x\} \\ &= \{x \in R \mid \text{There exists } a, b \in R \text{ such that } ax = 1 = xb\} \end{aligned}$$

is called the *group of units* of R. [Why is it a group?] In particular, a ring is a division ring if and only if $R^\times = R \setminus \{0\}$ (cf. Property 26.4(3c) below).

Definition 26.2. A ring R is called a *commutative ring* if for all $a, b \in R$, we have

$$ab = ba.$$

A commutative ring R is called an *(integral) domain* if

(i) $0 \neq 1$.
(ii) Let $a, b \in R$. If $ab = 0$, then either $a = 0$ or $b = 0$.

A commutative division ring is called a *field*.

Remark 26.3. Check that a field is a domain.

Rings satisfy the following basic properties (which we leave as an exercise):

Properties 26.4. Let R be a rng and $a, b, c \in R$. Then the following are true:

(1) $a \cdot 0 = 0 = 0 \cdot a$.
(2) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$.
(3) If R is a ring, then for all $a \in R$:
  (a) $(-1) \cdot a = -a$.
  (b) $(-1) \cdot (-1) = 1$.
  (c) $|R| > 1$ if and only if $1 \neq 0$.
(4) If R is a commutative ring with $1 \neq 0$, then R is a domain if and only if R satisfies the *Cancellation Law*: If a, b, c lie in R, then

$$a \cdot b = a \cdot c \text{ with } a \neq 0 \text{ implies } b = c.$$

We already know many examples of rings. We recall some of these.

Examples 26.5. 1. A *trivial ring* is a ring with one element. By the above, this is the case if and only if $1 = 0$ in R.

2. $\mathbb{Z}$ is a domain.

3. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields.

4. The subset of $\mathbb{C}$ given by
$$\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Z}\}$$
is a domain under the restriction of the $+$ and $\cdot$ of $\mathbb{C}$ called the *Gaussian integers*. If $z \in \mathbb{Z}[\sqrt{-1}]$, there exist unique integers a and b satisfying $z = a + b\sqrt{-1}$. (Cf. the next example.)
5. The subset of $\mathbb{C}$ given by
$$\mathbb{Q}[\sqrt{-1}] = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Q}\}$$
is a field under the restriction of the $+$ and $\cdot$ of $\mathbb{C}$. [Can you show this?] Note that if $z \in \mathbb{Q}[\sqrt{-1}]$, there exist unique rational numbers a and b satisfying $z = a + b\sqrt{-1}$. Why?
6. $\mathbb{Z}/n\mathbb{Z}$, with $n \in \mathbb{Z}^+$, is a commutative ring. It is a domain if and only if n is a prime if and only if it is a field.
7. If R is a ring so is $\mathbb{M}_n(R)$. If R is a commutative ring with $|R| > 1$, then $\mathbb{M}_n(R)$ is commutative if and only if $n = 1$. (If R is a rng then so is $\mathbb{M}_n(R)$ — obvious definition.)
8. If R is a ring, then $R[t]$, the set of polynomials with coefficients in R, is a ring under the usual $+$ and $\cdot$ of polynomials.
[What are the usual $+$ and $\cdot$ of polynomials? What is $1_{R[t]}$?]
9. Let R be a ring. Define the *ring of (formal) power series* over R to be the set
$$R[[t]] := \{\sum_{i=0}^{\infty} a_i t^i \mid a_i \in R \text{ for all } i\}$$
with properties and operations:
$$0_{R[[t]]} = 0_R \text{ and } 1_{R[[t]]} = 1_R, \text{ with } t^0 = 1_R$$
$$\sum_{i=0}^{\infty} a_i t^i = \sum_{i=0}^{\infty} b_i t^i \text{ if and only if } a_i = b_i \text{ for all } i$$
$$\sum_{i=0}^{\infty} a_i t^i + \sum_{i=0}^{\infty} b_i t^i := \sum_{i=0}^{\infty} (a_i + b_i) t^i$$
$$\sum_{i=0}^{\infty} a_i t^i \cdot \sum_{i=0}^{\infty} b_i t^i := \sum_{i=0}^{\infty} c_i t^i \text{ with } c_i = \sum_{j=0}^{i} a_j b_{i-j} \text{ for all } i.$$
Since we do not have to worry about convergence, it is easy to define power series (inductively) to satisfy conditions.

Just as for groups, we need to define the maps of interest between rings.

Definition 26.6. A map $\varphi : R \to S$ of rings is called a *ring homomorphism* if for all $a, b \in R$, the map φ satisfies:

(1) $\varphi(0_R) = 0_S$.
(2) $\varphi(a + b) = \varphi(a) + \varphi(b)$.
(3) $\varphi(1_R) = 1_S$.
(4) $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$.

I.e., φ is both a group homomorphism $\varphi : (R, +) \to (S, +)$ and a 'monoid homomorphism' $\varphi : (R, \cdot) \to (S, \cdot)$. Note that we know that (1) is automatic for group homomorphisms. [If we omit (3), we call φ a *rng homomorphism*.]
A ring homomorphism $\varphi : R \to S$ is called

i. a *ring monomorphism* or *monic* or *mono* if φ is injective.
ii. a *ring epimorphism* or *epic* or *epi* if φ is surjective.
iii. a *ring isomorphism* if φ is bijective with inverse a ring homomorphism.
iv. a *ring automorphism* if $R = S$ and φ is a ring isomorphism.

Of course, if R and S are fields, etc., then φ is called a *field homomorphism*, etc.
If $\varphi : R \to S$ is a ring homomorphism, we let

$$\ker \varphi := \{x \in R \mid \varphi(x) = 0_S\} \subset R$$

called the *kernel* of φ, and

$$\operatorname{im} \varphi := \{\varphi(x) \mid x \in R\} \subset S$$

called the *image* of φ.

Remarks 26.7. 1. A ring homomorphism $\varphi : R \to S$ is a ring isomorphism if and only if φ is bijective. [Proof?]

2. We say two rings R and S are *isomorphic* if there exists an isomorphism $\varphi : R \to S$. If this is the case, we write $R \cong S$ and often write the isomorphism as $\varphi : R \xrightarrow{\sim} S$.
3. A ring homomorphism $\varphi : R \to S$ is monic if and only if $\ker \varphi = 0 := \{0\}$.
4. Let $n \in \mathbb{Z}^+$, then the canonical map $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is a ring epimorphism with $\ker {}^- = n\mathbb{Z} = \{ny \mid y \in \mathbb{Z}\}$.

Remark 26.8. Let $\varphi : R \to S$ be a ring homomorphism. We leave it as an exercise to show that φ is a monomorphism if and only if given any ring homomorphisms $\psi_1, \psi_2 : T \to R$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$. (Cf. Exercise 1.13(7).) However, the analogue of Exercise 1.13(8) is false. In particular, suppose that a ring homomorphism $\varphi : R \to S$ has the property that whenever there exists ring homomorphisms $\theta_1, \theta_2 : S \to T$ with compositions satisfying $\theta_1 \circ \varphi = \theta_2 \circ \varphi$ that $\theta_1 = \theta_2$. Then this does not necessarily imply that φ is surjective. For example the inclusion map of $\mathbb{Z}$ into $\mathbb{Q}$ is a ring homomorphism that satisfies this condition, but is not a surjective ring homomorphism. (Cf. Exercise 27.20(8).) In category theory, being epi is defined by this condition. For this reason, it is no longer common to define a ring epimorphism as a surjective ring homomorphism. However, for consistency of notation, we shall use this older definition, i.e., call a surjective ring homomorphism a ring epimorphism.

We also need the analogues from group theory of subgroups and normal subgroups.

Definition 26.9. Let R be a ring, $S \subset R$ a subset with S a ring. We say S is a *subring* of R if the inclusion map $inc : S \to R$ (often written $inc : S \hookrightarrow R$) is a ring monomorphism, so the restriction of the identity map on R to S is the inclusion map of S into R with $1_S = 1_R$ and S is *closed* under the ring operations $+$ and $\cdot$ on R, meaning that the restriction of the operations on R to S have image in S, viz., $+|_{S\times S} : S \times S \to S$ and $\cdot|_{S\times S} : S \times S \to S$. [We also have the obvious notion of a *subrng* of a rng, if we ignore, in this case, the condition on the 1's, if any.]

Examples 26.10. 1. $\mathbb{Z} \subset \mathbb{Q}$ and $\mathbb{Z}, \mathbb{Q}, \mathbb{R} \subset \mathbb{C}$ are subrings.

2. If R is a domain, e.g., a field, and $S \subset R$ a subring, then S is a domain.
3. If R is not the zero ring, then $0 \subset \mathbb{Z}$ is not a subring. It is a subrng.
4. Let $\mathbb{Z} \times \mathbb{Z}$ be a ring under componentwise $+$ and $\cdot$, so $1_{\mathbb{Z}\times\mathbb{Z}} = (1_{\mathbb{Z}}, 1_{\mathbb{Z}})$ and $\mathbb{Z} \times \mathbb{Z}$ is commutative ring, but not a domain. We also have $\mathbb{Z} \times 0 := \mathbb{Z} \times \{0\} \subset \mathbb{Z} \times \mathbb{Z}$ is not a subring but is a subrng.
5. Let R be a ring. The *center* of R is the set

$$Z(R) := \{a \in R \mid ax = xa \text{ for all } x \in R\}.$$

It is a commutative subring of R. For example, the center of a division ring is a field.

Remark 26.11. The last example allows us to generalize our terminology. Let R be a commutative ring, A a ring and $\varphi : R \to A$ a ring homomorphism. We call A an *R-algebra* via φ if $\varphi(R) \subset Z(A)$. If B is another R-algebra via τ, then a ring homomorphism $\psi : A \to B$ is called an *R-algebra homomorphism* if

$$\begin{tikzcd} A \arrow[rr, "\psi"] & & B \\ & R \arrow[ul, "\varphi"] \arrow[ur, "\tau"'] & \end{tikzcd}$$

commutes. Such a τ is called an *R-algebra monomorphism* (respectively, an *R-algebra epimorphism R-algebra isomorphism*) if it is a ring monomorphism (respectively, epimorphism, isomorphism).

We need the analogue of normal subgroups in ring theory, i.e., those subobjects of a ring that are kernels of ring homomorphisms. If a ring homomorphism $\varphi : R \to S$ satisfies $\varphi(1_R) = 0_S$, then $S = 0$ as $\varphi(1_R) = 1_S$. In particular, $\ker\varphi$ cannot be a subring of R unless $\ker\varphi = R$; so in general kernels are not subrings. They are, however, always subrngs. Just as in the case when being a subgroup of a group is not sufficient to be the kernel of a group homomorphism, neither is being a subrng sufficient. Rather we need the following objects:

Definition 26.12. Let R be a ring and $\mathfrak{A}$ a nonempty subset of R. We call $\mathfrak{A}$ a (2-*sided*) *ideal* of R if $(\mathfrak{A}, +) \subset (R, +)$ is a subgroup and for all r in R and a in $\mathfrak{A}$, we have both ar and ra lie in $\mathfrak{A}$, i.e., $\cdot : R \times \mathfrak{A} \to \mathfrak{A}$ and $\cdot : \mathfrak{A} \times R \to \mathfrak{A}$. So an ideal of R is a subrng of R closed under multiplication by R on both the right and left. If $\mathfrak{A}$ only satisfies being a subrng and closed under multiplication by R on the left (respectively, on the right), we call $\mathfrak{A}$ a *left ideal* (respectively, *right ideal*). Of course, if R is a commutative ring, the notions of (2-sided) ideal, right ideal, and left ideal coincide, making life much simpler.

Examples 26.13. Let R be a nontrivial ring.

1. $0 = \{0\}$ is an ideal of R called the *trivial ideal* and R is an ideal of R called the *unit ideal*. If these are the only ideals of R, then we call R a *simple ring*. (Cf. simple groups.)

2. Let $a \in R$. Define
$$(a) = RaR := \{\sum_{i=1}^{n} x_i a y_i \mid n \in \mathbb{Z}^+,\ x_i, y_i \in R,\ 1 \leq i \leq n\},$$
an ideal of R called the *principal ideal generated* by a. It is the smallest ideal in R containing a. [Cf. this with a cyclic subgroup of a group.] If R is commutative, this is much simpler, as (using the distributive laws), it is just
$$(a) = Ra := \{ra \mid r \in R\}.$$

3. Let $a_1, \dots, a_n \in R$. The smallest ideal *generated* by $a_1, \dots, a_n$ is
$$\begin{aligned}(a_1, \dots, a_n) &= Ra_1R + \cdots + Ra_nR := \\ &\{\sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} a_i y_{ij} \mid x_{ij}, y_{ij} \in R \text{ all } i, j \text{ some } n, m_i \in \mathbb{Z}^+\}.\end{aligned}$$
If R is commutative, this is simply
$$(a_1, \dots, a_n) = Ra_1 + \cdots + Ra_n := \{\sum_{i=1}^{n} r_i a_i \mid r_i \in R,\ 1 \leq i \leq n\},$$
i.e., it is all *R-linear combinations* of $a_1, \dots, a_n$. [Cf. this with the span of finitely many vectors in a vector space.]
4. Let I be an (indexing) set, $A = \{a_i \mid i \in I\}$. Then the smallest ideal in R containing A is
$$\langle A \rangle = (a_i)_I := \{a \in R \mid \text{There exist } a_{i_1}, \dots, a_{i_n} \in A, \text{ some } n \text{ and } i_j \in I, \text{ satisfying } a \in (a_{i_1}, \dots, a_{i_n})\}.$$
We say that A *generates* I. [So $\langle A \rangle$ is the union of all the ideals generated by $a_{i_1}, \dots, a_{i_n}$ in A with $i_1, \dots, i_n \in I$, for some n.]
5. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Then
$$\begin{aligned}\mathfrak{A} + \mathfrak{B} &:= (\mathfrak{A} + \mathfrak{B}) = \{a + b \mid a \in \mathfrak{A} \text{ and } b \in \mathfrak{B}\}\\ \mathfrak{A} \cap \mathfrak{B} &:= \{x \mid x \in \mathfrak{A} \text{ and } x \in \mathfrak{B}\}\\ \mathfrak{A}\mathfrak{B} &:= (\{ab \mid a \in \mathfrak{A},\ b \in \mathfrak{B}\})\\ &= \{\sum_{i=1}^{n} a_i b_i \mid a_i \in \mathfrak{A},\ b_i \in \mathfrak{B},\ 1 \leq i \leq n, \text{ some } n\}\end{aligned}$$
are ideals in R. Moreover, we have
$$\mathfrak{A}\mathfrak{B} \subset \mathfrak{A} \cap \mathfrak{B} \subset \mathfrak{A} + \mathfrak{B}$$
(and they are usually all different).
Note the formal difference between the definition of $\mathfrak{A}\mathfrak{B}$ for rings and HK, with H, K subgroups of G, although this difference vanishes if H or K is a normal subgroup (the analogue of an ideal). So $\mathfrak{A}\mathfrak{B}$ is the ideal in R generated by ab with $a \in \mathfrak{A}$, $b \in \mathfrak{B}$.

More generally, we can recursively define the product of finitely many ideals. [Note that infinite products do not make sense in a ring.]

Note $\mathfrak{A} \cup \mathfrak{B}$ is usually not an ideal (cf. groups and vector spaces). Can you give a condition that will guarantee that it is?

6. If $\varphi : R \to S$ is a ring homomorphism, then $\ker \varphi$ is an ideal of R. But in general $\operatorname{im} \varphi$ is not an ideal of S. Indeed, it is an ideal if and only if φ is surjective. It is, however, a subring of S
7. If $\mathfrak{A}$ is an ideal of R and $\mathfrak{A} \cap R^{\times} \neq \emptyset$, then $\mathfrak{A} = R$, the unit ideal (and conversely). Indeed, if $x \in \mathfrak{A} \cap R^{\times}$ and $r \in R$, then $r = r1 = rx^{-1}x \in \mathfrak{A}$.

We now give some basic examples.

Specific Examples 26.14. Let R be a ring.

1. If R is a division ring, then R is simple. Indeed, if $0 < \mathfrak{A} \subset R$ is an ideal, then $\mathfrak{A} \cap R^{\times} \neq \emptyset$.
2. Suppose R is not the trivial ring. If R is commutative, then R is simple if and only if R is a field. Indeed if R is simple, $0 \neq a \in R$, then $(a) = R$; so there exists an $r \in R$ satisfying $1 = ar = ra$ showing a is a unit. The converse is the previous example.
3. $\mathbb{M}_n(\mathbb{R})$, for $n > 1$, is simple and not a division ring. In fact, if R is simple, so is $\mathbb{M}_n(R)$ which is never a division ring if $n > 1$. [We leave this as an exercise.]
4. Let $R = \mathbb{Z}$ and $0 < \mathfrak{A} \subset \mathbb{Z}$ an ideal. Then $(\mathfrak{A}, +) \subset (\mathbb{Z}, +)$ is a subgroup, so $\mathfrak{A} = \mathbb{Z}n$ for some $n \in \mathbb{Z}^+$ as an additive group. Clearly, $(n) = \mathbb{Z}n \subset \mathbb{Z}$ is an ideal. Therefore, every ideal in $\mathbb{Z}$ is of the form (n) for some $n \geq 0$ in $\mathbb{Z}$. In particular, every ideal in $\mathbb{Z}$ is principal, where an ideal $\mathfrak{A}$ is called *principal* if there exists an $a \in \mathfrak{A}$ such that $\mathfrak{A} = (a)$.
5. A domain R is called a *principal ideal domain* or a *PID* if every ideal in R is principal, e.g., $\mathbb{Z}$ is a PID.
6. We give further examples of PIDs (without proof):
   (*i*) Any field.
   (*ii*) $F[t]$, if F is a field. [Can you show this?]
   (*iii*) Let
   $$\mathbb{Z}[\theta] := \{a + b\theta \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}$$
   with
   $$\theta = \sqrt{-1},\ \sqrt{-2},\ \frac{-1-\sqrt{-3}}{2},\ \frac{-1-\sqrt{-7}}{2},\ \frac{-1-\sqrt{-11}}{2},$$
   $$\frac{-1-\sqrt{-19}}{2},\ \frac{-1-\sqrt{-43}}{2},\ \frac{-1-\sqrt{-67}}{2},\ \text{or}\ \frac{-1-\sqrt{-163}}{2}.$$
   Under the $+$ and $\cdot$ in $\mathbb{C}$, these are rings and each is a PID.

   Gauss had conjectured that these were the only PID's of this form (with negative square roots). It was proved by Harold Stark.
7. $\mathbb{Z}[\sqrt{-5}] := \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}$ is a domain but not a PID.
8. $\mathbb{Z}[t]$ is a domain but not a PID.

We begin our study by generalizing concepts that we investigated when studying the integers.

Definition 26.15. Let R be a commutative ring, a, b in R with $a \neq 0$. We say that a *divides* b and write $a \mid b$, if there exists an x in R satisfying $b = ax$.

Properties 26.16. Let R be a commutative ring and a, b, c elements in R.

(1) If $a \mid b$ and $a \mid c$, then $a \mid bx + cy$ for all x, y in R.

(2) If R is a domain, then $a \mid b$ and $b \mid a$ if and only if there exists a unit u in R satisfying $a = ub$.

Note the second property generalizes the corresponding result for the integers, as ± 1 are the only units in $\mathbb{Z}$.

Proof. The proof of the first property is the same as for the integers, so we need only prove the second. If $a = ub$ for a unit u then $b = u^{-1}a$. Conversely, if $a \mid b$ and $b \mid a$ then there exist $x, y \in R$ satisfying $b = ax$ and $a = by$. As $a \neq 0$ in the domain R and $a = by = axy$, we have $1 = xy$ by the Cancellation Law, so $x, y \in R^{\times}$. $\square$

Property (2) motivates the following definition: If R is a domain, a and b nonzero elements in R, we say that a is an *associate* of b and write $a \approx b$ if there exists a unit u in R such that $a = ub$. Thus a is an associate of b if and only if $a \mid b$ and $b \mid a$. [Check that $\approx$ is an equivalence relation.] For example, if $R = \mathbb{Z}$, then the only associates of n in $\mathbb{Z}$ are $\pm n$. (We also let 0 be the associate of 0.)

Remarks 26.17. Let R be a commutative ring. Then the relationship between division and principal ideal is given by the following: Let a, b, and c be nonzero elements in R.

1. $a \in (b)$ if and only if $b \mid a$ if and only if $(a) \subset (b)$.
2. $(a) = (b)$ if and only if $a \mid b$ and $b \mid a$.
3. $a \mid b$ and $a \mid c$ if and only if $(b, c) \subset (a)$.
4. $(b, c) = (a)$ if and only if $a \mid b$ and $a \mid c$ and there exist x and y in R satisfying $a = bx + cy$.
5. Suppose that R is a domain. Then
   (a) $a \approx b$ if and only if $(a) = (b)$.
   (b) If $a = rb$ in R, then $a \approx b$ if and only if $r \in R^{\times}$. In particular, if $a \not\approx b$ and $a = rb$ in R, then $(a) < (b)$. Indeed, if $ub = a = rb$ with $u \in R^{\times}$, then $u = r$ as R is a domain.

If $R = \mathbb{Z}$ and $p \neq \pm 1$ is an integer, we know that p is a prime if and only if $p \mid ab$ implies $p \mid a$ or $p \mid b$. This motivates the most important definition concerning ideals.

Definition 26.18. Let R be a commutative ring and $\mathfrak{A} < R$ an ideal. We call $\mathfrak{A}$ a *prime ideal* in R if

$$ab \in \mathfrak{A} \text{ implies that } a \in \mathfrak{A} \text{ or } b \in \mathfrak{A}.$$

It is called a *maximal ideal* in R if

$$\mathfrak{A} < \mathfrak{B} \subset R \text{ is an ideal, then } \mathfrak{B} = R.$$

[Note this definition of maximal ideal (even left maximal ideal — definition?) makes sense in any ring, commutative or not.]

Examples 26.19. Let R be a nontrivial commutative ring.

1. R is a domain if and only if $0 = (0)$ is a prime ideal in R:

   We have 0 is a prime ideal if and only if $ab \in (0)$ implies $a \in (0)$ or $b \in (0)$ if and only if $ab = 0$ implies $a = 0$ or $b = 0$.

2. Every maximal ideal $\mathfrak{m} < R$ is a prime ideal:

   Let $ab \in \mathfrak{m}$ with $a \notin \mathfrak{m}$. Then $\mathfrak{m} < \mathfrak{m} + Ra \subset R$ is an ideal, hence $R = \mathfrak{m} + Ra$. Therefore, there exist $m \in \mathfrak{m}$ and $r \in R$ satisfying $1 = m + ra$. Hence $b = bm + rab \in \mathfrak{m} + \mathfrak{m} \subset \mathfrak{m}$.

3. Suppose that R is a PID. Then every nonzero prime ideal in R is maximal:
   [Warning: This is not true in general, e.g., in $\mathbb{Z}[t]$ – why?]

   Let $0 < \mathfrak{p} < R$ be a prime ideal and suppose that $\mathfrak{p} < \mathfrak{A} \subset R$ is an ideal. By the definition of a PID, there exist $p \in \mathfrak{p}$ and $a \in \mathfrak{A}$ satisfying $\mathfrak{p} = (p)$ and $\mathfrak{A} = (a)$. Therefore, we have $p = ra$ for some $r \in R$ and $a \notin (p)$. As $ra = p \in \mathfrak{p}$, we have $r \in \mathfrak{p}$. Write $r = sp$, with $s \in R$. Then $p = ra = psa$ in the domain R, so $1 = sa$. Hence $a \in R^\times$ and $\mathfrak{A} = R$. This shows that $\mathfrak{p}$ is maximal.

4. The prime ideals in $\mathbb{Z}$ are:

$$\begin{aligned} 0 = (0) &\qquad \text{prime, not maximal} \\ p\mathbb{Z} = (p) &\qquad p \text{ a prime, maximal.} \end{aligned}$$

We want to view prime ideals as an ideal theoretical property rather than a property about elements in the ideal. This is easily done and is the direct analogue of Euclid's Lemma for ideals.

Lemma 26.20. *Let R be a commutative ring and $\mathfrak{p} < R$ an ideal. Then $\mathfrak{p}$ is a prime ideal if and only if for all ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R satisfying $\mathfrak{A}\mathfrak{B} \subset \mathfrak{p}$, either $\mathfrak{A} \subset \mathfrak{p}$ or $\mathfrak{B} \subset \mathfrak{p}$. Moreover, if $\mathfrak{A} < R$ is not a prime ideal, then there exists ideals $\mathfrak{A} < \mathfrak{B}_i$ in R with $i = 1, 2$ satisfying $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$.*

Proof. ($\Leftarrow$): If $xy \in \mathfrak{p}$ with $x, y \in R$ then $(x)(y) \subset \mathfrak{p}$, so either $x \in (x) \subset \mathfrak{p}$ or $y \in (y) \subset \mathfrak{p}$.

($\Rightarrow$): Suppose that the ideal $\mathfrak{A} \not\subset \mathfrak{p}$ but $\mathfrak{A}\mathfrak{B} \subset \mathfrak{p}$ with $\mathfrak{B}$ an ideal. Then there exists an element $a \in \mathfrak{A} \setminus \mathfrak{p}$. As $ab \in \mathfrak{p}$ for all $b \in \mathfrak{B}$, it follows that $\mathfrak{B} \subset \mathfrak{p}$.

As for the last statement, suppose that $\mathfrak{A} < R$ is not a prime ideal. Then there exist ideals $\mathfrak{B}_i \not\subset \mathfrak{A}$, $i = 1, 2$ but $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$. Let $\mathfrak{B}_i{}' = \mathfrak{B}_i + \mathfrak{A} > \mathfrak{A}$, then $\mathfrak{B}_1{}'\mathfrak{B}_2{}' \subset \mathfrak{B}_1\mathfrak{B}_2 + \mathfrak{A} \subset \mathfrak{A}$. □

Exercises 26.21.

1. Prove if R is a domain so is the ring of formal power series $R[[t]]$.
2. Let R be a commutative ring. Show that if $f = 1 + \sum_{i=1}^{\infty} a_i t^i$ is a formal power series in $R[[t]]$, then one can determine $b_1, \dots, b_n, \dots$ such that $g = 1 + \sum_{i=1}^{\infty} b_i t^i$ is the

multiplicative inverse of f in $R[[t]]$. In particular,

$$R[[t]]^{\times} = \{a_0 + \sum_{i=1}^{\infty} a_i t^i \in R[[t]] \mid a_0 \in R^{\times}\}.$$

3. Let R be a commutative ring and $A = R[[t]]$, the power series over R. Let $\mathfrak{P}$ be a prime ideal in A not containing t and $\varphi : A \to R$ the homomorphism induced by $t \mapsto 0$. Suppose that $\varphi(\mathfrak{P}) = (a_1, \dots, a_n)$. Let $f_i = a_i+$ higher terms in t for $i = 1, \dots, n$. Prove if $g \in \mathfrak{P}$, then $g = f_1 h_1 + \cdots + f_n h_n$ for some elements $h_1, \dots, h_n$ in A.
4. Let A be an additive group and let $\mathrm{End}(A)$ denote the set of group homomorphisms of A to A. Prove that $\mathrm{End}(A)$ is a ring under addition and composition of elements in $\mathrm{End}(A)$. Also show the unit group of $\mathrm{End}(A)$ is the group of group automorphisms of A.
5. Let F be a field. Show that $\mathbb{M}_n(F)$ is a simple ring. If $n > 1$ then $\mathbb{M}_n(F)$ is not a division ring.
6. Show that a ring homomorphism $\varphi : R \to S$ is a monomorphism if and only if given any ring homomorphisms $\psi_1, \psi_2 : T \to R$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$.
7. Show if R is a commutative ring, then, for any ideal $\mathfrak{B}$ in $\mathbb{M}_n(R)$, there exists an ideal $\mathfrak{A}$ in R satisfying $\mathfrak{B} = \mathbb{M}_n(\mathfrak{A})$ (obvious definition). In particular, if R is simple, so is $\mathbb{M}_n(R)$. Show that $\mathbb{M}_n(R)$ is never a division ring if $n > 1$.
8. Let R be a commutative ring. Prove the *Binomial Theorem*: Let a and b be elements of R and n a positive integer. Then

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$

9. If R is a ring satisfying $x^2 = x$ for all x in R, then R is commutative.
10. If R is a rng satisfying $x^3 = x$ for all x in R, then R is commutative.
11. Let R be a commutative ring and $\mathfrak{A}$ be an ideal in R satisfying

$$\mathfrak{A} = \mathfrak{m}_1 \cdots \mathfrak{m}_r = \mathfrak{n}_1 \cdots \mathfrak{n}_s$$

with all the $\mathfrak{m}_i$ distinct maximal ideals and all the $\mathfrak{n}_j$ distinct maximal ideals. Show that $r = s$ and there exists a $\sigma \in S_r$ satisfying $\mathfrak{m}_i = \mathfrak{n}_{\sigma(i)}$ for all i.
12. Let R be a commutative ring and $\mathfrak{A}$ an ideal of R. Suppose that every element in $R \setminus \mathfrak{A}$ is a unit of R. Show that $\mathfrak{A}$ is a maximal ideal of R and that, moreover, it is the only maximal ideal of R.
13. Let R be the set of all continuous functions $f : [0,1] \to \mathbb{R}$. Then R is a commutative ring under $+$ and $\cdot$ of functions. Show that any maximal ideal of R has the form $\{f \in R \mid f(a) = 0\}$ for some fixed a in $[0,1]$.
14. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. Show that if $\mathfrak{B}$ is an ideal (respectively, prime ideal) of S then $\varphi^{-1}(\mathfrak{B})$ is an ideal (respectively, prime ideal) of R. Give an example with $\mathfrak{B}$ a maximal ideal of S but $\varphi^{-1}(\mathfrak{B})$ not a maximal ideal of R.

15. Show $F[t]$, with F a field, is a PID.
16. Show that $\mathbb{Z}[t]$, the ring of polynomials with coefficients in $\mathbb{Z}$, is not a PID. Show this by finding a non-principal maximal ideal in $\mathbb{Z}[t]$. Also find a nonzero prime ideal in $\mathbb{Z}[t]$ that is not maximal and prove it is such.
17. Let $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ be ideals in R, at least $n-2$ of which are prime. Let $S \subset R$ be a subrng (it does not have to have a 1) contained in $\mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$. Then one of the $\mathfrak{A}_j$'s contains S. In particular, if $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ are prime ideals in R and $\mathfrak{B}$ is an ideal properly contained in S satisfying $S \setminus \mathfrak{B} \subset \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, then S lies in one of the $\mathfrak{p}_i$'s.

27. Factor Rings and Rings of Quotients

We begin by generalizing the construction of the integers modulo n.

Definition 27.1. Let R be a ring, $\mathfrak{A}$ an ideal in R. If a and b are elements in R, write $a \equiv b \mod \mathfrak{A}$ if $a - b$ is an element of $\mathfrak{A}$.

A proof analogous proof to that for the integers modulo n, establishes the following:

Proposition 27.2. *Let R be a ring and $\mathfrak{A}$ an ideal in R. Then $\equiv \mod \mathfrak{A}$ is an equivalence relation. Suppose that the elements a, a', b, b' in R satisfy:*

$$a \equiv a' \mod \mathfrak{A} \qquad \textit{and} \qquad b \equiv b' \mod \mathfrak{A},$$

then

$$\begin{aligned} a + b &\equiv a' + b' \mod \mathfrak{A}, \\ a \cdot b &\equiv a' \cdot b' \mod \mathfrak{A}. \end{aligned}$$

Let $\overline{R} = R/\mathfrak{A} := R/(\equiv \mod \mathfrak{A})$. If $a \in R$, set

$$\overline{a} = \{b \in R \mid b \equiv a \mod \mathfrak{A}\} = a + \mathfrak{A}.$$

If

$$^{-} : R \to R/\mathfrak{A} \text{ given by } a \mapsto \overline{a}$$

denotes the canonical surjection and a, b lie in R, define

(i) $$\overline{a} + \overline{b} := \overline{a + b}$$

(ii) $$\overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

Then under these operations, $\overline{R}$ is a ring with $0_{\overline{R}} = \mathfrak{A}$ and $1_{\overline{R}} = 1_R + \mathfrak{A}$ with $\overline{R}$ commutative if R is. The canonical surjection is an epimorphism with kernel $\mathfrak{A}$. We call $\overline{R}$ the factor or quotient ring of R by $\mathfrak{A}$.

Proofs analogous to those in Group Theory show:

Theorem 27.3. (First Isomorphism Theorem) *Let $\varphi : R \to S$ be a ring homomorphism. Then we have a commutative diagram of rings and ring homomorphisms*

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[d, "-"'] & S \\ R/\ker\varphi \arrow[r, "\overline{\varphi}"'] & \operatorname{im}\varphi \arrow[u, "\mathrm{inc}"'] \end{tikzcd}$$

with $^-$ *a ring epimorphism,* $\overline{\varphi}$ *a ring isomorphism between* $R/\ker\varphi$ *and* $\operatorname{im}\varphi$*, and* inc *a ring monomorphism.*

and

Theorem 27.4. (Correspondence Principle) *Let* $\varphi : R \to S$ *be a ring epimorphism. Then*

$$\{\mathfrak{B} \mid \mathfrak{B} \text{ an ideal in } R \text{ with } \ker\varphi \subset \mathfrak{B}\} \longrightarrow \{\mathfrak{C} \mid \mathfrak{C} \text{ an ideal in } S\}$$

given by $\mathfrak{B} \mapsto \varphi(\mathfrak{B})$ *is an order preserving bijection.*

We give some applications.

Proposition 27.5. *Let* R *be a nontrivial commutative ring and* $\mathfrak{A}$ *an ideal of* R*. Then*

(1) $\mathfrak{A}$ *is a prime ideal of* R *if and only if* $R/\mathfrak{A}$ *is a domain.*
(2) $\mathfrak{A}$ *is a maximal ideal of* R *if and only if* $R/\mathfrak{A}$ *is a field.*

PROOF. Let $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism, so $\ker{}^- = \mathfrak{A}$.
(1): $\overline{R}$ is a domain if and only if $(\overline{0})$ is a prime ideal in $\overline{R}$ if and only if whenever $a, b \in R$ satisfy $\overline{a} \cdot \overline{b} = \overline{0}$, then $\overline{a} \in (\overline{0})$ or $\overline{b} \in (\overline{0})$ if and only if whenever $a, b \in R$ satisfy $a \cdot b \in \mathfrak{A}$, then $a \in \mathfrak{A}$ or $b \in \mathfrak{A}$ if and only if $\mathfrak{A}$ is a prime ideal.
(2): $(\overline{0}) \neq \overline{R}$ is a field if and only if $\overline{R}$ is simple if and only if $(\overline{0})$ and $\overline{R}$ are the only ideals of $\overline{R}$ if and only if $\{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} \subset R \text{ is an ideal}\} = \{\mathfrak{A}, R\}$ if and only if $\mathfrak{A} < R$ is maximal by the Correspondence Principle. □

Question Is (2) true if the ring R is not necessarily commutative and we replace field by division ring?

Compare the next application with the Third Isomorphism Theorem of Groups.

Proposition 27.6. *Let* R *be a ring and* $\mathfrak{A} \subset \mathfrak{B} \subset R$ *ideals. Then*

$$\mathfrak{B}/\mathfrak{A} := \{b + \mathfrak{A} \mid b \in \mathfrak{B}\} \subset R/\mathfrak{A}$$

is an ideal in $R/\mathfrak{A}$ *and*

$$R/\mathfrak{B} \cong (R/\mathfrak{A})/(\mathfrak{B}/\mathfrak{A}).$$

PROOF. It is easy to check that $R/\mathfrak{A} \to R/\mathfrak{B}$ given by $r + \mathfrak{A} \mapsto r + \mathfrak{B}$ is a well-defined epimorphism with kernel $\mathfrak{B}/\mathfrak{A}$. □

Proposition 27.7. *Let* $\varphi : R \to S$ *be a ring epimorphism of commutative rings with kernel* $\mathfrak{A}$*. Then the map*

$$\{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal with } \mathfrak{A} \subset \mathfrak{p}\} \to \{\mathfrak{P} \mid \mathfrak{P} < S \text{ a prime ideal}\}$$

given by $\mathfrak{p} \mapsto \varphi(\mathfrak{p})$ *is a bijection.*

PROOF. We may assume that S is not the trivial ring for if not then both sets of prime ideals would be empty. By the First Isomorphism Theorem, we may assume that $S = R/\mathfrak{A}$ and φ is the canonical epimorphism $^{-} : R \to R/\mathfrak{A}$. Let $\mathfrak{A} \subset \mathfrak{B} \subset R$ be an ideal. Then $\mathfrak{B}$ is a prime ideal in R if and only if $R/\mathfrak{B}$ is a domain if and only if $(R/\mathfrak{A})/(\mathfrak{B}/\mathfrak{A})$ is a domain if and only if $\mathfrak{B}/\mathfrak{A}$ is a prime ideal in $R/\mathfrak{A}$. $\square$

Warning 27.8. If $\varphi : R \to S$ is a ring homomorphism of rings with $\mathfrak{p}$ a prime ideal in R, it does not follow that the ideal generated by image of $\mathfrak{p}$, $S\varphi(\mathfrak{p})$, is a prime ideal in S. For example, the inclusion $\mathbb{Z} \subset \mathbb{Q}$ is a ring homomorphism and (2) is a prime ideal in $\mathbb{Z}$, but its image (2) is the unit ideal in $\mathbb{Q}$.

Next we want to determine the smallest subring of a ring.

Construction 27.9. Let R be a nontrivial ring. As any ring homomorphism must take the multiplicative identity to the multiplicative identity, there exists a unique ring homomorphism from $\mathbb{Z}$ to R. It is given by

$$\iota : \mathbb{Z} \to R \text{ given by } m \mapsto m1_R = \begin{cases} \underbrace{1_R + \cdots + 1_R}_{m} & \text{if } m \geq 0 \\ \underbrace{-1_R + \cdots + -1_R}_{-m} & \text{if } m < 0. \end{cases}$$

Since $\mathbb{Z}$ is a PID, there exists a unique integer $n \geq 0$ such that $\ker \iota = (n)\,[= (-n)]$. We call n the *characteristic* of the ring R and write $\operatorname{char}(R)$ (or just $\operatorname{char} R$) for this integer. We shall only be interested in the characteristic of commutative rings.

Remarks 27.10. Let R be a ring and $\iota : \mathbb{Z} \to R$ the unique ring homomorphism.

1. $\operatorname{im} \iota$ is the unique smallest subring of R.
2. $\operatorname{char}(R) = 0$ if and only if $m1_R \neq 0$ for all nonzero integers m if and only if ι is monic.
3. If $\operatorname{char}(R) \neq 0$ then, by the Division Algorithm, $\operatorname{char}(R)$ is the least positive integer n such that $n1_R = 0$.
4. If $\varphi : \mathbb{Z}/n\mathbb{Z} \to R$ and $\psi : \mathbb{Z}/m\mathbb{Z} \to R$ are ring monomorphisms with n, m positive integers, then $n = \operatorname{char}(R) = m$ and $\varphi = \psi$ are induced by ι.

Examples 27.11. 1. $\mathbb{Z}$ is a domain, and $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields of characteristic zero.

2. If $n \in \mathbb{Z}^+$ then $\operatorname{char}(\mathbb{Z}/n\mathbb{Z}) = n$.
3. If $\varphi : R \to S$ is a ring monomorphism, then $\operatorname{char}(R) = \operatorname{char}(S)$.
4. The canonical epimorphism $^{-} : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ shows that, in general, the characteristic of a ring is not preserved under a ring homomorphism.

Proposition 27.12. *Let R be a domain. Then either*

(1) $\operatorname{char}(R) = 0$ *and there exists a ring monomorphism* $\mathbb{Z} \to R$ *or*
(2) $\operatorname{char}(R) = p$, *$p$ a prime, and there exists a ring monomorphism* $\mathbb{Z}/p\mathbb{Z} \to R$.

PROOF. Let $\iota : \mathbb{Z} \to R$ be the unique ring homomorphism. Then $\ker \iota = (n)$ for some non-negative integer n, so $\operatorname{char}(R) = n$. As R is a domain so is $\operatorname{im} \iota \cong \mathbb{Z}/n\mathbb{Z}$. Therefore, $n = 0$ or n is a prime and the induced map $\bar{\iota} : \mathbb{Z}/n\mathbb{Z} \to R$ is a monomorphism. $\square$

The proposition says that if R is a domain, we may view $\mathbb{Z} \subset R$ if $\operatorname{char}(R) = 0$ and $\mathbb{Z}/p\mathbb{Z} \subset R$ if $\operatorname{char}(R) = p$. Of course, the same discussion shows that if R is a ring, then the unique smallest subring of R is isomorphic to $\mathbb{Z}/(\operatorname{char} R)$

We have now generalized the construction of $\mathbb{Z}/n\mathbb{Z}$ to $R/\mathfrak{A}$, where $\mathfrak{A}$ is an ideal in R. Our next goal is to construct $\mathbb{Q}$ from $\mathbb{Z}$. Our construction will work for any domain.

Construction 27.13. Let R be a domain. We wish to construct fractions from R, i.e., if $a, b, c, d \in R$ with $b \neq 0$ and $d \neq 0$, we want to have the following:

$$(i) \qquad \frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}. \quad \text{[Note } bd \neq 0 \text{ in the domain } R.]$$

$$(ii) \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},$$

and we know that we must have

$$(iii) \qquad \frac{a}{b} = \frac{c}{d} \text{ if and only if } ad = bc \text{ in } R.$$

We do this construction as follows: Let

$$\mathcal{S} = \{(a, b) \mid a \in R,\, 0 \neq b \in R\} = R \times (R \setminus \{0\})$$

and

$$(a, b) \sim (c, d) \text{ in } \mathcal{S} \text{ if } ad = bc.$$

Check that $\sim$ is an equivalence relation.

Let

$$\frac{a}{b} := [(a, b)]_\sim, \text{ the equivalence class of } (a, b) \text{ under } \sim \text{ and}$$

$$K := \mathcal{S}/{\sim} = \{\frac{a}{b} \mid (a, b) \in \mathcal{S}\} = \{\frac{a}{b} \mid a, b \in R,\, b \neq 0\}.$$

Then K satisfies (iii) above. Define $+ : K \times K \to K$ and $\cdot : K \times K \to K$ by (i) and (ii) above respectively.

Claim. $+$ and $\cdot$ are well-defined:

Suppose that $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ in $\mathcal{S}$. We must show

$$(ad + bc)b'd' = (a'd' + b'c')bd$$
$$(ac)(b'd') = (a'c')(bd)$$

knowing that

$$ab' = a'b \text{ and } cd' = c'd,$$

which is easily checked. Thus K has a $+$ and $\cdot$. It is routine to check that K is a commutative ring with $0_K = \frac{0}{b}$, and $1_K = \frac{b}{b}$ with $0 \neq b \in R$. If $\frac{a}{b} \neq 0_K$ then $a \neq 0$ in R, hence $\frac{b}{a}$ is defined and is the inverse of $\frac{a}{b}$, so K is a field called the *field of quotients* or the *quotient field* of R. We write $qf(R)$ for K.

Example. $\mathbb{Q} = qf(\mathbb{Z})$.

This construction of $qf(R)$ is essentially unique as $qf(R)$ has the following *universal property*:

Theorem 27.14. (The Universal Property of the Field of Quotients) *Let R be a domain. Then there exists a field K, and a ring monomorphism $i : R \to K$ satisfying the following Universal Property: If F is a field, then for any ring monomorphism $\varphi : R \to F$, there exists a unique ring monomorphism $\psi : K \to F$ such that the diagram*

(†)
$$\begin{tikzcd} R \arrow[r, "i"] \arrow[dr, "\varphi"'] & K \arrow[d, "\psi"] \\ & F \end{tikzcd}$$

commutes.

In particular, if K' is a field and $j : R \to K'$ a ring monomorphism also satisfying the universal property (†), then there exists a unique (field) isomorphism $\sigma : K \to K'$ such that the diagram

$$\begin{tikzcd} R \arrow[r, "i"] \arrow[dr, "j"'] & K \arrow[d, "\sigma"] \\ & K' \end{tikzcd}$$

commutes.

Proof. Let $K = qf(R)$, then $i : R \to K$ given by $r \mapsto \frac{r}{1}$ is a ring homomorphism. Let b be a nonzero element in R. Then $\frac{r}{1} = \frac{0}{b}$ in K if and only if $rb = 1 \cdot 0 = 0$ in the domain R if and only if $r = 0$, so this homomorphism is injective.

Now suppose that we are given a ring monomorphism $\varphi : R \to F$ with F a field. Define

$$\psi : K \to F \text{ by } \psi(\frac{a}{b}) = \varphi(a)\varphi(b)^{-1}.$$

This makes sense, as $\varphi(b) \neq 0$, since φ is monic and F is a field.

Claim 1. ψ is well-defined and monic.

$\dfrac{a}{b} = \dfrac{a'}{b'}$ in K if and only if $ab' = a'b$ in R if and only if $\varphi(a)\varphi(b') = \varphi(a')\varphi(b)$ in F as φ is monic if and only if $\varphi(a)\varphi(b)^{-1} = \varphi(a')\varphi(b')^{-1}$ in the field F if and only if $\psi(\frac{a}{b}) = \psi(\frac{a'}{b'})$ in F. It is easily seen that ψ is a homomorphism. The Claim follows.

Claim 2. ψ is unique.

If ψ' is another monomorphism making the diagram (†) commute, then

$$\psi'(\frac{a}{1}) = \varphi(a) = \psi(\frac{a}{1_K}) \text{ for all } a \in R$$

in the field F. In particular, if $b \neq 0$ in R, we have

$$\psi'(\frac{b}{1})\psi'(\frac{1}{b}) = \psi'(1) = 1_F = \psi(\frac{b}{1})\psi(\frac{1}{b})$$

in F, so

$$\psi'(\frac{1}{b}) = \psi(\frac{1}{b}),$$

hence, for all a and b in R with $b \neq 0$,

$$\psi'(\frac{a}{b}) = \psi'(\frac{a}{1})\psi'(\frac{1}{b}) = \psi(\frac{a}{1})\psi(\frac{1}{b}) = \psi(\frac{a}{b}).$$

This proves the second claim. Finally, we show:

Claim 3. The map $i : R \to K$ satisfies the uniqueness property relative to the diagram (†).

Suppose that $j : R \to K'$ also satisfies (†). Then there exist unique ring monomorphisms $\psi : K \to K'$ and $\psi' : K' \to K$ such that we have commutative diagrams

$$\begin{tikzcd} R \arrow[r, "i"] \arrow[dr, "j"'] & K \arrow[d, "\psi"] \\ & K' \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} R \arrow[r, "j"] \arrow[dr, "i"'] & K' \arrow[d, "\psi'"] \\ & K. \end{tikzcd}$$

By (†), we have $\psi'\psi = 1_K$ and $\psi\psi' = 1_{K'}$ as

$$\begin{tikzcd} R \arrow[r, "i"] \arrow[dr, "i"'] & K \arrow[d, "1_K"] \\ & K \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} R \arrow[r, "j"] \arrow[dr, "j"'] & K' \arrow[d, "1_{K'}"] \\ & K' \end{tikzcd} \qquad \text{commute.}$$

It follows that ψ and ψ' are inverse isomorphisms. $\square$

Notation 27.15. If R is a domain, we shall always view $R \subset qf(R)$, i.e., we shall identify R with its image $\{\frac{r}{1} \mid r \in R\}$ in $qf(R)$, e.g., $\mathbb{Z} \subset \mathbb{Q}$.

Remarks 27.16. In general, it is very difficult to define a map between objects. We may have an idea where generators of the domain object must go, but to show the relations of the domain are preserved usually presents problems, i.e., to show the putative map is well-defined. Universal properties are methods to address this in certain cases. One tries to construct an object with given properties together with a map, so that there is automatically a map (that preserves the desired structure) between it and any other other object that satisfies the given properties. Moreover, that map is unique. Therefore, if you have such a universal property, you automatically obtain unique maps to other appropriate objects. The above is an example of such. Another, that you have seen in linear algebra, is given a vector space together with a given basis, any linear transformation from it to another vector space is completely determined where the basis maps.

The following is a useful observation:

Observation 27.17. Let F be a field and R a nontrivial ring. If $\varphi : F \to R$ is a ring homomorphism, then φ is monic as $\ker \varphi < F$ and F is simple.

If F is a field, then we can say a bit more.

Corollary 27.18. *Let F be a field of characteristic n.*

(1) If $n = 0$, then there exists a (unique) monomorphism $\mathbb{Q} \to F$.
(2) If $n > 0$, then $n = p$ is a prime and there exists a (unique) monomorphism $\mathbb{Z}/p\mathbb{Z} \to F$.

If K is a field, there exists a unique smallest subfield Δ_K of K, viz., the intersection of all the subfields of K. This field is called the *prime subfield* of K. By the corollary, we know that

$$\Delta_K \cong \begin{cases} \mathbb{Q} & \text{if } \operatorname{char}(K) = 0. \\ \mathbb{Z}/p\mathbb{Z} & \text{if } \operatorname{char}(K) = p. \end{cases}$$

For example, let K be a finite field with N elements. Then there exists a unique prime p such that $\operatorname{char} K = p$ and a unique monomorphism $\mathbb{Z}/p\mathbb{Z} \to K$. We also have K is a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$, say of dimension n. So $N = p^n$. We know that $K^\times$ is a group of order $N-1$, so $x^{N-1} = 1$ for all x in $K^\times$ and $x^N = x$ for all $x \in K$. This shows that the analogue of Fermat's Little Theorem holds for all finite fields. We shall show later that $K^\times$ is, in fact, a cyclic group as well as showing that for every prime p and positive integer m there exists a field of cardinality p^m, and that this field is unique up to an isomorphism of fields. We also point out that there are many infinite fields of characteristic $p > 0$. Indeed, the ring $(\mathbb{Z}/p\mathbb{Z})[t]$ is an infinite domain (why?), hence so is its quotient field.

Next we turn to a generalization of the Chinese Remainder Theorem.

If R_i, $i \in I$, are rings, then the cartesian product of the R_i, $R = \bigtimes_I R_i$ becomes a ring via componentwise operations. We also write this as $\prod R_i$. It is commutative if all the R_i are commutative. It is easy to check that $R^\times = \prod_I (R_i)^\times$. [As usual, we are sloppy about writing elements in $\prod_I R_i$.]

Theorem 27.19. (Chinese Remainder Theorem) *Suppose that R is a commutative ring with $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ ideals in R that are comaximal, i.e., $\mathfrak{A}_i + \mathfrak{A}_j = R$ for all $i, j = 1, \dots, n$ with $i \neq j$. Let*

$$\varphi : R \to \prod_{i=1}^{n} R/\mathfrak{A}_i \ \textit{ be given by } \ r \mapsto (r + \mathfrak{A}_1, \dots, r + \mathfrak{A}_n).$$

Then φ is a ring epimorphism with

$$\ker \varphi = \mathfrak{A}_1 \cap \dots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cdots \mathfrak{A}_n,$$

hence induces isomorphisms

$$\begin{aligned} \overline{\varphi} :& R/(\mathfrak{A}_1 \cdots \mathfrak{A}_n) \to R/\mathfrak{A}_1 \times \cdots \times R/\mathfrak{A}_n \\ \overline{\varphi} :& (R/(\mathfrak{A}_1 \cdots \mathfrak{A}_n)^\times \to (R/\mathfrak{A}_1)^\times \times \cdots \times (R/\mathfrak{A}_n)^\times. \end{aligned}$$

Proof. Clearly we have φ is a ring homomorphism with $\ker \varphi = \mathfrak{A}_1 \cap \dots \cap \mathfrak{A}_n$. By this and the First Isomorphism Theorem, we need only show that $\mathfrak{A}_1 \cap \dots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cdots \mathfrak{A}_n$ and φ is onto, i.e., if $x_1, \dots, x_n$ lie in R, there exists an $x \in R$ satisfying $x \equiv x_i \mod \mathfrak{A}_i$ for all i. We prove this by induction on n.

$n = 2$: We have $R = \mathfrak{A}_1 + \mathfrak{A}_2$, so $1 = a_1 + a_2$ for some $a_i \in \mathfrak{A}_i$, $i = 1, 2$. Therefore, $a_1 \equiv 1 \mod \mathfrak{A}_2$ and $a_2 \equiv 1 \mod \mathfrak{A}_2$; hence $x = x_1 a_2 + x_2 a_1 \equiv x_i \mod \mathfrak{A}_i$ for $i = 1, 2$. As $\mathfrak{A}_1\mathfrak{A}_2 \subset \mathfrak{A}_1 \cap \mathfrak{A}_2$, we need only show that $\mathfrak{A}_1 \cap \mathfrak{A}_2 \subset \mathfrak{A}_1\mathfrak{A}_2$ to complete the $n = 2$ case. If $b \in \mathfrak{A}_1 \cap \mathfrak{A}_2$, then $b = b \cdot 1 = ba_1 + ba_2 \in \mathfrak{A}_2\mathfrak{A}_1 + \mathfrak{A}_1\mathfrak{A}_2 \subset \mathfrak{A}_1\mathfrak{A}_2$.

[**Warning**: If $\mathfrak{B}_1$, $\mathfrak{B}_2$ are ideals in a ring, usually $\mathfrak{B}_1\mathfrak{B}_2 < \mathfrak{B}_1 \cap \mathfrak{B}_2$.]

$n > 2$: By hypothesis, we have $R = \mathfrak{A}_1 + \mathfrak{A}_i$ for $i > 1$, so $1 = a_i + b_i$ for some $a_i \in \mathfrak{A}_1$ and $b_i \in \mathfrak{A}_i$ for each $i > 1$. Consequently, we have $1 = \prod_{i=2}^{n}(a_i + b_i)$ lies in $\mathfrak{A}_1 + (\mathfrak{A}_2 \cdots \mathfrak{A}_n)$ (why?), so $R = \mathfrak{A}_1 + (\mathfrak{A}_2 \cdots \mathfrak{A}_n)$. By the $n = 2$ case and induction, we conclude that $\mathfrak{A}_1 \cap \cdots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cap (\mathfrak{A}_2 \cdots \mathfrak{A}_n) = \mathfrak{A}_1 \cdots \mathfrak{A}_n$.

By induction, there exists a y in R satisfying

$$y \equiv x_i \mod \mathfrak{A}_i \quad \text{for } i = 2, \dots, n;$$

and by the $n = 2$ case, there exists an x in R satisfying

$$\begin{aligned} x &\equiv x_1 \mod \mathfrak{A}_1 \\ x &\equiv y \mod \mathfrak{A}_2 \cdots \mathfrak{A}_n. \end{aligned}$$

As $\mathfrak{A}_2 \cdots \mathfrak{A}_n \subset \mathfrak{A}_i$ for $i = 2, \dots n$, we have $x \equiv x_i \mod \mathfrak{A}_i$ for $i = 1, \dots, n$, proving that φ is surjective.

We leave the isomorphism of multiplicative groups as an exercise. □

Exercises 27.20.

1. Let $\varphi : R \to S$ be a ring epimorphism (of rings). Do an analogous analysis before The First Isomorphism Theorem for Groups 12.1 for $\varphi : R \to S$ to see how ideals and factor rings arise naturally.
2. Let $R = (\mathbb{Z}/2\mathbb{Z})[t]$, $f = f(t) = t^2 + t + 1$, and $g = t^2 + 1$. Show all of the following:
   (*i*) $R/(f)$ is a field with four elements.
   (*ii*) $R/(g)$ is not a domain and has four elements.
   (*iii*) Neither $R/(f)$ nor $R/(g)$ is isomorphic to the ring $\mathbb{Z}/4\mathbb{Z}$.
3. Show that the ideals (2) and (t) are prime ideals in $\mathbb{Z}[t]$
4. Let R be a commutative ring. Suppose for every element x in R there exists an integer $n = n(x) > 1$ such that $x^n = x$. Show that every prime ideal in R is maximal.
5. Let R be a commutative ring. If $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in R and $\overline{} : R \to R/\mathfrak{A}$ is the canonical epimorphism, show that this induces an isomorphism $R/(\mathfrak{A} + \mathfrak{B}) \to \overline{R}/\overline{\mathfrak{B}}$.
6. (Second Isomorphism Theorem) Let $R \subset S$ be a subring, $\mathfrak{A} \subset S$ an ideal. Then $\mathfrak{A} \cap R$ is an ideal in R, $R + \mathfrak{A}$ is a subring of S, and $(R + \mathfrak{A})/\mathfrak{A} \cong R/(R \cap \mathfrak{A})$.
7. Let R be a rng and $\mathfrak{A}$ an ideal in R. Then there exists an isomorphism of rngs, $(R + \mathfrak{A})/\mathfrak{A} \cong \mathfrak{A}/(R \cap \mathfrak{A})$.
8. Show that the inclusion map $i : \mathbb{Z} \subset \mathbb{Q}$ satisfies the property that if $\psi_1 \circ i = \psi_2 \circ i$, for $\psi_1, \psi_2 : \mathbb{Q} \to R$ ring homomorphisms, then $\psi_1 = \psi_2$. This shows that a surjective ring homomorphism is not equivalent to this property. (Cf. Remark 9.5.)
9. Let R be a commutative ring and $A = R[[t]]$, the power series over R. Show if $\mathfrak{A}$ is an ideal in R, then $\mathfrak{A}$ is a prime ideal in R if and only $A\mathfrak{A}$ is a prime ideal in A. Moreover, if $\mathfrak{m}$ is a maximal ideal in R, then $A\mathfrak{m} + At$ is a maximal ideal in A and the unique maximal ideal in A containing $\mathfrak{m}$. In particular, if R has precisely one maximal ideal, then so does A. [A commutative ring having precisely one maximal ideal is called a *local ring*.]
10. Let R be a commutative ring of characteristic $p > 0$, p a prime. Prove that the map $R \to R$ by $x \mapsto x^p$ is a ring homomorphism. It is called the *Frobenius homomorphism*.

In particular, the *Children's Binomial Theorem holds*, i.e., $(x+y)^p = x^p + y^p$ in R for all x and y in R.

11. Let R be a finite domain. Show that $\operatorname{char}(R) = p$ is a prime and R is a field. Moreover, R is a vector space over $\mathbb{Z}/p\mathbb{Z}$.
12. Show that if R is a domain, so is the polynomial ring $R[t]$. In particular, show that there exists fields properly containing the complex numbers. Does the field that you constructed have the property that every non-constant polynomial over it has a root? Prove or disprove this.
13. Let $\varphi : R \to S$ be a ring isomorphism. Show φ induces a group isomorphism $R^\times \to S^\times$ by restriction. We also write this map as φ.
14. Prove the isomorphism statement about the multiplicative groups in the Chinese Remainder Theorem 27.19

28. Zorn's Lemma

One would like to have a method of proof generalizing induction. There are many equivalent ways of doing this: Well-Ordering Principle, Transfinite Induction, Zorn's Lemma. These are also equivalent to the Axiom of Choice, Tychonoff's Theorem, the theorem that all vector spaces have bases. In algebra, the most useful form is Zorn's Lemma, which generalizes the Well-Ordering Principle. These are independent of the other axioms in Zermelo-Frankel Set Theory, and we accept them as an axiom. [In Appendix A, we show that the Axiom of Choice, Zorn's Lemma, and the Well-Ordering Principle are all equivalent.]

Definition 28.1. Let S be a set with a relation R on S. We call S a *partially ordered set* or *poset* (under R) if the following hold: For all a, b, c in S

(1) $a\mathrm{R}a$. [Reflexitity]

(2) If $a\mathrm{R}b$ and $b\mathrm{R}a$ then $a = b$.

(3) If $a\mathrm{R}b$ and $b\mathrm{R}c$ then $a\mathrm{R}c$. [Transitivity]

We also say (S, R) is a poset, if the relation is not obvious. So the second condition replaces symmetry in the definition of an equivalence relation. If S is a poset (under R), a, b elements in S, then a and b may be *incomparable*, i.e., neither $a\mathrm{R}b$ nor $b\mathrm{R}a$. We say that the poset S is a *chain* or *totally ordered* if for all a and b in S, we have either $a\mathrm{R}b$ or $b\mathrm{R}a$, i.e., all elements are *comparable*.

Examples 28.2. 1. Recall if T is a set, then the *power set* of T is defined by $\mathcal{P}(T) := \{A \mid A \subset T\}$. It becomes a poset under $\subset$ (or $\supset$). It is almost never a chain. [When is it?]

2. $(\mathbb{Z}, \leq)$ and $(\mathbb{R}, \geq)$ are chains.
3. As $\{1, \sqrt{-1}\}$ is a basis for $\mathbb{C}$ as a vector space over $\mathbb{R}$, for each $\alpha \in \mathbb{C}$, there exist unique $x, y \in \mathbb{R}$ such that $\alpha = x + y\sqrt{-1}$. Define the *lexicographic order* on $\mathbb{C}$ by $\alpha = x + y\sqrt{-1} \leq_L \beta = w + z\sqrt{-1}$ if $\alpha = \beta$ and if not this, then $x < w$ and if not these, then ($x = w$ and $y < z$) in $\mathbb{R}$. Then $(\mathbb{C}, \leq_L)$ is a chain.

Note: If $\alpha \leq_L \beta$ and $\gamma \leq_L \delta$, then $\alpha + \gamma \leq_L \beta + \delta$, but $\alpha\gamma >_L \beta\delta$ is possible (where $>_L$ is the obvious relation), e.g., $0 \leq_L \sqrt{-1}$ but $\sqrt{-1}\sqrt{-1} <_L 0$.
[One can generalize the lexicographic order on $\mathbb{C}$ to any finite dimensional vector space V over $\mathbb{R}$ with ordered basis $\{v_1, \ldots, v_n\}$. How?]

Definition 28.3. Let (S, R) be a poset, T a subset of S. An element a in S is called an *upper bound* of T if $x\mathrm{R}a$ for all $x \in T$, and s is called a *maximal element* of S if $s\mathrm{R}y$ with $y \in S$ implies that $s = y$. We say that the poset S is *inductive* if every chain in S has an upper bound in S.

Examples 28.4. 1. 1 is a maximal element in $([0, 1], \leq)$.
2. 0 is a maximal element in $([0, 1], \geq)$. (Of course, it is usually called a *minimal element*.)
3. $(\mathbb{Z}, \leq)$ is not inductive.
4. $(\mathbb{Z}^+, \geq)$ is inductive. (Note here 'upper bound' is usually called 'lower bound' for obvious reasons.)

We can now state the axiom that we shall use as an extension of finite induction.

Lemma 28.5. (Zorn's Lemma) *Let S be a nonempty inductive poset. Then S contains a maximal element.*

This is indeed an extension of finite induction, for if S is a subset of $\mathbb{Z}^+$, then $(S, \geq)$ has a maximal element, which is exactly the Well-Ordering Principle. As mentioned, we cannot prove this lemma; we accept it as an axiom.

We give some applications of Zorn's Lemma. The first is the result that vector spaces always have bases.

Proposition 28.6. *Let V be a nonzero vector space over a field F and S a linearly independent subset of V. Then S extends to a basis of V, i.e., is part of a basis for V.*

Proof. Let
$$\mathcal{S} = \{T \mid S \subset T \subset V \text{ with } T \text{ linearly independent}\},$$
a nonempty poset under $\subset$. Let $\mathcal{C}$ be a chain in $\mathcal{S}$.
Claim. $A = \bigcup_{\mathcal{C}} T$ is an upper bound for $\mathcal{C}$ in $\mathcal{S}$:
Suppose that A is linearly dependent. By the definition of linear dependence, there exists a finite linearly dependent subset $\{v_1, \ldots, v_n\}$ of A. For each i, $1 \leq i \leq n$, there exists a $T_i \in \mathcal{C}$ with $v_i \in T_i$. Since $\mathcal{C}$ is a chain and n finite, there exists a T in $\mathcal{C}$ containing T_i for $i = 1, \ldots, n$, hence containing $\{v_1, \ldots, v_n\}$, contradicting the fact that T is linearly independent. This shows that A is indeed an upper bound of $\mathcal{C}$ in $\mathcal{S}$.

As the hypotheses of Zorn's Lemma have been fulfilled, there exists a maximal element T in $\mathcal{S}$. We show that T is a basis for V. If the span, $\langle T \rangle$, of T is not V, then there exists $w \in V \setminus \langle T \rangle$. Let $T' = T \cup \{w\}$. As T' must be linearly dependent by maximality of T, there exist β and β_v in F for $v \in T$, with *almost all* $\beta_v = 0$, i.e., $\beta_v = 0$ except for finitely many v in T, but not all β, β_v zero that satisfies $\sum_T \beta_v v + \beta w = 0$. As T is linearly independent, $\beta \neq 0$, so $w = -\sum_T \beta^{-1}\beta_v v$ lies in $\langle T \rangle$, a contradiction. Consequently, T spans V, so is a basis for V. $\square$

Remark 28.7. From linear algebra, we know that any two bases for a finite dimensional vector space have the same cardinality. As mentioned before, this is also true if the vector space is not finite dimensional using the same proof that all linear transformations are completely determined by where a basis is mapped (which does not need finiteness) and the Schroeder-Bernstein Theorem A.13 which says if X and Y are sets and there is an injective map $X \to Y$, then $|X| \leq |Y|$ defines a linear ordering with equality if and only if there is a bijection $X \to Y$.

Proposition 28.8. *Let V be a nonzero vector space over a field F and S a spanning set for V. Then a subset of S is a basis of V.*

We leave a proof of this as an exercise. You should try the obvious proof. It does not work. Can you figure out what this means? Rather you will have to modify the proof above.

The two proposition imply, just as in the finite dimensional case, the following:

Corollary 28.9. *Let V be a nonzero vector space over a field F and $\mathcal{B}$ a subset of V. Then the following are equivalent:*

(1) $\mathcal{B}$ is a basis for V.
(2) $\mathcal{B}$ s a maximal linearly independent set in V (relative to $\subset$).
(3) $\mathcal{B}$ is a minimal spanning set for V (relative to $\supset$).

As a further application of Zorn's Lemma, we next give a proof of the crucial result Theorem 15.13 needed in Section 15. The proof is a typical Zorn's Lemma argument used to extend maps.

Proposition 28.10. *Let G be a divisible additive group and $\theta : A \to B$ a group monomorphism of additive groups. If $\varphi : A \to G$ is a group homomorphism, then there exists a group homomorphism $\psi : B \to G$ satisfying $\varphi = \psi \circ \theta$.*

Proof. We may replace A by $\theta(A)$, i.e., assume that $A \subset B$. Let

$$\begin{aligned}\mathcal{S} = \{(C, \psi_C) \mid A \subset C \subset B,\\ \psi_C : C \to G \text{ a group homomorphism with } \varphi_C|_A = \varphi\}.\end{aligned}$$

Partially order $\mathcal{S}$ by

$$(C, \psi_C) \leq (C', \psi'_{C'}) \text{ if } C \subset C' \text{ and } \psi_{C'}|_C = \psi_C.$$

The set $\mathcal{S}$ is not empty, since $(A, \varphi) \in \mathcal{S}$. Let $\mathcal{C}$ be a chain in $\mathcal{S}$ and set $C = \bigcup_{(C_\alpha, \psi_{C_\alpha}) \in \mathcal{C}} C_\alpha$. Define $\psi_C : C \to G$ by $x \mapsto \psi_{C_\alpha}(x)$ if $x \in C_\alpha$. Since $\mathcal{C}$ is a chain, ψ_C is well-defined. Therefore, $(C, \psi_C) \in \mathcal{S}$ is an upper bound for $\mathcal{C}$. By Zorn's Lemma, there exist a maximal element $(C, \psi_C) \in \mathcal{S}$.

Claim. $C = B$:

Suppose the claim does not hold. Then there exists an $x \in B \setminus C$.

Case 1. $\langle x \rangle \cap C = 0$:

In this case $C + \langle x \rangle = C \oplus \langle x \rangle$. Define $\widetilde{\psi} = \psi \oplus 0$, i.e., $(c, rx) \mapsto \psi_C(c)$. Then $(C, \psi_C) < (C \oplus \langle x \rangle, \widetilde{\psi})$ in $\mathcal{S}$, contradicting the maximality of (C, ψ).

Case 2. $\langle x \rangle \cap C \neq 0$:
Choose $n \in \mathbb{Z}^+$ minimal such that $nx \in C$. Since G is divisible, there exists $y \in G$ satisfying $ny = \psi_C(nx)$. Define the map $\widetilde{\psi} : C + \langle x \rangle \to G$ defined by $c + rx \mapsto \psi_C(c) + ry$. Then $\widetilde{\psi}$ is a group homomorphism extending ψ to $C + \langle x \rangle$. As $(C, \psi_C) < (C + \langle x \rangle, \widetilde{\psi})$ in $\mathcal{S}$, this contradicts the maximality of (C, ψ). $\square$

We next apply Zorn's Lemma to Ring Theory.

Proposition 28.11. *Let R be a nontrivial ring and $\mathfrak{A} < R$ an ideal. Then there exists a maximal ideal $\mathfrak{m}$ in R such that $\mathfrak{A} \subset \mathfrak{m}$. In particular, any nontrivial ring contains maximal ideals.*

Proof. Let

$$S = \{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} < R \text{ an ideal}\}.$$

The set S is a poset under $\subset$, and S is nonempty as $\mathfrak{A} \in S$. Let $\mathcal{C}$ be a chain in S, so

$$\text{for all } \mathfrak{B}' \text{ and } \mathfrak{B}'' \text{ in } \mathcal{C} \text{ either } \mathfrak{B}' \subset \mathfrak{B}'' \text{ or } \mathfrak{B}'' \subset \mathfrak{B}'.$$

Claim. $\bigcup_{\mathfrak{B} \in \mathcal{C}} \mathfrak{B} < R$ is an ideal and hence an upper bound for $\mathcal{C}$ in S:
Let $r \in R$ and $x, y \in \bigcup_{\mathcal{C}} \mathfrak{B}$. Then there exist $\mathfrak{B}'$ and $\mathfrak{B}''$ in $\mathcal{C}$ with $x \in \mathfrak{B}' \in \mathcal{C}$ and $y \in \mathfrak{B}'' \in \mathcal{C}$. As $\mathcal{C}$ is a chain, we may assume that $\mathfrak{B}' \subset \mathfrak{B}''$, so $rx, xr, x \pm y \in \mathfrak{B}'' \subset \bigcup_{\mathcal{C}} \mathfrak{B}$, showing $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal. If $\bigcup_{\mathcal{C}} \mathfrak{B}$ is the unit ideal, there exists a $\mathfrak{B}'$ in $\mathcal{C}$ such that $1 \in \mathfrak{B}'$. This means that $\mathfrak{B}'$ is the unit ideal, so not in S, a contradiction. By Zorn's Lemma, there exists an ideal $\mathfrak{m} \in S$ that is a maximal element. So $\mathfrak{m}$ satisfies $\mathfrak{A} \subset \mathfrak{m} < R$ is an ideal, maximal with respect to this property. Thus $\mathfrak{m}$ is a maximal ideal of R containing $\mathfrak{A}$. $\square$

Remark 28.12. The reason the above worked was the claim, i.e.,

(i) The union of ideals in a chain is an ideal under the partial order $\subset$.
[The same is true for chains of subgroups, subspaces of a vector space, and other algebraic objects.]
Note. This does not work for arbitrary unions.

(ii) There exists an element lying in none of the elements in any chain (in this case the element 1).
In fact, if R is a nontrivial rng (even commutative), it may not have maximal ideals.

We next want to give further application of Zorn's Lemma to commutative ring theory that generalizes the existence of maximal ideals and which is quite useful. We begin with an important definition.

Definition 28.13. Let R be a nontrivial commutative ring, S a nonempty subset of R. We call S a *multiplicative set* if

(i) $1 \in S$.

(ii) If s_1 and s_2 are elements of S, then $s_1 s_2$ lies in S, i.e., S is closed under multiplication.

If T is a nonempty subset of R, we say that T *excludes* S if $S \cap T = \emptyset$.

Examples 28.14. Let R be a nontrivial commutative ring, $a \in R$, and $\mathfrak{A} < R$ an ideal.
1. $\mathfrak{A}$ is a prime ideal if and only if $R \setminus \mathfrak{A}$ is a multiplicative set.

2. $\{a^n \mid n \geq 0\}$ is a multiplicative set.
3. Let S be a multiplicative set with $0 \notin S$. Then S contains no *nilpotent elements*, i.e., an element $x \in R$ that satisfies $x^n = 0$ for some $n \in \mathbb{Z}^+$.
4. $R^\times$ is a multiplicative set.
5. An element $x \in R$ is called a *zero divisor* of R if there exists a nonzero element $y \in R$ satisfying $xy = 0$. Let

$$\mathrm{zd}(R) := \{x \in R \mid x \text{ is a zero divisor}\}.$$

Then $R \setminus \mathrm{zd}(R)$ is a multiplicative set.

Using the concept of multiplicative sets, we can now prove the generalization of the existence of maximal ideals that we seek. The proof will be just as the one for the existence of maximal ideals but using the ideal analogue for prime ideals of Euclid's Lemma given in Lemma 26.20.

Theorem 28.15. (Krull) *Let R be a commutative ring and S a multiplicative set. Suppose that $\mathfrak{A} < R$ is an ideal in R excluding S. Then there exists an ideal $\mathfrak{B} \subset R$ containing $\mathfrak{A}$ excluding S and maximal with respect to these two properties. Moreover, any such $\mathfrak{B}$ is a prime ideal.*

Note: If R is a nontrivial commutative ring, applying the above to the multiplicative set $R^\times$ shows the existence of maximal ideals in a commutative ring.

Proof. Let

$$\mathcal{I} = \{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} < R \text{ an ideal excluding } S\}$$

We know that $\mathcal{I}$ is not empty as $\mathfrak{A} \in \mathcal{I}$. Partially order $\mathcal{I}$ by $\subset$. Let $\mathcal{C}$ be a chain in $\mathcal{I}$.

Claim. $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal in R excluding S (hence not the unit ideal). In particular, $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an upper bound for $\mathcal{C}$ in $\mathcal{I}$:

$\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal exactly as before. If $b \in S \cap \bigcup_{\mathcal{C}} \mathfrak{B}$, then there exists an ideal $\mathfrak{B} \in \mathcal{C}$ such that $b \in S \cap \mathfrak{B}$, a contradiction. This establishes the claim.

By Zorn's Lemma, there exists an ideal $\mathfrak{B}$ in $\mathcal{I}$ satisfying

(*i*) $\mathfrak{B} \cap S = \emptyset$.
(*ii*) $\mathfrak{A} \subset \mathfrak{B}$.
(*iii*) If $\mathfrak{B} < \mathfrak{B}' \subset R$ is an ideal, then $\mathfrak{B}' \cap S \neq \emptyset$.

To finish, we must show that $\mathfrak{B}$ is a prime ideal. Suppose this is not the case. Then by Lemma 26.20, there exist ideals $\mathfrak{B} < \mathfrak{B}_i$, $i = 1, 2$, satisfying $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{B}$. By (*iii*), there exist $s_i \in \mathfrak{B}_i \cap S$, $i = 1, 2$. As S is a multiplicative set, $s_1 s_2 \in \mathfrak{B} \cap S$, a contradiction. □

Note the last part of the proof is structurally similar to factoring integers into primes. This is an argument that is often used.

The theorem allows us to begin to see the importance of prime ideals in understanding the structure of a commutative ring.

Definition 28.16. Let R be a nontrivial commutative ring. The *nilradical* of R is the set

$$\mathrm{nil}(R) := \{a \mid a \text{ is a nilpotent element in } R\}.$$

Corollary 28.17. *Let R be a nontrivial commutative ring. Then* $\mathrm{nil}(R)$ *is an ideal in* R *and*

$$\mathrm{nil}(R) = \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}.$$

PROOF. We leave it as an exercise to prove $\mathrm{nil}(R)$ is a ideal.

$\mathrm{nil}(R) \subset \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}$: If $a \in \mathrm{nil}(R)$, there exists an $n \in \mathbb{Z}^+$ such that $a^n = 0$ and $0 \in \mathfrak{A}$ for every ideal, in particular for prime ideals. Hence $a \in \mathfrak{p}$ for all prime ideals $\mathfrak{p}$ in R.

$\bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p} \subset \mathrm{nil}(R)$: Suppose this is not true, then there exists an element

$$a \in \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p} \setminus \mathrm{nil}(R).$$

In particular, the multiplicative set $S = \{a^n \mid n \geq 0\}$ does not contain 0 hence the trivial ideal (0) excludes S. By the theorem of Krull, there exists a prime ideal $\mathfrak{p}$ excluding S. Thus $a \notin \mathfrak{p}$, a contradiction.

Therefore, $\mathrm{nil}(R) = \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}$ as needed. □

Note. If R is a domain, then $\mathrm{nil}(R) = 0$. The converse is false, e.g., $\mathbb{Z} \times \mathbb{Z}$ has trivial nilradical, but is not a domain. A commutative ring whose nilradical is trivial is called a *reduced ring.*

Remark 28.18. As the intersection of a chain of prime ideals in a nonzero commutative ring R is clearly a prime ideal, there exist minimal elements in the set of all prime ideals by Zorn's Lemma. We call such a prime a *minimal prime ideal* of R. As every prime ideal in R contains a minimal prime, we have

$$\mathrm{nil}(R) = \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a minimal prime ideal}}} \mathfrak{p}.$$

Exercises 28.19.

1. Let V be a finite dimensional vector space over $\mathbb{R}$ with ordered basis $\{v_1, \ldots, v_n\}$. Define a lexicographic order of V relative to this ordered basis.
2. Prove Proposition 28.8.

3. Rngs may not have maximal ideals. This problem constructs one. Let p be a prime number. Let $\mathbb{Z}_{p^\infty}$ be the (additive) subgroup of $\mathbb{Q}/\mathbb{Z}$ consisting of all elements having order some power of p, i.e., $x = \alpha + \mathbb{Z}$, $\alpha \in \mathbb{Q}$, lies in $\mathbb{Z}_{p^\infty}$ if and only if $p^r x = 0$ in $\mathbb{Q}/\mathbb{Z}$, i.e., $p^r\alpha$ is an integer, for some positive integer r.
   (i) Show that the set $\{\frac{1}{p^r} + \mathbb{Z} \mid r \text{ a non-negative integer }\}$ generates $\mathbb{Z}_{p^\infty}$.
   (ii) Show the subgroup $\left\langle \frac{1}{p^r} + \mathbb{Z} \right\rangle$ of $\mathbb{Z}_{p^\infty}$ is isomorphic to $\mathbb{Z}/p^r\mathbb{Z}$.
   (iii) Show that any subgroup of $\mathbb{Z}_{p^\infty}$ is $\left\langle \frac{1}{p^r} + \mathbb{Z} \right\rangle$ for some non-negative integer r and the subgroups of $\mathbb{Z}_{p^\infty}$ form a chain under set inclusion.
   (iv) Make $\mathbb{Z}_{p^\infty}$ into a rng by defining $x \cdot y = 0$ for all $x, y \in \mathbb{Z}_{p^\infty}$. Show that $\mathbb{Z}_{p^\infty}$ has no maximal ideals
4. Let $\mathcal{C}$ be a chain of prime ideals under inclusion in a commutative ring R. Show that $\bigcap_{\mathcal{C}} \mathfrak{p}$ and $\bigcup_{\mathcal{C}} \mathfrak{p}$ are prime ideals.
5. Let R be a commutative ring and $\mathfrak{A}$ an ideal in R contained in a prime ideal $\mathfrak{P}$. Prove that there exists a prime ideal $\mathfrak{p}$ in R satisfying $\mathfrak{A} \subset \mathfrak{p} \subset \mathfrak{P}$ and with $\mathfrak{p}$ minimal among all prime ideals containing $\mathfrak{A}$.
6. Let $\mathfrak{p} < \mathfrak{P}$ be prime ideals in a commutative ring R. Show that there exist prime ideals $\mathfrak{p}_0$ and $\mathfrak{P}_0$ satisfying $\mathfrak{p} < \mathfrak{p}_0 < \mathfrak{P}_0 < \mathfrak{P}$ with no primes properly between $\mathfrak{p}_0$ and $\mathfrak{P}_0$.
7. Let R be a commutative ring. Prove the following:
   (i) Let $\mathcal{S}$ be the set of non finitely generated ideals in R and suppose that it is not empty. Let $\mathfrak{A}$ be a maximal element in $\mathcal{S}$. Then $\mathfrak{A}$ is a prime ideal.
   (ii) Let $\mathfrak{A}$ be a non finitely generated ideal in R. Suppose an ideal $\mathfrak{B}$ in R has the property that it contains $\mathfrak{A}$, is not finitely generated, and is maximal with respect to this property. Show that there exists a prime ideal in R containing $\mathfrak{A}$ that is not finitely generated.
8. Let R be a commutative ring and $A = R[[t]]$, the power series over R. Let $\varphi : A \to R$ be the ring epimorphism induced by $t \mapsto 0$ (evaluation at 0). Show the following:
   (i) If $\mathfrak{A}$ is an ideal in R, then $\varphi^{-1}(\mathfrak{A}) = \mathfrak{A}A + tA$.
   (ii) φ induces a bijection between the set of maximal ideals in A and the set of maximal ideals in R.
9. As in the previous problem, let R be a commutative ring, $A = R[[t]]$, and $\varphi : A \to R$ be the ring epimorphism induced by $t \mapsto 0$. Show the following:
   (i) Let $\mathfrak{P}$ be a prime ideal in A. Suppose that $\varphi(\mathfrak{P}) = (a_1, \dots, a_n)$. Then $\mathfrak{P} = Aa_1 + \cdots + Aa_n + At$ if $t \in \mathfrak{P}$ and $\mathfrak{P} = (f_1, \dots, f_n)$ for $f_i \in A$ with $f_i = a_i +$ higher terms in t if $t \notin \mathfrak{P}$.
   (ii) Every prime ideal in A can be generated by finitely many elements if and only if every prime ideal in R can be generated by finitely many elements.
10. Let R be a nonzero commutative ring. Show that the set of nonzero divisors is a multiplicative group.
11. Let R be a commutative ring. Prove
   (i) If x is nilpotent in R, then $1 + x$ is a unit in R.

(*ii*) The nilradical of R is an ideal.
(*iii*) Compute the nilradical of the rings: $\mathbb{Z}/12\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$, $n > 1$, and $\mathbb{Z}$.

12. Let R be a commutative ring. The *Jacobson radical* of R is defined to be $\mathrm{rad}(R) := \bigcap_{\mathrm{Max}(R)} \mathfrak{m}$, the intersection of all maximal ideals in R. Show that x lies in $\mathrm{rad}(R)$ if and only if $1 - yx$ is a unit in R for all y in R.
13. Let R be an infinite domain that has only finitely many units. Show that R must have infinitely many maximal ideals.
14. Let R be a commutative ring and $\mathfrak{A} < R$ an ideal. Define the *radical* of $\mathfrak{A}$ to be the set
$$\sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ for some } n \in Z^+\}.$$
Show the following:
(*i*) $\sqrt{\mathfrak{A}}$ is an ideal and
$$\sqrt{\mathfrak{A}} = \bigcap_{\substack{\mathfrak{A} \subset \mathfrak{p} < R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}.$$
(*ii*) Let $^{-} : R \to R/\mathfrak{A}$ be the canonical ring epimorphism, then $\mathrm{nil}(\overline{R}) = \sqrt{\mathfrak{A}}/\mathfrak{A}$.
15. Let R be a commutative ring, $\mathfrak{A} < R$ an ideal, and $^{-} : R \to R/\mathfrak{A}$ be the canonical epimorphism. We say that $\mathfrak{A}$ is a *primary ideal* if $ab \in \mathfrak{A}$ implies that $a \in \mathfrak{A}$ or $b^n \in \mathfrak{A}$ for some positive integer n. Let $\mathfrak{A} < R$ be an ideal. Show both of the following:
(*i*) $\mathfrak{A}$ is a primary ideal if and only if every zero divisor of $R/\mathfrak{A}$ is nilpotent.
(*ii*) If $\mathfrak{A}$ is primary, then its radical, $\sqrt{\mathfrak{A}}$, (see the previous exercise) is a prime ideal.
16. Determine all primary ideals if R is a PID. (Cf. the previous exercise for the definition of a primary ideal.)
17. Let R be a commutative ring. An element e of R is called an *idempotent* if $e^2 = e$. For example, if S is another commutative ring the element $(1_R, 0_S)$ is an idempotent in the ring $R \times S$. The object of this exercise is to prove a converse. Let e be an idempotent of R. Then prove
(*i*) $e' := 1 - e$ is an idempotent of R.
(*ii*) The principal ideal Re of R is a ring with identity $1_{Re} = e$. [Note that Re is not a subring of R since Re will not have the same identity as R unless $e = 1$.]
(*iii*) R is ring isomorphic to $Re \times Re'$.

29. Localization

In section 27, we showed that a field of quotients exists for any domain by defining fractions. The purpose was to create a ring containing the given ring R in which every nonzero element of R has a multiplicative inverse. [Note if 0 has an inverse then the ring must be trivial.] Instead, we can try to invert only some elements. This turns out to be a crucial idea, and the method is to invert elements in a multiplicative set. But to make it effective, we must weaken what we mean when we say two 'fractions' are equal.

Construction 29.1. We shall leave most of the details of this construction as an exercise, an exercise that you should do. Let R be a commutative ring and S a multiplicative set in R. Let
$$\mathcal{S} = \{(a, s) \mid a \in R,\, s \in S\} = R \times S.$$

Define a relation on $\mathcal{S}$ as follows: Write

$$(a, s) \sim (a', s') \text{ in } \mathcal{S}$$

if there exists an $s'' \in \mathcal{S}$ satisfying $s''(as' - a's) = 0$ in R.
Then $\sim$ is a equivalence relation. Let

$$\frac{a}{s} \text{ denote the equivalence class of } (a, s)$$

and set

$$S^{-1}R = \{\frac{a}{s} \mid a \in R,\, s \in S\} = S/\sim.$$

Define

$$\frac{a}{s_1} + \frac{b}{s_2} := \frac{as_2 + bs_1}{s_1 s_2}$$
$$\frac{a}{s_1} \cdot \frac{b}{s_2} := \frac{ab}{s_1 s_2}$$

for all $a, b \in R$ and for all $s_1, s_2 \in S$. Then

$$+ : S^{-1}R \times S^{-1}R \to S^{-1}R \quad \text{and} \quad \cdot : S^{-1}R \times S^{-1}R \to S^{-1}R$$

are well-defined maps making $S^{-1}R$ into a commutative ring, called the *localization* of R at S with $0_{S^{-1}R} = \frac{0}{s}$ and $1_{S^{-1}R} = \frac{s}{s}$, for any $s \in S$. The map

$$\varphi : R \to S^{-1}R \text{ given by } r \mapsto \frac{r}{1}$$

is a ring homomorphism.

Examples 29.2. Let R be a commutative ring and S a multiplicative set in R.

1. If $0 \in S$, then $S^{-1}R$ is the trivial ring.
2. Suppose that R is a domain. Then $T = R \setminus \{0\}$ is a multiplicative set and $T^{-1}R$ is the quotient field of R.
3. Let $\varphi : R \to S^{-1}R$ be the ring homomorphism given by $r \mapsto \frac{r}{1}$. Then
   (*i*) $\varphi(S) \subset (S^{-1}R)^\times$.
   (*ii*) $\ker \varphi = \{x \in R \mid \text{there exists an } s \text{ in } S \text{ such that } sx = 0\}$.
   (*iii*) φ is a ring monomorphism if and only if $S \cap \mathrm{zd}(R) = \emptyset$.
   (*iv*) If R is a domain and S does not contain zero, then φ is a ring monomorphism. Note if this is the case that $\frac{a}{s} = \frac{b}{s}$ in $S^{-1}R$ if and only if $a = b$. We then view φ as the inclusion map, so $R \subset S^{-1}R \subset qf(R)$ (cf. (2)).
4. Let $\mathfrak{p}$ be a prime ideal in R. Then $T = R \setminus \mathfrak{p}$ is a multiplicative set and the localization $T^{-1}R$ is denoted $R_\mathfrak{p}$. For example, if p is a prime in $\mathbb{Z}$ and $T = \{n \in \mathbb{Z} \mid p \not| n\}$, then $\mathbb{Z}_{(p)} = \{\frac{n}{m} \mid n, m \in \mathbb{Z} \text{ relatively prime with } p \not| m\}$.
5. Let f be an element of R. Then $T = \{f^n \mid n \geq 0\}$ is a multiplicative set and the localization $T^{-1}R$ is denoted R_f. For example, if p is a prime in $\mathbb{Z}$, then $\mathbb{Z}_p = \{\frac{n}{m} \mid n, m \in \mathbb{Z} \text{ relatively prime with } m = p^k \text{ some } k \geq 0\}$.

The existence of the localization of a commutative ring at a multiplicative set is a basic construction in commutative algebra whose importance cannot be overestimated. Indeed it, together with the construction of polynomial rings and quotient rings over a commutative ring are the basic constructions in commutative algebra. It is especially important when the multiplicative set is the complement of a prime ideal. For example, the ring $\mathbb{Z}_{(p)}$ in Example 29.2(4) has only one nonzero prime ideal and all prime integers relatively prime to p become units. One can view $\mathbb{Z}_{(p)} \subset \mathbb{Q}$. It turns out that in this ring every element can be written up^n for some unit u in $\mathbb{Z}_{(p)}$ and some positive integer n. In the more general case that R is a commutative ring and $\mathfrak{p}$ a prime ideal in R, let $\mathfrak{p}R_{\mathfrak{p}}$ denote the image of $\mathfrak{p}$ in $R_{\mathfrak{p}}$. Then $\mathfrak{p}R_{\mathfrak{p}}$ is the unique maximal ideal in $R_{\mathfrak{p}}$. One can obtain valuable information for R from $R_{\mathfrak{p}}$ via the natural map $R \to R_{\mathfrak{p}}$ just as one get valuable information for R from the canonical epimorphism $R \to R/\mathfrak{p}$.

Localization serves as the algebraic analogue of the geometric study of germs of functions at a point and of functions in neighborhoods of a point that we shall now discuss.

Example 29.3. Let $C([0,1])$ denote the ring of continuous real-valued functions on $[0,1]$. Exercise 26.21(13) said that the maximal ideals in $C([0,1])$ are in one to one correspondence with $[0,1]$, the correspondence given by

$$a \mapsto \mathfrak{m}_a = \{f \in C([0,1]) \mid f(a) = 0\}.$$

[This is true with $[0,1]$ replaced by any finite closed interval (or any compact subset) of the real line.] However, if we look at $C((0,1])$, we have problems with functions that should have a limit as $x \to 0^-$. For example, the function may be unbounded or oscillate wildly, e.g., $\sin(1/x)$. A continuous function defined on an interval I is called *uniformly* continuous if given any $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $x, y \in I$ satisfies $|x - y| < \delta$. If a function is continuous on a closed bounded interval in $\mathbb{R}$, it is uniformly continuous. Let

$$C_u((0,1]) := \{f \in C((0,1]) \mid f \text{ uniformly continuous on } (0,1]\},$$

a subring of $C((0,1])$. If $f \in C_u((0,1])$, then $\lim_{x\to 0^-} f(x)$ exists and

$$\mathfrak{A} = \{f \in C_u((0,1]) \mid \lim_{x\to 0^-} f(x) = 0 \text{ exists}\}$$

will putatively be the maximal ideal that would have been associated to the point 0 if f was defined at 0 or at least contained in a maximal ideal by Krull's theorem 28.15.

We look at this a bit more algebraically. Limits are local phenomena, so we want to determine an equivalence relation that determines when two functions are 'close' at point to replace the maximal ideals $\mathfrak{m}_a$ that we found in $C([0,1])$.

If f is a continuous function defined on a nonempty open subset U of $[0,1]$, there is always a *global* continuous function $g \in C([0,1])$ such that $g|_U = f$. This type of statement is not true in general. Indeed there may not be sufficiently many global functions. For example, if $X = \mathbb{C}$, by a theorem of Liouville, the only analytic (holomorphic) functions defined on all of $\mathbb{C}$ are the constant functions. So we wish to look at functions on a topological space X that do not necessarily have this property.

Let X be an arbitrary topological space. Let R be a set of functions each function defined on some nonempty open subset of X and $R(U)$ be the set of functions defined on an open set U of X. Suppose that $x \in X$ and f and g are two functions in R such that there exists a nonempty open neighborhood U of x on which both f and g are defined. Write $f \sim g$ if there exists a nonempty open neighborhood $V \subset U$ of x such that $f|_V = g|_V$. This is an equivalence relation, and the equivalence classes of such functions at x is called the *germ* of functions at x. Suppose for every open set U in X the set $R(U)$ is a commutative ring, e.g., X is a topological or differentiable manifold with continuous, respectively differentiable functions. If $x \in X$ let $\mathfrak{p}_x$ be the set of all the functions zero at x. Then $\mathfrak{p}_x$ is a prime ideal and the set of germs at $x \in U$, is the maximal ideal in $R(U)_{\mathfrak{p}_x}$, the localization $R(U)$ at $\mathfrak{p}_x$. This includes the case $X = [0, 1]$ if $x \in X$. In the case of $X = (0, 1)$, we needed to look at uniformly continuous functions. Usually, one needs to look at functions with what are called compact support. [A closed and bounded subset of $\mathbb{R}^n$ is a compact set by the Bolzano-Weierstraß Theorem.] If $R(U)$ is a ring for all open U in X as above and $R(U)_{\mathfrak{p}_x}$ exists then (X, R) is called a *locally ringed space*. Topological and geometric manifolds as well as algebraic varieties are examples of such. The R having the properties for each such U is an example of as *sheaf* of functions. The local rings are called *stalks*.

Exercises 29.4.

1. Prove all of the claims in Construction 29.1.
2. Let R be a commutative ring and S a multiplicative set in R. Let $\varphi : R \to S^{-1}R$ be given by $r \mapsto \frac{r}{1}$. Show that this satisfies the following *universal property*: If $\psi : R \to R'$ is a ring homomorphism with R' commutative and $\psi(S)$ a subset of the unit group of R', then there exists a unique ring homomorphism $\theta : S^{-1}R \to R'$ such that

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[dr, "\psi"'] & S^{-1}R \arrow[d, "\theta"] \\ & R' \end{tikzcd}$$

   commutes.
3. Let R be a commutative ring, S a multiplicative set in R. Show if $\mathfrak{p}$ is a prime ideal in R excluding S, then $S^{-1}\mathfrak{p}$ is a prime ideal, and every prime in $S^{-1}R$ is of this form. Use this to show if $0 \notin S$, then there exists a prime ideal in R excluding S. (Cf. this to Theorem 28.15.)
4. Let $\mathfrak{p}$ be a prime ideal in a commutative ring R. Show that $R_{\mathfrak{p}}$ has a unique maximal ideal. What is it? A commutative ring with a unique maximal ideal is called a *local ring*.
5. Let R be a nonzero commutative ring and $S \subset R$ be the multiplicative set of nonzero divisors in R. Show that $\varphi : R \to S^1R$ by $r \mapsto \frac{r}{1}$ is a ring monomorphism and $S^{-1}R$ is the largest ring with this property. We call this localization of R the *total quotient ring* of R or the *total ring of fractions* of R and view $R \subset S^{-1}R$.
6. Let $\mathfrak{A}$ be an ideal in a commutative ring R and S a multiplicative set in R. Define
$$S^{-1}\mathfrak{A} := \{\frac{a}{s} \mid a \in \mathfrak{A},\ s \in S\}.$$
   Show that $S^{-1}\mathfrak{A}$ is an ideal in $S^{-1}R$ and not the unit ideal if $\mathfrak{A}$ excludes S.

[Note. If $\dfrac{x}{s}$ lies in $S^{-1}\mathfrak{A}$, this does not mean that $x \in \mathfrak{A}$, only that there exist an $a \in \mathfrak{A}$ and an $s' \in S$ such that $\dfrac{x}{s} = \dfrac{a}{s'}$.]

7. Let R be a commutative ring, S a multiplicative set in R, and $\varphi : R \to S^{-1}R$ the ring homomorphism given by $r \mapsto \frac{r}{1}$. Suppose that $\mathfrak{B}$ is an ideal in $S^{-1}R$. Show that $\mathfrak{A} = \varphi^{-1}(\mathfrak{B})$ is an ideal in R and $S^{-1}\mathfrak{A} = \mathfrak{B}$.
8. Let R be a domain, S a multiplicative set in R not containing 0. Then we can view $S^{-1}R$ as a subring of the quotient field of R. Prove that

$$R = \bigcap_{\operatorname{Max}(R)} R_{\mathfrak{m}}$$

where $\operatorname{Max}(R)$ is the set of maximal ideals in R.

CHAPTER VI

Domains

In this chapter, we study in greater depth special domains. The main goal is to study those domains that satisfy the analogue of the Fundamental Theorem of Arithmetic called Unique Factorization Domains or UFDs and explicit examples of these domains. The UFDs include PIDs and the historically interesting special class of PIDs that satisfy the division algorithm called *euclidean domains*. The explicit example of a euclidean domain, the Gaussian integers, is used to prove the classical two square theorems in Number Theory. We also prove Lagrange's Four Square Theorem to indicate the method of proof developed by Fermat's called Infinite Descent (a form of induction) to prove this theorem. We also introduce a more general class of commutative rings called Noetherian rings. These are the most important commutative rings in algebra and in algebraic geometry (as well as historically the most important because of the work of Hilbert). We shall study these rings in greater detail in a later chapter. For purposes here they represent a generalization of PIDs. From the viewpoint of induction, they represent rings where finite induction has a direct analogue. Theorems proven using Zorn's Lemma over a commutative ring, usually do not need it if the ring is Noetherian.

30. Special Domains

We have seen that prime elements in $\mathbb{Z}$ are those elements that satisfy one of two equivalent conditions: Let $a, b \in \mathbb{Z}$. Then $p \neq 0$ is a prime if

1. whenever $p = ab$, then either $a = \pm 1$ or $b = \pm 1$, i.e., either a or b is a unit or
2. whenever $p \mid ab$ then $p \mid a$ or $p \mid b$.

The first was the original definition, the second Euclid's Lemma. We have generalized the concept of a prime element in $\mathbb{Z}$ to the concept of a prime ideal, but have not generalized the first equivalence in this way, nor will we, as its importance is related to elements. Elements satisfying the generalization of the first condition will be called *irreducible elements*. That elements in a domain factor into a product of irreducible elements (up to units) holds for a large class of domains. What does not hold in general and what makes the two conditions not equivalent is an appropriate uniqueness statement. We shall carefully study domains in which we do have such a uniqueness statement. In general the second condition above is the more important as it gives more information about rings.

Definition 30.1. Let R be a domain and a a nonzero nonunit in R. We say a is *irreducible* if whenever $a = bc$ with $b, c \in R$, then either b or c is a unit in R (if a is not irreducible, we say that it is *reducible*) and is a *prime element* if whenever $a \mid bc$ with $b, c \in R$, then $a \mid b$ or $a \mid c$.

Note that any associate of an irreducible element (respectively, prime element) is irreducible (respectively, prime).

Remark 30.2. Let R be a domain and $p \in R$ a prime element. Then p is irreducible i.e., being a prime element is stronger than being an irreducible element.

PROOF. Suppose that $a = bc$ with $b, c \in R$. As a is a prime element, either $a \mid b$ or $a \mid c$, say $a \mid b$. Then $b = ax$ some nonzero element $x \in R$. Therefore, we have $b = ax = bcx$ in the domain R, so cancellation implies that $1 = cx$, i.e., $c \in R^\times$ as needed. □

We want to generalize the Fundamental Theorem of Arithmetic to a wider class of domains. Before, our approach relied on the concept of greatest common divisors. In general, our notion of a gcd of two elements below in a domain may not exist. Moreover, for integers, a gcd was always unique because we could make it positive as $\mathbb{Z}^\times = \{\pm 1\}$. In general, we cannot do this, so must omit this condition that a gcd is unique if it exists.

Definition 30.3. Let R be a domain, a, b and d nonzero elements in R. We call d a *greatest common divisor* or *gcd* of a and b if:

(i) $d \mid a$ and $d \mid b$ in R.

(ii) If c in R satisfies $c \mid a$ and $c \mid b$ in R, then $c \mid d$ in R.

If a gcd of a and b exists and is a unit, then we say that a and b are *relatively prime*.

Remarks 30.4. Let R be a domain and a, b nonzero elements in R.

1. As mentioned above, a gcd of a, b may not exist. For example, there exist elements in the subring $R := \{x + 2y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ of the Gaussian integers $\mathbb{Z}[\sqrt{-1}] := \{x + y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ not having a gcd.
2. If a gcd d of a, b exists it may not satisfy $d = ax + by$ for some $x, y \in R$, as it did for $\mathbb{Z}$.
3. If $d \in R$ satisfies $(d) = (a, b)$ then d is a gcd of a and b (and, of course, there exist $x, y \in R$ satisfying $d = ax + by$).
4. If d and d' are two gcd's of a and b in R, then $d \mid d'$ and $d' \mid d$, hence $d \approx d'$, i.e., $(d) = (d')$. Moreover, if this is the case, then any *generator* c of (d), i.e., an element c in R satisfying $(c) = (d)$, is a gcd of a and b. In particular, a gcd of a and b, if it exists, is unique up to units in R.

We codify some of the remarks above in the following:

Proposition 30.5. *Let R be a PID, a and b nonzero elements in R. Let $(d) = (a, b)$. Then d is a gcd of a and b in R, unique up to units. In particular, gcd's of nonzero elements in a PID always exist and if d' is a gcd of a and b, then $d' = ax + by$ for some $x, y \in R$.*

The proposition allows us to extend the analogue of Euclid's Lemma to any PID.

Corollary 30.6. (Euclid's Lemma) *Let R be a PID and $r \in R$. Then r is irreducible if and only if r is a prime element.*

Proof. We have already shown that prime elements are irreducible. Conversely, assume that r is an irreducible element in R and satisfies $r \mid ab$ with $a, b \in R$. As R is a PID, there exists a $c \in R$ satisfying $(c) = (r, a)$. Write $r = xc$ with $x \in R$. As r is irreducible, either c is a unit or x is a unit in R. If c is a unit, then (c) is the unit ideal, hence $1 = ry + az$ in R for some $y, z \in R$. It follows that $r \mid ryb + abz = b$ as needed. If x is a unit then $(r) = (c) = (r, a)$, so $a \in (r)$, i.e., $r \mid a$. □

The Fundamental Theorem of Arithmetic relied on the equivalence of the notion of irreducible and prime elements. We wish to generalize this theorem to a wider class of domains, so we must say what we mean by uniqueness of factorization. For the integers this meant up to ± 1, the units in $\mathbb{Z}$, so we will use the following generalization.

Definition 30.7. Let R be a domain. We call R a *unique factorization domain* or *UFD* if for any nonzero nonunit r in R, we have

(i) $r = f_1 \cdots f_n$ for some irreducible elements $f_1, \ldots, f_n$ in R for some $n \in \mathbb{Z}^+$.

(ii) If $r = f_1 \cdots f_n = g_1 \cdots g_m$ with $f_1, \ldots, f_n, g_1, \ldots, g_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$, then $n = m$ and there exists a permutation $\sigma \in S_n$ satisfying $f_i \approx g_{\sigma(i)}$ for all i.

[Note that if $r = uf_1 \cdots f_n$ with u a unit in R and $f_i \in R$ for all i, then we can 'absorb' the unit by replacing any one of the factors with an associate.]

Remark 30.8. If R is a UFD, then, as desired, an element in R is prime if and only if it is irreducible.

Proof. We need only show if r is an irreducible element in R that it is a prime. So suppose that $r \mid ab$ with $a, b \in R$ nonunits. Write $a = f_1 \cdots f_n$ and $b = g_1 \cdots g_m$, with $f_1, \ldots, f_n, g_1, \ldots, g_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$. Since $ab = rx$ for some $x \in R$ by assumption, the uniqueness property in the definition of UFD's implies $r \approx f_i$ some i or $r \approx g_j$ some j, i.e., $r \mid a$ or $r \mid b$. □

The uniqueness property for a UFD is really just that irreducible elements are prime:

Proposition 30.9. (Euclid's Argument) *Let R be a domain. Suppose that $p_1 \cdots p_n = f_1 \cdots f_m$ with $p_1, \ldots, p_n$ prime elements in R and $f_1, \ldots, f_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$. Then $n = m$ and there exists a permutation $\sigma \in S_n$ satisfying $p_i \approx f_{\sigma(i)}$. In particular, each f_i is a prime element in R.*

Proof. As $p_1 \mid f_1 \cdots f_m$ in R, there exists an i such that $f_i = up_1$ for some element u in R. Since f_i is irreducible and p_1 not a unit, u is a unit in R. Now $p_1 \cdots p_n = up_1 f_2 \cdots f_m$ in the domain R. If $n = 1$ or $m = 1$ we must have $m = n = 1$ by cancellation, since the product of irreducibles is not a unit. So we may assume that $m, n > 1$. Cancellation shows $p_2 \cdots p_n = uf_2 \cdots f_m$, and the result follows by induction. □

An immediate consequence is the following:

Corollary 30.10. *Let R be a domain. Suppose that every nonzero nonunit in R is a product of (finitely many) prime elements, i.e., $a = p_1 \cdots p_n$, with $p_1, \ldots, p_n$ prime elements in R. Then R is a UFD.*

The most important class of commutative rings is the following:

Definition 30.11. A commutative ring R is called a *Noetherian ring* if one of the following equivalent conditions hold:

(i) R satisfies the *ascending chain condition* or *ACC*, viz., given a (countable) chain of ideals

$$\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \cdots \subset \mathfrak{A}_n \subset \cdots$$

in R, there exists a positive integer N such that $\mathfrak{A}_{N+i} = \mathfrak{A}_N$ for all $i \geq 0$.
[We say that the chain *stabilizes*.]
Equivalently, there exists no infinite chain of ideals in R,

$$\mathfrak{B}_1 < \mathfrak{B}_2 < \cdots < \mathfrak{B}_n < \cdots .$$

(ii) Every ideal $\mathfrak{A}$ in R is *finitely generated* (*fg*), i.e., there exist $a_1, \ldots, a_n \in R$, some n, such that $\mathfrak{A} = (a_1, \ldots, a_n)$.

(iii) R satisfies the *Maximal Principle*, i.e., any nonempty set of ideals S in R contains a maximal element relative to the relation $\subset$.
[Cf. this condition to the conclusion of Zorn's Lemma.]

Remark 30.12. We leave it as an important exercise (cf. Exercise 30.22(11)) to show these three conditions are equivalent. However, the proof of this really needs the Axiom of Choice (Appendix A (A.8)) to prove ACC implies the Maximal Principle. [In Appendix A, it is shown that Zorn's Lemma is equivalent to the Axiom of Choice.] In fact, it is also known that the Axiom of Choice is also equivalent to the implication that ACC implies the Maximal Principle. For Noetherian rings one usually uses the Maximal Principle instead of directly using Zorn's Lemma in proofs.

Using the second condition we have:

Examples 30.13. 1. Every PID is a Noetherian domain.

2. Let R be a commutative ring. By Exercise 28.19(7), R is Noetherian if and only if every prime ideal is finitely generated. Let $A = R[[t]]$. By Exercise 28.19(9), prime ideals in A are finitely generated if prime ideals in R are. [The converse is also true.] Hence if R is Noetherian, so is $R[[t]]$.

Noetherian rings are important as they are the rings that arise in classical algebraic geometry. The collection of Noetherian rings also has the properties of being closed under basic ring constructions (and combinations of them): viz.,

(i) The homomorphic image of a Noetherian ring is Noetherian.

(ii) The localization of a Noetherian ring is Noetherian.

(iii) (Hilbert Basis Theorem) If R is a Noetherian ring so is $R[t]$.

We shall leave the first two as exercises and prove the last in Section 41. (Cf. Theorem 41.1). It is not true, however, that a subring of a Noetherian ring is Noetherian. Indeed as any field is a Noetherian domain, one need only find a non-Noetherian domain to provide a counterexample. Can you do so?

We illustrate the importance of the Noetherian condition by the following interesting result.

Theorem 30.14. *Let R be a Noetherian domain and r a nonzero non-unit in R. Then r is a product of (finitely many) irreducible elements.*

Proof. Let

$$S = \{(a) \mid (0) < (a) < R \text{ with } a \text{ not a product of irreducible elements}\}.$$

Suppose that S is nonempty, i.e., the result is false. Then there exists a principal ideal $(a) \in S$, a maximal element by the Maximal Principle. We call (a) a *maximal counterexample*. Certainly, a cannot be irreducible, so $a = bc$ for some non-units $b, c \in R$. But then $(a) < (b)$ and $(a) < (c)$, so by maximality, both b and c are products of irreducible elements, hence so is $a = bc$. □

The above proof is called proof by *Noetherian induction*. Note the similarity between this proof and that showing positive integers factor into primes. Further application of this type of decomposition into 'smaller pieces' for Noetherian rings is given in the exercises.

Theorem 30.15. *Let R be a PID. Then R is a UFD.*

Proof. As R is a PID, it is a Noetherian domain. Hence every nonzero nonunit factors into a product of irreducible elements. As every irreducible element in a PID is a prime, factorization is (essentially) unique by Euclid's Argument. □

Corollary 30.16. *$\mathbb{Z}$ is a UFD.*

Next we generalize the Division Algorithm, to study a special type of PID.

Definition 30.17. Let R be a domain. Then R is called a *euclidean domain* if there exists a function

$$\partial : R \setminus \{0\} \to \mathbb{Z}^+ \cup \{0\}$$

satisfying the *Division Algorithm*, i.e., for all $a, b \in R$ with b nonzero, there exist $q, r \in R$ satisfying

(i) $$a = bq + r.$$

(ii) $$r = 0 \text{ or } \partial r < \partial b.$$

We call ∂ a *euclidean function*.

Remark 30.18. If we agree that $-\infty + n = -\infty$ for all integers n and let $\partial 0 = -\infty$, we can write (ii) as

$$\partial r < \partial b.$$

[This is often done.]

Remarks 30.19. 1. A euclidian function is sometimes defined to have the further property that

(iii) $$\partial(ab) \geq \partial(a) \text{ for all nonzero } a, b \in R.$$

If (iii) also holds, we shall call ∂ a *strong euclidean function*.
[If (iii) holds, then $\partial(ab) > \partial(a)$ if b is not a unit and $\partial(u) = \partial(1)$ for all units $u \in R^\times$.]

A euclidean domain may have more than one euclidean function. It can be shown that if a euclidean function ∂ satisfies (i) and (ii), then $\partial' : R \setminus \{0\} \to \mathbb{Z}^+ \cup \{0\}$ defined by $\partial'(x) = \min_{0 \neq r \in R} \partial(rx)$ is a strong euclidean function. In particular, every euclidean domain satisfies (iii) under some strong euclidean function.

2. Historically, a euclidean function ∂ on a euclidean domain was also required to be multiplicative, i.e., $\partial(ab) = \partial(a)\partial(b)$ for any nonzero $a, b \in R$. Such a euclidean function is called a *normed euclidean function*.

Theorem 30.20. *If R is a euclidean domain, then R is a PID, hence a UFD.*

PROOF. Let $0 < \mathfrak{A} \subset R$ be an ideal and

$$\emptyset \neq \mathcal{S} = \{\partial a \mid 0 \neq a \in \mathfrak{A}\} \subset \mathbb{Z}^+ \cup \{0\}$$

(as $\mathfrak{A} \neq 0$), By the Well-ordering Principle, there exists an element a in $\mathfrak{A}$ with $\partial a \in \mathcal{S}$ minimal.

Claim. $\mathfrak{A} = (a)$ [Cf. the proof that $\mathbb{Z}$ is a PID.]:

Let $b \in \mathfrak{A}$, then $b = aq + r$ in R with $r = 0$ or $\partial r < \partial a$. If $r \neq 0$, then $0 \neq r = b - aq$ lies in $\mathfrak{A}$, contradicting the minimality of ∂a so $r = 0$ and $\mathfrak{A} = (a)$. $\square$

Remarks 30.21. (no justification)

1. Any field is euclidean.
2. If F is a field, then the polynomial ring $F[t]$ is euclidean with normed euclidean function deg.
3. $\mathbb{Z}[t]$ is a UFD but not a PID.
4. $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.
5. $\mathbb{Z}[\sqrt{-1}]$, $\mathbb{Z}[\sqrt{-2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-3}}{2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-7}}{2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-11}}{2}]$ are normed euclidean domains.
6. Motzkin showed that $\mathbb{Z}[\frac{-1-\sqrt{-19}}{2}]$, was a PID but not a euclidean domain under a normed euclidean function. It is not a euclidean domain under a strong euclidean function.
7. $\mathbb{Z}(\sqrt{14}$ is a euclidean domain, but not a euclidean domain under a normed euclidean function.

Exercises 30.22.

1. Let R be a domain and a a nonzero nonunit in R. Show that a is irreducible if and only if the principal ideal (a) is maximal in the set $\{(b) \mid b \text{ a nonzero nonunit in } R\}$. In particular, if R is a PID, then every irreducible element in R is a prime element.
2. Produce elements a and b in the domain $R := \{x + 2y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ having no gcd. Prove your elements do not have a gcd.
3. Let $\mathbb{Z} \subset R$ be a subring with R a UFD. Let d be the gcd of two nonzero integers a and b in $\mathbb{Z}$. Show that d is still a gcd of a and b in R.
4. Show 1 is a gcd for 2 and t in $\mathbb{Z}[t]$, but there are no polynomials $f, g \in \mathbb{Z}[t]$ satisfying $1 = 2f + tg$.

5. Let R be a UFD and a a nonzero element in R. Show that the nilradical of $R/(a)$ is the intersection of a finite number of prime ideals in $R/(a)$, say $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$, and if $\mathfrak{P}$ is any prime ideal in $R/(a)$ then there exists an i such that $\mathfrak{p}_i \subset \mathfrak{P}$. (Cf. Exercise 28.19 (14).)
6. A domain R is called a *Bézout domain* if every finitely generated ideal in R is principal. [A Bézout domain may not be a PID. For example, The ring of entire (holomorphic) functions $f : \mathbb{C} \to \mathbb{C}$ can be shown to be a Bézout domain.] Show that every two nonzero elements in a Bézout domain has a gcd.
7. A domain R with quotient field F is called a *valuation ring* if for any nonzero element x in F, either $x \in R$ or $x^{-1} \in R$. Show that a valuation ring has a unique maximal ideal and is a Bézout domain.
8. A domain R is called a *GCD-domain* if every pair of nonzero elements in R has a gcd. Let R be a GCD-domain. If a, b are nonzero in R, write $[a, b]$ for a gcd of a, b. Of course, this is only unique up to units. Show all of the following holds (up to units) for all nonzero a, b, c, d in R:
   (i) $[ab, ac] = a[b, c]$.
   (ii) If $[a, d] = d$, then $[a/d, b/d] = 1$.
   (iii) If $[a, b] = [a, c] = d$, then $[ab, cd] = 1$.
9. Show if R is a GCD-domain, then an element in R is a prime element if and only if it is an irreducible element.
10. Let R be a domain with quotient field F. We say that R is a *normal domain* or an *integrally closed domain* if whenever a monic polynomial $f \in R[t]$ has a root x in F, then $x \in R$. (Cf. $R = \mathbb{Z}$.) Prove that if R is a GCD-domain, then it is a normal domain.
11. Prove that the three conditions defining a Noetherian ring in (30.11) are indeed equivalent.

    [Note: Technically to prove ACC implies the Maximal Principle, one needs the Axiom of Choice as mentioned in Remark 30.12. If you see where, you can invoke it.]
12. If in the previous exercise, you do not want to use the Axiom of Choice, but rather use Zorn's Lemma, show that if all ideals of a commutative ring are finitely generated, then the Maximal Principle holds using Zorn's Lemma.
13. Prove Theorem 30.14 using ACC and not the Maximal Principle.
14. Prove that any Noetherian valuation ring is a PID.
15. A commutative ring R is called *Artinian* if it satisfies the following condition (called the *descending chain condition* or *DCC*): Any chain of ideals

$$\mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n \supset \cdots$$

(countable) in R *stabilizes*, i.e, there exists an integer N such that $\mathfrak{A}_{N+i} = \mathfrak{A}_N$ for all $i \geq 0$. Equivalently, there exist no infinite chains

$$\mathfrak{B}_1 > \mathfrak{B}_2 > \cdots > \mathfrak{B}_n > \cdots .$$

Show that R is Artinian if and only if it satisfies the *Minimal Principle* which says that any nonempty collection of ideals in R has a minimal element (under set inclusion). (Cf. See the Note in Exercise (11) above. It applies here as well.)

16. If R is a domain, show that it is Artinian (cf. previous exercise) if and only if it is a field.
17. Let R be a Noetherian ring. Show if $\varphi : R \to S$ is a ring epimorphism, then S is Noetherian.
18. If R is a Noetherian ring and S a multiplicative set in R, then the localization $S^{-1}R$ is a Noetherian ring. (Cf. Exercise 29.4(7).)
19. Let R be a Noetherian domain. Show that any nontrivial ideal of R contains a finite product of nonzero prime ideals, i.e., if $0 < \mathfrak{A} < R$ is an ideal, then there exist nonzero prime ideals $\mathfrak{p}_1, ..., \mathfrak{p}_n$ in R such that $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_n \subset \mathfrak{A}$.
20. Let R be a Noetherian ring. Show that any ideal $\mathfrak{A} < R$ contains a finite product of prime ideals.
21. An ideal $\mathfrak{C}$ in a commutative ring R is called *irreducible* if whenever $\mathfrak{C} = \mathfrak{A} \cap \mathfrak{B}$ for some ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R, then either $\mathfrak{C} = \mathfrak{A}$ or $\mathfrak{C} = \mathfrak{B}$. Show if R is Noetherian, then every ideal $\mathfrak{A} < R$ is a finite intersection of irreducible ideals of R, i.e., $\mathfrak{A} = \mathfrak{C}_1 \cap \cdots \cap \mathfrak{C}_n$, for some irreducible ideals $\mathfrak{C}_i$ in R.
22. Let R be a Noetherian ring and $\mathfrak{A} < R$ an irreducible ideal (cf. Exercise 21 above). Show that $\mathfrak{A}$ is a primary ideal. (Cf. Exercise 28.19(15) for the definition of a primary ideal).
23. Let f, g be polynomials with coefficients in a commutative ring R. Suppose the leading coefficient of f is a unit (i.e., the coefficient of the highest degree term in f is a unit). Show that there are polynomials q and r with coefficients in R such that $g = fq + r$ with either $r = 0$ or the degree of r is less than the degree of f. This says the Division Algorithm holds when dividing by a polynomial with unit leading term.
24. Let F be a field. Show that $F[t]$, the ring of polynomials with coefficients in F, is a euclidean domain, hence a PID, hence a UFD.
25. Prove that a domain in which every prime ideal is principal is a PID.
26. Let $R = \mathbb{Z} + t\mathbb{Q}[t]$, i.e., polynomials with rational coefficients in which the constant is an integer. Prove that R is a Bézuit domain that is not Noetherian nor a UFD.
27. Suppose that R is a euclidean domain under the euclidean function ∂. Show that R is a euclidean domain under the strong euclidean function $\partial' : R \setminus \{0\} \to \mathbb{Z}^+ \cup \{0\}$ defined by $\partial'(x) = \min_{0 \neq r \in R} \partial(rx)$.
28. Let R be a euclidean domain under a strong euclidean function ∂, a, b nonzero elements in R, Show that b is a unit if and only if $\partial(b) = \partial(1)$ and if b is a nonunit, then $\partial(ab) > \partial(a)$.
29. Let R be a euclidean domain. Show directly, i.e., without using facts about PIDs or Noetherian domains, the following:
    (a) Every irreducible element in R is a prime element.
    (b) If R is a euclidean domain (under a strong euclidean function), then R is a UFD.

30. Let R be a UFD with quotient field K. Let S be a multiplicative set in R not containing zero. Show that the localization $S^{-1}R$ of R at S is a UFD. In particular, if $\mathfrak{p}$ is a prime ideal in R, then the localization $R_{\mathfrak{p}} := S^{-1}R$ of R at $S = R \setminus \mathfrak{p}$,
$$R_{\mathfrak{p}} := \{\frac{a}{b} \mid a, b \in R,\ b \notin \mathfrak{p}\} \subset K$$
is also a UFD.

31. Characterization of UFDs

We prove a generalization of our result that a PID is a UFD by giving Kaplansky's characterization of UFD's.

Theorem 31.1. (Kaplansky) *Let R be a domain. Then R is a UFD if and only if every nonzero prime ideal contains a prime element.*

Proof. ($\Rightarrow$): Let $0 < \mathfrak{p} < R$ be a prime ideal and a a nonzero element in $\mathfrak{p}$. Then $a = f_1 \cdots f_r$, with $f_1, \dots, f_r$ irreducible in R. Since $\mathfrak{p}$ is a prime ideal, there exist an i such that $f_i \in \mathfrak{p}$. As R is a UFD, f_i is a prime element.
($\Leftarrow$): Let
$$S = \{r \in R \mid r \neq 0 \text{ and } r \text{ is a unit or a product of prime elements}\}.$$
Clearly, $S \neq \emptyset$ is a multiplicative set and (0) excludes S. If we show $S = R \setminus \{0\}$, then R is a UFD by Euclid's Argument, i.e., we shall be done.
Claim. S is a *saturated multiplicative set*, i.e., if $a, b \in R$, then
$$a, b \in S \text{ if and only if } ab \in S:$$
As S is a multiplicative set, we need only show if $ab \in S$, then a and b lie in S. Note that $R^\times$ is a saturated multiplicative set, and by definition a subset of S, so we may assume that neither a nor b is a unit. Write $ab = p_1 \cdots p_r$ with $p_1, \dots, p_r$ primes hence each p_i lies in S. As $p_1 \mid ab$ and p_1 is a prime element, $p_1 \mid a$ or $p_1 \mid b$, say $p_1 \mid a$. Write $a = p_1 a_1$ with $a_1 \in R$, hence $p_1 a_1 b = p_1 p_2 \cdots p_r$ in the domain R. By cancellation, we conclude that $a_1 b = p_2 \cdots p_r$ in S. By induction on r, we have $a_1, b \in S$. As S is a multiplicative set, $a \in S$. This proves the claim.

As (0) excludes S, by Krull's Theorem, there exists a prime ideal $\mathfrak{p} \in R$ excluding S and maximal with respect to excluding S. If $0 < \mathfrak{p}$, then there exists a prime element $p \in \mathfrak{p}$, hence $p \in S \cap \mathfrak{p}$, a contradiction. Consequently, $(0) = \mathfrak{p}$ is maximal with respect to excluding S. In particular, if $a \in R$ is nonzero and satisfies $(0) < (a)$, we have $(a) \cap S \neq \emptyset$. We conclude that there exists an $r \in R$ such that $ar \in S$. As S is saturated by the claim, $a \in S$ as needed. □

Definition 31.2. Let R be a commutative ring. We call a prime ideal $\mathfrak{p}$ in R to be of *height n* if there exists a chain of prime ideals $\mathfrak{p}_0 < \cdots < \mathfrak{p}_n = \mathfrak{p}$ and none smaller.

Corollary 31.3. *Let R be a UFD. Then every nonzero prime in R is of height one.*

Proof. Every nonzero prime $\mathfrak{P}$ (if any) contains a nonzero prime ideal $\mathfrak{p}$ that is principal. In particular, it $\mathfrak{P}$ is prime, $\mathfrak{P} = \mathfrak{p}$. □

Remarks 31.4. 1. If R is a Noetherian ring, then, *a priori*, we do not need Zorn's Lemma to prove Krull's Theorem, hence in the proof above, if R is a Noetherian domain. However, as mentioned in Remark 30.12 this is really illusory.

2. We shall see that $\mathbb{C}[t_1, \dots, t_n]$ is a UFD, Using this one can show that the geometric meaning of Kaplansky's Theorem in the case of $\mathbb{C}^n$ is: Any hypersurface in $\mathbb{C}^n$ given by polynomial equations (in several variables) satisfying an irreducible condition can be defined by a single irreducible polynomial.

Using the argument of Kaplansky's Theorem 31.1, one shows:

Corollary 31.5. *If every prime ideal of a commutative ring R is principal, then every ideal in R is principal.*

This corollary was generalized by I. S. Cohen.

Theorem 31.6. (Cohen) *Let R be a commutative ring. Then R is Noetherian if and only if every prime ideal in R is finitely generated.*

We leave its proof as an exercise.

Exercises 31.7.

1. Prove Krull's Theorem 28.15 if R is a Noetherian ring and Kaplansky's Theorem 31.1 if R is a Noetherian domain without directly using Zorn's Lemma (cf. Remark 30.12).
2. Let $S \subset R$ be a multiplicative set such that $0 \notin S$. As above, we say that S is *saturated* if $ab \in S$ implies that $a, b \in S$. Prove that S is a saturated multiplicative set if and only if $R \setminus S$ is a union of prime ideals. In particular, the set of zero divisors in a commutative ring is a union of prime ideals.
3. Let R be a UFD and $\mathfrak{p}$ a nonzero prime in R. Then $\mathfrak{p}$ properly contains no nonzero prime ideal if and only if $\mathfrak{p}$ is principal and generated by a prime element. Moreover prime elements generate all such prime ideals in R.
4. Show 31.5.
5. Prove Cohen's Theorem 31.6.

32. Gaussian Integers

In this section, we study the Gaussian integers $\mathbb{Z}[\sqrt{-1}]$, a domain as it is a subring of $\mathbb{C}$. The map $^{-} : \mathbb{C} \to \mathbb{C}$ given by $x + y\sqrt{-1} \mapsto x - y\sqrt{-1}$, $x, y \in \mathbb{R}$, will denote *complex conjugation*. (As $\mathbb{C}$ is a real vector space on $\{1, \sqrt{-1}\}$, any element z in $\mathbb{C}$ can be written $z = x + y\sqrt{-1}$ for some unique real numbers x and y. This is a field automorphism that *fixes* the reals, i.e., is the identity on the reals. In fact, $\overline{z} = z$ if and only if $z \in \mathbb{R}$. Clearly, it restricts to a ring automorphism of the Gaussian integers. Let $N : \mathbb{C} \to \mathbb{R}^+ \cup \{0\}$ be the map defined by $z \mapsto z\overline{z}$. It is called the *norm map*.

Remarks 32.1. 1. If R is a ring, we let $\mathrm{Aut}_{ring}(R)$ denote the set of ring automorphisms of R. It is a group under composition. As $\overline{\overline{z}} = z$ for all $z \in \mathbb{C}$, complex conjugation has order two in $\mathrm{Aut}_{ring}(\mathbb{C})$. A ring automorphism of commutative rings of order one or two is called an *involution*. So complex conjugation is an involution on $\mathbb{C}$ and on $\mathbb{Z}[\sqrt{-1}]$.

2. Let $x, y \in \mathbb{R}$, then $N(x + y\sqrt{-1}) = x^2 + y^2$, which is positive unless $x = 0 = y$. This indicates that the norm map should be of interest when studying sums of two squares.
3. $N : (\mathbb{C}^\times, \cdot) \to (\mathbb{R}^\times, \cdot)$ is a monoid homomorphism, i.e., $N(1) = 1$ and it is *multiplicative*:
$$N(z_1 z_2) = N(z_1)N(z_2) \text{ for all } z_1, z_2 \in \mathbb{C}^\times.$$
In particular, if $x_1, x_2, y_1, y_2 \in \mathbb{C}$, we have the two square identity:
$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2,$$
the product of sums of two squares is a sum of two squares. Of course, once we have this identity, we see that this holds in any commutative ring by multiplying out.
4. $N(z) = N(\overline{z})$ for all $z \in \mathbb{C}$.

Lemma 32.2. *$\mathbb{Z}[\sqrt{-1}]$ is a domain and satisfies the following:*

(1) $N|_{Z[\sqrt{-1}]} : \mathbb{Z}[\sqrt{-1}] \to \mathbb{Z}^+ \cup \{0\}$ and satisfies $N|_{\mathbb{Z}[\sqrt{-1}]}(\alpha) = 0$ if and only if $\alpha = 0$. We shall write N for $N|_{\mathbb{Z}[\sqrt{-1}]}$.

(2) $\mathbb{Z}[\sqrt{-1}]^\times = \{\pm 1, \pm\sqrt{-1}\}$.

(3) Let $a, b \in \mathbb{Z}$. Then $a + b\sqrt{-1}$, is a root of the monic polynomial (i.e., leading coefficient is one), $t^2 - 2at + (a^2 + b^2) \in \mathbb{Z}[t]$.

Proof. We only prove (2) leaving the rest as an easy check. If $uv = 1$ with $u, v \in \mathbb{Z}[t]$, then $1 = N(1) = N(u)N(v)$, so we have $N(u) \in \mathbb{Z}^\times \cap \mathbb{Z}^+ = \{1\}$. As $x^2 + y^2 = 1$ with $x, y \in \mathbb{Z}$ if and only if $x = \pm 1$, $y = 0$ or $x = 0, y = \pm 1$, we conclude that
$$\mathbb{Z}[\sqrt{-1}]^\times \subset \{\beta \mid N(\beta) = 1\} \subset \{\pm 1, \pm\sqrt{-1}\}.$$
Clearly, $\{\pm 1, \pm\sqrt{-1}\} \subset \mathbb{Z}[\sqrt{-1}]^\times$, establishing the lemma. □

Theorem 32.3. *The domain $\mathbb{Z}[\sqrt{-1}]$ is a euclidean domain, hence a PID and so a UFD.*

Proof. **Claim**. N is a normed euclidean function on $\mathbb{Z}[\sqrt{-1}]$. (Consequently, $\mathbb{Z}[\sqrt{-1}]$ is a strong euclidean domain under N.):
Let α and β be elements in $\mathbb{Z}[\sqrt{-1}]$ with β nonzero. We must show that there exist a γ and ρ in $\mathbb{Z}[\sqrt{-1}]$ satisfying $\alpha = \beta\gamma + \rho$ with $\rho = 0$ or $N(\rho) < N(\beta)$.
In $\mathbb{C}$, we can write
$$\frac{\alpha}{\beta} = \frac{\alpha}{\beta}\frac{\overline{\beta}}{\overline{\beta}} = \frac{\alpha\overline{\beta}}{N(\beta)} = r + s\sqrt{-1}$$
for some $r, s \in \mathbb{Q}$ (as $N(\beta) \in \mathbb{Z}$). Choose $m, n \in \mathbb{Z}$ satisfying
$$|r - m| \le 1/2 \text{ and } |s - n| \le 1/2.$$
Set $\gamma = m + n\sqrt{-1}$ and $\rho = \alpha - \beta\gamma$ in $\mathbb{Z}[\sqrt{-1}]$. If $\rho = 0$, we are done, so suppose not. As $N : \mathbb{C}^\times \to \mathbb{R}^+$ and is multiplicative, we have
$$\begin{aligned} N(\rho) &= N(\alpha - \beta\gamma) = N\big(\beta(\frac{\alpha}{\beta} - \gamma)\big) = N(\beta)N(\frac{\alpha}{\beta} - \gamma) \\ &= N(\beta)N\big((r - m) + (s - n)\sqrt{-1}\big) = N(\beta)[(r - m)^2 + (s - n)^2] \\ &\le N(\beta)(\frac{1}{4} + \frac{1}{4}) < N(\beta). \end{aligned}$$
□

Note in the last line of the proof, our estimate on $N(\rho)$ was $\leq \frac{1}{2}N(\beta)$, so we had some room to work. This explains why the domains in Remark 30.21(5) work — compute the restriction of N to each of these domains.

As mentioned before, we can use the Gaussian integers to study sums of squares of integers. In particular, $m = x^2 + y^2$ with x, y integers if and only if $m = N(x + y\sqrt{-1})$. We wish to determine all the non-negative integers that are sums of two squares. We have seen that a product of two integers each a sum of two squares of integers is itself a sum of two squares. By the Fundamental Theorem of Arithmetic, each non-negative integer factors as $p_1^{e_1} \cdots p_r^{e_r}$ with the p_i distinct positive primes and the $e_i \geq 1$. If e_i is even then $p_i^{e_i}$ is a square (so a sum of two squares). Hence we are reduced to determining which products of distinct primes are sums of two squares of integers. In particular, we must determine which primes are sums of two squares of integers. As 2 is a sum of two squares, we are reduced to odd primes. If $x \in \mathbb{Z}$, then $x^2 \equiv 0, 1 \mod 4$. In particular, a sum of two integer squares can only be congruent to $0, 1$ or 2 modulo four. In particular, no integer n satisfying $n \equiv 3 \mod 4$ can be a sum of two squares of integers, e.g., 3 and 7 are not sums of two squares.

The key to the solution is the following:

Lemma 32.4. *Let p be a positive prime in $\mathbb{Z}$ satisfying $p \equiv 1 \mod 4$. Then there exists an integer x satisfying $p \mid x^2 + 1$ in $\mathbb{Z}$. In particular, -1 is a square modulo p if $p \equiv 1 \mod 4$.*

PROOF. Let $x = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot \frac{p-1}{2}$ in $\mathbb{Z}$. By Wilson's Theorem (Exercise 10.16(14)), we have

$$\begin{aligned}
-1 \equiv (p-1)! &= (1 \cdot 2 \cdot 3 \cdots \frac{p-1}{2})(p - \frac{p-1}{2} \cdots p-3 \cdot p-2 \cdot p-1) \\
&\equiv (1 \cdot 2 \cdot 3 \cdots \frac{p-1}{2})(-\frac{p-1}{2} \cdots -3 \cdot -2 \cdot -1) = (-1)^{\frac{p-1}{2}} x^2 \\
&\equiv x^2 \mod p,
\end{aligned}$$

so $p \mid x^2 + 1$. □

Theorem 32.5. (Fermat) *Let p be a positive prime congruent to 1 modulo 4. Then there exist integers x and y such that $p = x^2 + y^2$.*

PROOF. By the lemma there exists an integer x such that $p \mid x^2 + 1$. Consequently, $p \mid (x + \sqrt{-1})(x - \sqrt{-1})$ in the UFD $\mathbb{Z}[\sqrt{-1}]$.

Claim. p is not irreducible (i.e., it is *reducible*) in $\mathbb{Z}[\sqrt{-1}]$:

Suppose that p is irreducible in $\mathbb{Z}[\sqrt{-1}]$. Then as $\mathbb{Z}[\sqrt{-1}]$ is a UFD, it is a prime. Hence $p \mid x + \sqrt{-1}$ or $p \mid x - \sqrt{-1}$. If $p \mid x + \sqrt{-1}$, then $p(a + b\sqrt{-1}) = x + \sqrt{-1}$ for some integers a and b. It follows that $pb = 1$ in $\mathbb{Z}$ (why?), which is impossible. [Or if $p \mid x + \sqrt{-1}$, then $p = \overline{p} \mid x - \sqrt{-1}$ also. It follows that $p \mid 2x$ hence $p \mid x$ as p is odd.] Similarly, $p \nmid x - \sqrt{-1}$, so p is not irreducible in $\mathbb{Z}[\sqrt{-1}]$, proving the claim.

Using the claim, we can write $p = \pi\alpha$ for some $\pi, \alpha \in \mathbb{Z}[\sqrt{-1}]$, with π irreducible in $\mathbb{Z}[\sqrt{-1}]$ and α not a unit in $\mathbb{Z}[\sqrt{-1}]$. Taking the norm of this equation yields

$$p^2 = N(p) = N(\pi)N(\alpha)$$

in $\mathbb{Z}$. As $N(\pi)$ and $N(\alpha)$ lie in $\mathbb{Z}^+$ with $\mathbb{Z}$ a UFD, we must have

$$\pi\overline{\pi} = N(\pi) = p = N(\alpha)$$

is a sum of two squares, proving the theorem. In addition, we have also shown, in view of this last equation, that $\overline{\pi} = \alpha$ is also irreducible and $p = \pi\overline{\pi}$ is a factorization of p in $\mathbb{Z}[\sqrt{-1}]$. $\square$

To categorize those integers that are sums of two squares, we need the following lemma:

Lemma 32.6. *Let a and b be two nonzero relatively prime integers and p a (positive) odd prime such that $p \mid a^2 + b^2$. Then $p \equiv 1 \mod 4$.*

Proof. As $p \mid a^2 + b^2$, we have $p \mid a$ if and only if $p \mid b$. Since a and b are relatively prime, p cannot divide a or b. By Fermat's Little Theorem 10.14, we know that $b^{p-3}b^2 = b^{p-1} \equiv 1 \mod p$. Consequently, $a^2 + b^2 \equiv 0 \mod p$, i.e., $a^2 \equiv -b^2 \mod p$, and $p-3$ even imply that

$$(b^{\frac{p-3}{2}}a)^2 = b^{p-3}a^2 \equiv -b^{p-3}b^2 \equiv -1 \mod p.$$

Let $x = b^{\frac{p-3}{2}}a$. Then $x^2 \equiv -1 \mod p$ which means that

$$1 \equiv x^{p-1} = (x^2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} \mod p.$$

If $p \equiv 3 \mod 4$, then we conclude that $1 \equiv -1 \mod p$, which is impossible. $\square$

The next result solves our stated problem. We leave its proof as an exercise (that you should do).

Theorem 32.7. *Let $n = p_1^{e_1} \cdots p_r^{e_r}$ in $\mathbb{Z}$ with $0 < p_1 < \cdots < p_r$ primes, and $e_i > 0$ for $i = 1, \dots, r$. Then n is a sum of two squares in $\mathbb{Z}$ if and only if e_i are even integers whenever $p_i \equiv 3 \mod 4$ for $i = 1, \dots, r$.*

Let p be a (positive) odd prime. By Lemma 32.6, we know that

$$-1 \text{ is } \begin{cases} \text{a square mod } p & \text{if } p \equiv 1 \mod 4. \\ \text{not a square mod } p & \text{if } p \equiv 3 \mod 4. \end{cases}$$

Definition 32.8. If a is an integer not divisible by p, define the *Legendre symbol*

$$\left(\frac{a}{p}\right) := \begin{cases} +1 \text{ if } a \text{ mod } p \text{ is a square.} \\ -1 \text{ if } a \text{ mod } p \text{ is not a square.} \end{cases}$$

It is also convenient to define $\left(\frac{b}{p}\right) = 0$ if $p \mid b$.

Example 32.9. $\left(\frac{0}{7}\right) = 0$, $\left(\frac{1}{7}\right) = \left(\frac{2}{7}\right) = \left(\frac{4}{7}\right) = 1$, and $\left(\frac{3}{7}\right) = \left(\frac{5}{7}\right) = \left(\frac{6}{7}\right) = -1$

We have proven:

Proposition 32.10. (Euler's Formula) *Let p be a (positive) odd prime. Then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} +1 \ \textit{if } p \equiv 1 \mod 4. \\ -1 \ \textit{if } p \equiv 3 \mod 4. \end{cases}$$

Corollary 32.11. *The set $\{4n + 1 \mid n \in \mathbb{Z}\}$ contains infinitely many primes.*

PROOF. We prove the stronger result that the set $\{8n+5 \mid n \in \mathbb{Z}\}$ contains infinitely many primes.

Check. If $n \in \mathbb{Z}$, then $n^2 \equiv 0, 1, 4 \mod 8$. In particular, if n is odd, then $n^2 \equiv 1 \mod 8$. Let p be an odd prime and set $N = (3 \cdot 5 \cdot 7 \cdots p)^2 + 2^2 \equiv 5 \mod 8$. If q is an odd prime such that $q \mid N$, then by Lemma 32.6, we have $q \equiv 1 \mod 4$, hence $q \equiv 1 \mod 8$ or $q \equiv 5 \mod 8$. If N is a product of primes all congruent to 1 modulo 8, then $N \equiv 1 \mod 8$, a contradiction. Therefore, there exists a prime q satisfying $q \equiv 5 \mod 8$ and $q \mid N$. As none of the primes $3, 5, 7, \dots, p$ divides N, we must have $q > p$, so there exists a prime larger that p congruent to 5 modulo 8. $\square$

Remark 32.12. It is still an open question whether the set of integers $\{x^2 + 1 \mid x \in \mathbb{Z}\}$ contains infinitely many primes.

The following is a well-known theorem, although we shall not prove it. The usual proof uses complex analysis, although there is now an elementary (but tricky) proof that does not.

Theorem 32.13. (Dirichlet's Theorem on Primes in an Arithmetic Progression) *Let a and b be two nonzero relatively prime integers. Then there exist infinitely many primes p satisfying $p \equiv a \mod b$.*

We shall prove that this is true for $a = 1$ in Proposition 59.10 below by elementary means. Other questions arise about *arithmetic progressions*, i.e., equally spaced integers, e.g., $a, a+b, a+2b, a+3b, \dots$ with a and b integers. The distance between two consecutive integers in an arithmetic progression is called the *spacing* of the progression, e.g., b is the spacing in the example. Such a progression can be finite or infinite. The most spectacular theorem is the following proven in 2004:

Theorem 32.14. *(*Green-Tao) *Let N be a positive integer. Then there exists an arithmetic progression consisting of N primes.*

The theorem shows that for any N there exists infinitely many arithmetic progressions consisting of N primes, but it does not indicate anything about the spacing. Turning this theorem around, one can ask does there exist infinitely many primes in an arithmetic progression with a given spacing. The most famous example of this is

Question 32.15. *(*Twin Prime Conjecture) *Do there exist infinitely many twin prime, i.e., primes pairs of the form $p, p+2$.*

Euler had shown that $\sum\limits_{p \text{ a prime}} \dfrac{1}{p}$ was infinite, so there was some hope that $\sum\limits_{p \text{ a twin prime}} \dfrac{1}{p}$ was also. Brun showed this to be false in 1915.

We shall also study the Legendre symbol further proving the following wonderful (albeit seemingly strange) theorem of Gauss upon which you should meditate (and we shall prove it in Theorem 59.19 below):

Theorem 32.16. (Quadratic Reciprocity) *Let p and q be distinct odd primes. Then*

$$\Big(\frac{p}{q}\Big)\Big(\frac{q}{p}\Big) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Exercises 32.17.

1. Show that the *evaluation map* $e_{\sqrt{-1}} : \mathbb{Z}[t] \to \mathbb{C}$ defined by $f \mapsto f(\sqrt{-1})$ is a ring homomorphism with kernel (t^2+1) and image the Gaussian integers. [Hint: Use Exercise 30.22(23).]
2. Let $R = \mathbb{Z}[\sqrt{-1}]$ and $n = p_1^{e_1} \cdots p_r^{e_r}$ be the standard factorization of the integer $n > 1$. Show that the following are equivalent:
   (*i*) n is a sum of two squares.
   (*ii*) $n = N(\alpha)$ for some $\alpha \in R$.
   (*iii*) If $p_i \equiv 3 \pmod 4$, then e_i is even.
   [Hint: Use Lemma 32.6.]
3. Show the following
   (*i*) Let α be an element in $\mathbb{Z}[\sqrt{-1}]$ such that $N(\alpha)$ is a prime or the square of a prime in $\mathbb{Z}$. Then α is a prime element or a product of two prime elements in $\mathbb{Z}[\sqrt{-1}]$.
   (*ii*) Let π be an prime element in $\mathbb{Z}[\sqrt{-1}]$. Show $N(\pi)$ is a prime or the square of a prime in $\mathbb{Z}$.
4. Determine all prime elements, up to units, in $\mathbb{Z}[\sqrt{-1}]$.
5. Show that $\mathbb{Z}[\sqrt{-2}]$ is a euclidean domain.
6. Prove that $\mathbb{Z}\left[\dfrac{-1-\sqrt{-3}}{2}\right]$ and $\mathbb{Z}[\sqrt{3}\,]$ are euclidean domains.
7. Prove that $\mathbb{Z}\left[\dfrac{-1-\sqrt{-d}}{2}\right]$ is a euclidean domain for $d = 7$ and $d = 11$.
8. Let $R = \mathbb{Z}[\sqrt{-d}] = \{a + b\sqrt{-d} \mid a, b \in \mathbb{Z}\}$, a subring of $\mathbb{C}$ with d a positive square-free integer. Let $N : R \to \mathbb{Z}$ be the norm map, so $\alpha = a + b\sqrt{-d} \mapsto \alpha\overline{\alpha} = a^2 + db^2$. Show all of the following:
   (*i*) The field of quotients of R is $\mathbb{Q}[\sqrt{-d}] = \{a + b\sqrt{-d} \mid a, b \in \mathbb{Q}\}$.
   (*ii*) $N : R \setminus \{0\} \to \mathbb{Z}$ is a monoid homomorphism.
   (*iii*) $R^\times = \{\alpha \in R \mid N(\alpha) = 1\}$ and compute this group for all d.
   (*iv*) The element α is irreducible in R if $N(\alpha)$ is a prime. Is the converse true?
   (*v*) Suppose $d \geq 3$, then 2 is irreducible but not prime in R.
9. Let $R = \mathbb{Z}[\sqrt{-5}]$. Show the following:
   (*i*) The elements $2, 3, 1+\sqrt{-5}$, and $1-\sqrt{-5}$ are all irreducible but no two are associates.
   (*ii*) None of the elements $2, 3, 1+\sqrt{-5}$, and $1-\sqrt{-5}$ are prime. In particular, R is not a UFD.
10. Let $R = \mathbb{Z}[\sqrt{-5}]$. Let $\mathfrak{P} = (2, 1+\sqrt{-5})$. Show
    (*i*) $\mathfrak{P}^2 = (2)$ in R.
    (*ii*) $\mathfrak{P}$ is a maximal ideal.
    (*iii*) $\mathfrak{P}$ is not a principal ideal.

33. Addendum: The Four Square Theorem

In the previous section, we characterized those positive elements that are sums of two squares. In this section, we prove Lagrange's theorem that every positive integer is a sum of four squares. The proof is based on a computation of Euler that shows a product of two integers each a sum four integer squares is a sum of four integer squares. Although he did not know it at the time, this formula comes from the Hamiltonian quaternions, the first known division ring that was not commutative. We begin by constructing it.

Construction 33.1. Let $\mathcal{H}$ be a four dimensional real vector space with basis $\{1, i, j, k\}$. We view 1 as $1_{\mathbb{R}}$, so $\mathbb{R} \subset \mathcal{H}$. We make $\mathcal{H}$ into a ring by defining a multiplication on this basis and extend linearly to the whole space. The multiplication of the basis elements is given by:

$$i^2 = -1, \;\; j^2 = -1, \;\; ij = -ji = k.$$

Extending this linearly give a multiplication on $\mathcal{H}$ as follows:

$$\begin{aligned}(x_0 1 + x_1 i + x_2 j + x_3 k) &\cdot (y_0 1 + y_1 i + y_2 j + y_3 k)\\ &= (x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3)1 + (x_0y_1 + x_1y_0 + x_2y_3 - x_3y_2)i\\ &+ (x_0y_2 + x_2y_0 + x_3y_1 - x_1y_3)j + (x_0y_3 + x_3y_0 + x_1y_2 - x_2y_1)k.\end{aligned}$$

By a straight-forward (but arduous) computation, one now checks that this makes $\mathcal{H}$ into a ring with $1_{\mathcal{H}} = 1_{\mathbb{R}}$. Note that the subring generated by $\{1, i\}$ is isomorphic to $\mathbb{C}$ (as are the subrings generated by $\{1, j\}$ and $\{1, k\}$, respectively). The ring $\mathcal{H}$ is not commutative as its *center*, i.e., the subring $\{x \in \mathcal{H} \mid xy = yx \text{ for all } y \text{ in } \mathcal{H}\}$ of $\mathcal{H}$, is precisely $\mathbb{R}$. [Why is the center a subring?] Note also that $\mathcal{H}^\times$ contains the *quaternion group* $\{1, i, j, ij, -1, -i, -j, -ji\}$.

Analogous to $\mathbb{C}$, there is a *quaternion conjugation* map:

$$^- : \mathcal{H} \to \mathcal{H} \text{ given by } \overline{x_0 1 + x_1 i + x_2 j + x_3 k} := x_0 1 - x_1 i - x_2 j - x_3 k.$$

It is easy to check that for all x, y in $\mathcal{H}$, we have

$$\begin{aligned}\overline{1} &= 1\\ \overline{x+y} &= \overline{x} + \overline{y}\\ \overline{xy} &= \overline{y}\;\overline{x}.\end{aligned}$$

Therefore, the quaternion conjugation map satisfies the properties of a ring homomorphism except that it reverses multiplication. Such a map of rings is called a *ring antihomomorphism*. Clearly, this map is also a surjective linear transformation of a real vector space. Hence it is bijective, since $\mathcal{H}$ is a finite dimensional vector space over $\mathbb{R}$. Consequently, the map is, in fact, a *ring antiautomorphism*. (Note the composition of two ring antihomomorphisms is a ring homomorphism.) In addition, for all x in $\mathcal{H}$, we have

$$\overline{\overline{x}} = x,$$

so $^-$ is an *involution*, i.e., an antiautomorphism satisfying the composition $^- \circ {}^-$ is the identity. (Cf. with complex conjugation.) We also have a *norm map*:

$$N : \mathcal{H} \to \mathbb{R}^+ \cup \{0\} \text{ given by } z \mapsto z\overline{z},$$

that is checked to be multiplicative, i.e., $N(xy) = N(x)N(y)$ for all x, y in $\mathcal{H}$. [Note that if we identify $\mathbb{C}$ as the subring of $\mathcal{H}$ generated by $\{1, i\}$, then $N|_{\mathbb{C}}$ is the usual norm map on $\mathbb{C}$.] This is useful as the norm of $x_0 1 + x_1 i + x_2 j + x_3 k$ is

$$N(x_0 1 + x_1 i + x_2 j + x_3 k) = x_0^2 + x_1^2 + x_2^2 + x_3^2,$$

a sum of four squares. The multiplicativity of this norm map yields *Euler's Equation*: the four square identity

$$\begin{aligned}
&(x_0^2 + x_1^2 + x_2^2 + x_3^2) \cdot (y_0^2 + y_1^2 + y_2^2 + y_3^2) \\
&\quad = (x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3)^2 + (-x_0y_1 + x_1y_0 - x_2y_3 + x_3y_2)^2 \\
&\quad + (-x_0y_2 + x_2y_0 + x_1y_3 - x_3y_1)^2 + (-x_0y_3 + x_3y_0 - x_1y_2 + x_2y_1)^2,
\end{aligned}$$

the formula that a product of two sums of fours squares is a sum of four squares, generalizing the norm on the complex numbers giving a formula for the product of sums of two squares. This is the equation that we shall need to prove Lagrange's Theorem that any positive integer is a sum of four squares. Of course, we do not need the quaternions to establish Euler's Equation — we need only multiply the right hand side out to see that it holds in any commutative ring. Indeed Euler found this equation without knowing about quaternions. It does, however, explain why such an equation exists. It can also be shown that there is a formula that a product of two sums of eight squares is a sum of eight squares. It arises from a generalization of the quaternions called the *octonians.*, a algebraic structure that looks a division ring but does not satisfy the associative law for multiplication. So the two square identity arises from fields, the four square identity from the quaternions (tossed out commutativity). the eight square identity from the octonians (toss out commutativity and associativity). Is there a sixteen square identity arising from a generalization of these algebraic structures by throwing out some other property of rings? Hurwitz showed not.

As with the norm map on $\mathbb{C}$, it follows that $N(z) = 0$ if and only if $z = 0$. In particular, if z is nonzero, then $z\overline{z}/N(z) = 1 = \overline{z}z/N(z)$, so z has a multiplicative inverse. Therefore, $\mathcal{H}$ is a division ring called the *Hamiltonian quaternions*. By definition, $ij \neq ji$, so $\mathcal{H}$ is not a field.

Remark 33.2. Let V be a finite dimensional real vector space that is also a division ring and contains $\mathbb{R}$ in its center. Then it can be shown that V is ring isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathcal{H}$. In particular, $\dim_{\mathbb{R}} V = 1, 2$, or 4. We shall show this in Section 104.

To show that every positive integer is a sum of four squares, we shall use the following:

Lemma 33.3. *$1 \leq m < p$ and mp is a sum of four integer squares.*

Proof. Let

$$S_1 := \{0^2, 1^2, \ldots, \frac{(p-1)}{2}^2\}$$

$$S_2 := \{-0^2 - 1, -1^2 - 1, \ldots, -\frac{(p-1)}{2}^2 - 1\}.$$

If $x^2 = y^2 \mod p$ with x, y integers, then $p \mid x^2 - y^2 = (x + y)(x - y)$. Consequently, $p \mid x + y$ or $p \mid x - y$. It follows (check) that no two elements in S_1 are congruent modulo

p and no two elements in S_2 are congruent modulo p. Since

$$|S_1| + |S_2| = \Big(\frac{p-1}{2} + 1\Big) + \Big(\frac{p-1}{2} + 1\Big) = p + 1,$$

there exist $x^2 \in S_1$ and $-y^2 - 1 \in S_2$ satisfying $0 \neq x^2 + y^2 + 1$ and

$$x^2 + y^2 + 1 \equiv 0 \mod p.$$

Hence there exists a positive integer m such that

$$mp = x^2 + y^2 + 1 \leq \Big(\frac{p-1}{2}\Big)^2 + \Big(\frac{p-1}{2}\Big)^2 + 1 < \frac{p^2}{2} + 1 < p^2.$$

So $1 \leq m < p$. □

Remarks 33.4. 1. Of course, the proof shows that the element mp is a sum of three squares.

2. The same proof shows that -1 is a sum of two squares in $\mathbb{Z}/p\mathbb{Z}$ for any prime p. In fact, any element a in $\mathbb{Z}/p\mathbb{Z}$ is a sum of two square — replace the -1 in the proof by a in S_1.
3. A similar (but not obvious) proof works to show that any element in a finite field is a sum of two squares.
[Note that if $\operatorname{char}(F) = 2$, then the map $F \to F$ given by $x \mapsto x^2$ is a (field) monomorphism, so an automorphism if F is a finite field.]

Historically, one of the main forms of induction used to prove rather deep theorems was the idea of Fermat, called *Fermat descent*. We prove our theorem about four squares using this method.

Theorem 33.5. (Lagrange). *Every positive integer is a sum of four squares of integers.*

Proof. We know that $2 = 1+1+0+0$ is a sum of four squares, so by Euler's Formula and the Fundamental Theorem of Arithmetic it suffices to show every positive odd prime is a sum of four squares. [We even know this is true for primes congruent to 1 modulo 4, but will not use this.] Let p be an odd prime. By the lemma and the Well-ordering Principle there exists an element $(x_0, y_0, z_0, w_0, m) \in \mathbb{Z}^5$ with $1 \leq m < p$ and satisfying

$$(*) \qquad x_0^2 + y_0^2 + z_0^2 + w_0^2 = mp$$

with m minimial. To finish, it suffices to show that $m = 1$.

Case 1. m is even:

By (*), an even number of the integers x_0, y_0, z_0, w_0 must be odd (if any are odd). If precisely two of these are odd, we may assume that they are x_0, y_0. So we have $x_0 \equiv y_0 \mod 2$ and $z_0 \equiv w_0 \mod 2$, hence

$$(\frac{x_0 + y_0}{2})^2 + (\frac{x_0 - y_2}{2})^2 + (\frac{z_0 + w_0}{2})^2 + (\frac{z_0 - w_0}{2})^2 = \frac{m}{2}p,$$

contradiction the minimality of m.

Case 2. m is odd:

We use the modification of the Division Algorithm given by Exercise 4.24(3). Using this modified division algorithm, we have equations:

$$x_0 = mx_1 + x_2 \qquad \text{with } |x_2| < \frac{m}{2}.$$
$$y_0 = my_1 + y_2 \qquad \text{with } |y_2| < \frac{m}{2}.$$
$$z_0 = mz_1 + z_2 \qquad \text{with } |z_2| < \frac{m}{2}.$$
$$w_0 = mw_1 + w_2 \qquad \text{with } |w_2| < \frac{m}{2}.$$

Therefore, we have

$$x_2^2 + y_2^2 + z_2^2 + w_2^2 \equiv x_0^2 + y_0^2 + z_0^2 + w_0^2 \equiv 0 \mod m,$$

so

$$(\dagger) \qquad x_2^2 + y_2^2 + z_2^2 + w_2^2 = Mm$$

for some integer M.

Subcase 1. $M = 0$:

In this subcase, we must have $x_2 = y_2 = z_2 = w_2 = 0$ and $m \mid x_0$, $m \mid y_0$, $m \mid z_0$, $m \mid w_0$, so $m^2 \mid x_0^2 + y_0^2 + z_0^2 + w_0^2 = mp$. Hence $m \mid p$ with $1 \leq m < p$. If follows that $m = 1$, so we are done in this subcase.

Subcase 2. $M > 0$:

We have $(x_1^2 + y_1^2 + z_1^2 + w_1^2) \cdot (x_2^2 + y_2^2 + z_2^2 + w_2^2) = A^2 + B^2 + C^2 + D^2$ by Euler's Equation with $A = x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2$ and for some integers B, C, D. Consequently,

$$\begin{aligned} mp &= x_0^2 + y_0^2 + z_0^2 + w_0^2 \\ &= (mx_1 + x_2)^2 + (my_1 + y_2)^2 + (mz_1 + z_2)^2 + (mw_1 - w_2)^2 \\ &= m^2(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2mA + mM. \end{aligned}$$

Multiply this equation by M/m and use ($\dagger$) to obtain the following equation in $\mathbb{Z}$:

$$\begin{aligned} Mp &= mM(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2MA + M^2 \\ &= (x_2^2 + y_2^2 + z_2^2 + w_2^2)(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2MA + M^2 \\ &= A^2 + B^2 + C^2 + D^2 + 2MA + M^2 = (A + M)^2 + B^2 + C^2 + D^2 \end{aligned}$$

is a sum of four squares with

$$0 < mM = x_2^2 + y_2^2 + z_2^2 + w_2^2 < 4(\frac{m}{2})^2 = m^2.$$

Therefore, $0 < M < m$, contradicting the choice of m. $\square$

Remark 33.6. It is much harder to determine which positive integers are sums of three squares. This was determined by Gauss, who showed that a positive integer n is a sum of three squares if and only if $n \neq 4^m(8k + 7)$ for some non-negative integers m and k. [Note that $3 = 1 + 1 + 1$ and $13 = 2^2 + 3^2$ but $39 = 3 \cdot 13$ is a sum of four squares (as $39 = 1^2 + 2^2 + 3^2 + 5^2$), but not fewer.]

Hilbert proved a much more general theorem when he solved the *Waring Problem* that given any positive integer n there exists a minimal positive integer $g(n)$ such that every positive integer is a sum of $g(n)$ nth powers. His proof is not constructive, and it is a difficult problem (and open for most n) to determine $g(n)$.

Exercises 33.7.

1. Show if an integer n satisfies $n = 4^m(8k+7)$, then n is not a sum of three squares. [The converse is the hard part.]
2. Show $\mathcal{H}$ is a ring and prove quaternion conjugation is multiplicative.
3. Let $z = x_0 1 + x_1 i + x_2 j + x_3 k \in \mathcal{H}$ be a quaternion, $x_i \in \mathbb{R}$. We call z a *pure quaternion* if $x_0 = 0$. Suppose z is nonzero. Show that z is a pure quaternion if and only if z does not lie in $\mathbb{R}$ but z^2 does lie in $\mathbb{R}$. In particular, if z is a pure quaternion, then $z^2 = -x_1^2 - x_2^2 - x_3^2$. [Note that if z is a pure quaternion, then $\overline{z} = -z$.]
4. Show that $t^2 + 1 \in \mathbb{R}[t]$ has infinitely many roots in $\mathcal{H}$.
5. Let $U_0 := \{x \in \mathcal{H} \mid N(z) = 1\}$ and $\mathrm{SU}_2(\mathbb{C}) := \{A \in \mathrm{GL}_2(\mathbb{C}) \mid AA^* = I,\ \det A = 1\} = \{\begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) \mid \alpha, \beta \in \mathbb{C} \text{ with } \alpha\overline{\alpha} + \beta\overline{\beta} = 1\}$, the special unitary group. Show that the map
$$i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad j \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad k \mapsto \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$$
induces a group isomorphism $U_0 \to \mathrm{SU}_2(\mathbb{C})$. [We also have $\mathrm{SU}_2(\mathbb{C})/\{\pm I\} \cong \mathrm{SO}_3(\mathbb{R})$, the special orthogonal group.]
6. Generalize the construction of quaternions as follows: Let F be any field of characteristic different from two and a and b nonzero elements of F. Let $\left(\frac{a,b}{F}\right)$ be the 4-dimensional F-vector space on basis $\{1, i, j, k\}$. Show that $\left(\frac{a,b}{F}\right)$ becomes a ring by defining
$$i^2 = a, \;\; j^2 = b, \;\; ij = -ji = k$$
and extend linearly with all elements in F commuting with all elements in $\left(\frac{a,b}{F}\right)$. Such a ring with F commuting with all elements is called an *F-algebra*, and the F-algebra $\left(\frac{a,b}{F}\right)$ is called a *general quaternion algebra*. Define the *conjugate* of $z = x_0 1 + x_1 i + x_2 j + x_3 k$ in $\left(\frac{a,b}{F}\right)$, $x_0 . x_1, x_2, x_3 \in F$, by
$$\overline{x_0 1 + x_1 i + x_2 j + x_3 k} := x_0 1 - x_1 i - x_2 j - x_3 k$$
and the norm map
$$N : \left(\frac{a,b}{F}\right) \to F \;\text{ by }\; z \mapsto z\overline{z}.$$
Analogous to the Hamiltonian quaternions, show conjugation is an anti-automorphism and N is multiplicative. Moreover, show
$$N(x_0 1 + x_1 i + x_2 j + x_3 k) = x_0^2 - a x_1^2 - b x_2^2 + ab x_3^2.$$

In particular, $\left(\frac{a,b}{F}\right)$ is a division ring if it satisfies the property that $N(z) = 0$ if and only if $z = 0$. [If this does not happen, then $\left(\frac{a,b}{F}\right) \cong \mathbb{M}_2(F)$.] Can you give non-isomorphic examples, of quaternion algebras that are division rings over $\mathbb{Q}$?

7. Let F be a field of characteristic different from two. Show that
$$\left(\frac{a, b}{F}\right) \cong \left(\frac{ax^2, by^2}{F}\right)$$
are isomorphic as F-algebras, i.e., a ring isomorphism that fixes F. In particular, $\mathcal{H} \cong \left(\frac{-1,-1}{\mathbb{R}}\right)$.
8. Show that $\left(\frac{1,-1}{F}\right) \cong \mathbb{M}_2(F)$. In particular, the only quaternion algebras over $\mathbb{C}$ is $\mathbb{M}_2(\mathbb{C})$.

CHAPTER VII

Polynomial Rings

In this chapter we study polynomial rings over rings, concentrating on the case of a polynomial ring in one variable over a domain, e.g., a field. Most importantly, we prove Kronecker's Theorem that a non-constant polynomial over a field F (in one variable) has a root in a field containing F. This will be a basic result when we turn to studying Field Theory. We then investigate polynomial rings over a UFD. We prove that any such ring is itself a UFD using a classical result of Gauss. In the final section of this chapter, we attempt to motivate the commutative algebra needed for algebraic geometry that will be studied more systematically in a later chapter, that is the study of zeros of polynomials in many variables over a field, especially algebraically closed fields.

34. Introduction to Polynomial Rings

We start by explicitly defining polynomial rings. If R is a ring, let $R[t] := \{a_0 + a_1t + \cdots + a_nt^n \mid a_i \in R,\ i = 1, \ldots n, \text{ some } n \geq 0\}$ with $a_0 + a_1t + \cdots + a_nt^n = 0$ if and only if $a_i = 0$ for $i = 0, \ldots, n$. Therefore, if

$$f = a_0 + a_1t + \cdots + a_nt^n \text{ and } g = b_0 + b_1t + \cdots + b_mt^m,$$

then

$$f = g \text{ if and only if } a_i = b_i \text{ for all } i$$

(where we let $a_i = 0$ for all $i > n$ and $b_j = 0$ for all $j > m$).
With f and g as above, define

$$f + g := (a_0 + b_0) + (a_1 + b_1)t + \cdots = \sum (a_i + b_i)t^i$$

and

$$fg := \sum c_it^i \text{ with } c_i = \sum_{k=0}^{i} a_{i-k}b_k.$$

This makes $R[t]$ into a ring called the *polynomial ring* over R with $0_{R[t]} = 0_R + 0t + \cdots$ and $1_{R[t]} = 1_R + 0_Rt + \cdots$, where we view $R \subset R[t]$ by identifying r and $r1_{R[t]}$ for all $r \in R$, i.e., identifying the ring monomorphism $R \to R[t]$ given by $r \mapsto r1_{\mathbb{R}[t]}$ as the inclusion. So R is just the set (and subring) of constant polynomials in $R[t]$. Recursively define the *polynomial ring* in n *variables* $t_1, \ldots, t_n$ over R by

$$R[t_1, \ldots, t_n] := (R[t_1, \ldots, t_{n-1}])[t_n].$$

If $f = a_0 + a_1t + \cdots + a_nt^n$ is a nonzero polynomial with $a_n \neq 0$, we define the *degree* of f by $\deg f := n$ and the *leading coefficient* of f by $\operatorname{lead} f := a_n$. If $\operatorname{lead} f = 1$, we say f is *monic*. If $f = 0$ or $\deg f = 0$, we call f a *constant polynomial*. We let $\operatorname{lead} 0 = 0$

(and one can define $\deg 0 = -\infty$ if we let $\max(n, -\infty) = n$ and $n + (-\infty) = -\infty$ for any integer n).

Properties 34.1. Let R be a ring and f, g nonzero polynomials.
1. If R is commutative, then so is $R[t]$.
2. $\deg(f+g) \leq \max\{\deg f, \deg g\}$. (Strict inequality is possible, e.g., if $g = -f$.)
3. $\deg(fg) \leq \deg f + \deg g$ with equality if and only if

$$(*) \qquad \mathrm{lead}(fg) = \mathrm{lead}(f)\,\mathrm{lead}(g)$$

   is nonzero.
4. If R is a domain, then (*) holds, so $R[t]$ is a domain.

If we had defined the degree of the zero polynomial to be $-\infty$, these properties would still hold for all polynomials.

Definition 34.2. Let S be a commutative ring and $a_1, \ldots, a_n$ elements in S. If R is a subring of S, then the map

$$e_{a_1,\ldots,a_n} : R[t_1, \ldots, t_n] \to S \text{ given by } f \mapsto f(a_1 \ldots, a_n)$$

is called *evaluation* at $a_1, \ldots, a_n$. We denote the image of $e_{a_1,\ldots,a_n}$ by $R[a_1, \ldots, a_n]$, as elements in this image are sums of the form $ra_1^{i_1} \cdots a_n^{i_n}$ with $r \in R$ and $i_1, \ldots, i_n$ non-negative integers. The evaluation map $e_{a_1,\ldots,a_n}$ is a ring homomorphism and if $R = S$, it is surjective. If R is not commutative, then this map may not be a homomorphism. (Why?) For this reason, we shall only study polynomial rings over commutative rings. Of course if the subring of S generated by R and $a_1, \ldots, a_n$ is commutative, the evaluation map $e_{a_1,\ldots,a_n}$ is still a ring homomorphism as we may replace S by a commutative subring.

If $\varphi : R \to S$ is a ring homomorphism (respectively, ring monomorphism, ring epimorphism, ring isomorphism) of commutative rings, then φ induces a map

$$\begin{aligned} &\tilde{\varphi} : R[t_1, \ldots, t_n] \to S[t_1, \ldots, t_n] \\ &\quad \text{given by } \sum_{i_1} \cdots \sum_{i_n} a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n} \mapsto \sum_{i_1} \cdots \sum_{i_n} \varphi(a_{i_1,\ldots,i_n}) t_1^{i_1} \cdots t_n^{i_n}. \end{aligned}$$

and $\tilde{\varphi}$ is a ring homomorphism (respectively, ring monomorphism, ring epimorphism, ring isomorphism). For convenience, we write the iterated sum by $\sum$ or $\sum_{i_1,\ldots,i_n}$. Moreover, if $a_1, \ldots, a_n$ are elements in R, we have a commutative diagram:

$$\begin{tikzcd} R[t_1, \ldots, t_n] \arrow[r, "\tilde{\varphi}"] \arrow[d, "e_{a_1,\ldots,a_n}"'] & S[t_1, \ldots, t_n] \arrow[d, "e_{\varphi(a_1),\ldots,\varphi(a_n)}"] \\ R \arrow[r, "\varphi"'] & S. \end{tikzcd}$$

In particular, the above applies to the case that $S = R/\mathfrak{A}$ with $\mathfrak{A}$ an ideal in R and φ is the canonical epimorphism $^{-} : R \to R/\mathfrak{A}$. Assume that $\mathfrak{A} < R$ (otherwise we would get no further information). Then the canonical epimorphism induces a ring epimorphism, $\tilde{^{-}} : R[t_1, \ldots, t_n] \to (R/\mathfrak{A})[t_1, \ldots, t_n]$ given by $\sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n} \mapsto \sum_{i_1,\ldots,i_n} \overline{a_{i_1,\ldots,i_n}} t_1^{i_1} \cdots t_n^{i_n}$.

In this case, for simplicity of notation, we write $^-$ for $\tilde{}$. Hence if $(a_1, \dots, a_n) \in R^n$, we have a commutative diagram:

$$\begin{tikzcd} R[t_1, \dots, t_n] \arrow[r, "-"] \arrow[d, "e_{a_1,\dots,a_n}"'] & \overline{R}[t_1, \dots, t_n] \arrow[d, "e_{\overline{a_1},\dots,\overline{a_n}}"] \\ R \arrow[r, "-"] & \overline{R}, \end{tikzcd}$$

i.e.,

$$\overline{f}(\overline{a_1}, \dots, \overline{a_n}) = \overline{f(a_1, \dots, a_n)}.$$

In particular, if no $(a_1, \dots, a_n) \in \overline{R}^n$ satisfies $\overline{f}(a_1, \dots, a_n) = 0$, then no $(x_1, \dots, x_n) \in R^n$ satisfies $f(x_1, \dots, x_n) = 0$. Viewed geometrically, this says $f = 0$ in R^n has no solution if $\overline{f} = 0$ in $\overline{R}^n$ has no solution. This is a very useful idea that is used. It is used, for instance, in number theory to try to show a given Diophantine equation has no solution.

We return to the one variable case. For further remarks in this chapter about polynomials in many variables, see Addendum 36.

Lemma 34.3. *Let R be a domain. Then $R[t]^\times = R^\times$.*

Proof. Clearly, units in R remain units in $R[t]$. Conversely, let f, g be polynomials in $R[t]$ satisfying $fg = 1$. Then $\deg(fg) = \deg f + \deg g = 0$, so $\deg f = \deg g = 0$ and f, g lie in R. □

Note that $\mathbb{Z}/4\mathbb{Z}$ is not a domain and $(\overline{1} + \overline{2}t)^2 = \overline{1}$ in $\mathbb{Z}/4\mathbb{Z}[t]$.

By Exercise 30.22(23), whose solution mimics how you learned to divide polynomials, we have the following:

Theorem 34.4. (General Division Algorithm) *Let f, g be polynomials with coefficients in a commutative ring R. Suppose the leading coefficient of g is a unit. Then there are polynomials q and r with coefficients in R such that $f = gq + r$ with either $r = 0$ or the degree of r is less than the degree of g.*

As usual, if R is a commutative ring, we say that α in R is a *root* of a polynomial f in $R[t]$ if $f(\alpha) = e_\alpha(f) = 0$. The General Division Algorithm has many consequences, most which you have encountered in middle school algebra that we now state and prove.

Corollary 34.5. *Let F be a field. Then $F[t]$ is a euclidean domain. In particular, $F[t]$ is a PID, hence also a UFD.*

Corollary 34.6. (Remainder Theorem) *Let R be a commutative ring, α an element in R, and f a nonzero polynomial in $R[t]$. Then there exists a polynomial q in $R[t]$ satisfying $f = (t - \alpha)q + f(\alpha)$. Moreover,*

$$t - \alpha \mid f \textit{ in } R[t] \textit{ if and only if } f(\alpha) = 0, \textit{ i.e., } \alpha \textit{ is a root of } f.$$

Proof. Apply the General Division Algorithm with $g = t - \alpha$ to get $f = (t - \alpha)q + r$ in $R[t]$ for some $q, r \in R[t]$ with $r = 0$ or $\deg r < \deg(t - \alpha) = 1$. It follows that $r \in R$, hence $f(\alpha) = e_\alpha(f) = e_\alpha\big((t - \alpha)q\big) + e_\alpha(r) = r$. The result now follows easily. □

Corollary 34.7. *Let R be a domain, f a nonzero polynomial in $R[t]$, and $x_1, \ldots, x_n$ distinct roots of f in R. Then $\prod_{i=1}^{n}(t - x_i) \mid f$ in $R[t]$. In particular,* $\deg f \geq n$.

PROOF. If $n = 1$, the result follows by the Remainder Theorem, so assume that $n > 1$. We prove a stronger result by induction on the degree of f. By iterating the Remainder Theorem, we can write $f = (t - x_1)^{r_1} h$ in $R[t]$ for some r_1 in $\mathbb{Z}^+$ and h in $R[t]$ with $h(x_1)$ nonzero. We call r_1 the *multiplicity* of the root x_1 of f in R. As $R[t]$ is a domain, we have $\deg h \leq \deg f - r_1$. For each $i = 2, \ldots, n$, applying the evaluation homomorphism e_{x_i} shows that $0 = f(x_i) = (x_i - x_1)^{r_1} h(x_i)$ in the domain $R[t]$. Hence $h(x_i) = 0$ for $i = 2, \ldots, n$. By induction on the degree of f, we can write $h = \prod_{i=2}^{n}(t - x_i)^{r_i} h_1$ in $R[t]$ with $h_1 \in R[t]$ satisfying $h(x_i) \neq 0$ for $i = 2, \ldots, n$, i.e., x_i is a root of h of multiplicity r_i for $i = 2, \ldots, n$. It follows that $f = \prod_{i=1}^{n}(t - x_i)^{r_i} h_1$ in $R[t]$. The result follows. □

Note the proof shows that the degree of f above is greater than the number of roots of f in R "counted with multiplicity."

Corollary 34.8. (Lagrange) *Let R be a domain and f a nonzero polynomial of degree n in $R[t]$. Then f has at most n roots in R.*

Corollary 34.9. *Let R be a domain and f and g monic polynomials in $R[t]$. If $f(x_i) = g(x_i)$ for n distinct elements $x_1, \ldots, x_n$ in R with $n > \max(\deg f, \deg g)$, then $f = g$. In particular, if R is infinite and $f(x) = g(x)$ for all $x \in R$, then $f = g$.*

Remarks 34.10. 1. The polynomial $t^2 - 1$ in $(\mathbb{Z}/8\mathbb{Z})[t]$ has four roots: $\pm\overline{1}, \pm\overline{3}$, so the assumption that R be a domain in Corollary 34.7 is essential.

2. If R is not a commutative ring, evaluation is not necessarily a ring homomorphism, as we mentioned above. This occurs because we assume that the variable t commutes with all elements of R, but the element x at which we wish to evaluate will, in general, not commute with all elements in R. If x does commute with all elements of R, then evaluation at that element will be a ring homomorphism. The theory of polynomials over non-commutative rings is, therefore, quite different. For example, let $\mathcal{H}$ be the Hamiltonian quaternions, $a, b, c \in \mathbb{R}$ satisfy $1 = N(\alpha) = \alpha\overline{\alpha} = a^2 + b^2 + c^2$ with $\alpha = ai + bj + ck$. Then every such α is a root of the polynomial $t^2 + 1$ in $\mathcal{H}[t]$. So $t^2 + 1$ has infinitely many roots in $\mathcal{H}[t]$.
3. If R is an infinite domain, Corollary 34.9 says that polynomials in $R[t]$ and polynomial functions are essentially the same, where $p : R \to R$ is a *polynomial function* if there exists a polynomial f in $R[t]$ such that $p(x) = f(x)$ for all $x \in R$. We write p_f for this p. So the corollary says that if f and g are polynomials in $R[t]$, with R an infinite domain, then $f \approx g$ if and only if $p_f = p_g$. (Clearly, if $f = g$ then $p_f = p_g$, and the zero polynomial is the only polynomial with infinitely many roots.)
4. More generally if R is a infinite domain and $\varphi : R \to S$ a ring monomorphism, then polynomials in $R[t_1, \ldots, t_n]$ and polynomial functions $p_f : R^n \to S$ given by $p_f(a_1, \ldots, a_n) = e_{\varphi(a_1), \ldots, \varphi(a_n)}(f)$ where $f \in R[t_1, \ldots, t_n]$ satisfies $f \approx g$ if and only if $p_f = p_g$
5. Let p be a prime in $\mathbb{Z}^+$. Then the polynomials t^p and t in $(\mathbb{Z}/p\mathbb{Z})[t]$ are clearly distinct, but by Fermat's Little Theorem, $p_{t^p} = p_t$ in the notation of the previous remark, so the condition that R be infinite in the previous remark is crucial.

For the next corollary, we set up important notation. Let F be a field and $\mathfrak{A} < F[t]$ be an ideal. We have the canonical ring epimorphism

$$\overline{} : F[t] \to F[t]/\mathfrak{A} \text{ given by } g \mapsto \overline{g} = g + \mathfrak{A}.$$

As $\mathfrak{A} < F[t]$, the ring $F[t]/\mathfrak{A}$ is not the trivial ring. Since a field is simple, the canonical epimorphism induces a ring monomorphism

$$\overline{}|_F : F \to F[t]/\mathfrak{A} \text{ given by } a \mapsto \overline{a}.$$

We shall always view this as an inclusion, i.e., for all $a \in F$, we shall identify

$$a \text{ and } \overline{a}.$$

As $\overline{t} = t + \mathfrak{A}$, this identification means that the canonical epimorphism maps $g = \sum a_i t^i \mapsto \overline{g} = \sum \overline{a_i}\,\overline{t}^i = \sum a_i \overline{t}^i$, i.e., the canonical epimorphism under this identification is none other than the evaluation map at $\overline{t}$, i.e.,

$$\overline{} : F[t] \to F[t]/\mathfrak{A} \text{ is the map } g \mapsto g(\overline{t}). \tag{34.11}$$

Corollary 34.12. (Kronecker's Theorem) *Let F be a field and f an irreducible element in $F[t]$. Set $K = F[t]/(f)$. Then K is a field. Viewing $F \subset K$, i.e., as a subfield of K as above, we have $\overline{t}$ is a root of f in K.*

Proof. As f is irreducible in the UFD $F[t]$, it is a prime element. As $F[t]$ is a PID, the nonzero prime ideal (f) is maximal, so K is a field. As $0 = \overline{f} = f(\overline{t})$, by the above, the result follows. $\square$

Corollary 34.13. *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists a field K with $F \subset K$ a subfield such that f has a root in K.*

Proof. We can write $f = f_1 g$ with f_1 and g polynomials in $F[t]$ and f_1 irreducible. By Kronecker's Theorem, f_1 hence f has a root in $F[t]/(f_1)$, a field containing F. $\square$

Of course, to make Kronecker's Theorem and its corollary really useful, we would want to prove results by induction, the natural choice being the degree of a polynomial. That we can do this follows from the following stronger form of Kronecker's Theorem, which we leave as an exercise, but will be essential when we study field theory:

Proposition 34.14. *Let F be a field and f a non-constant polynomial of degree n in $F[t]$. Then $F[t]/(f)$ is a vector space over F of dimension n. In particular, there exists a field K containing F such that f has a root in K and the dimension of K as a vector space over F is at most n.*

We next prove a fundamental fact about finite fields. To do so, we use the First Sylow Theorem. This result is also important in general field theory.

Theorem 34.15. *Let F be a field and G a finite (multiplicative) subgroup of $F^\times$. Then G is cyclic. In particular, if F is a finite field, the group $F^\times$ is cyclic.*

Proof. Let p be a prime that divides $|G|$ and P a Sylow p-subgroup of G (so the only one as G is abelian). Choose x_P in P such that the order of the cyclic subgroup $\langle x_P \rangle$ generated by x_P in P is maximal. Let N be the order of x_P. Then N is a power of p and we must have $y^N = 1$ for all $y \in P$. (Why?) Therefore, the polynomial $t^N - 1$ has $|P|$ distinct roots in F, so $N = |P|$ by Corollary 34.8. It follows that P is cyclic. Let $x = \prod_{\substack{p \mid |G| \\ p \text{ a prime}}} x_P \in G$. As G is abelian, it follows from Exercise 13.7(9) that $G = \langle x \rangle$ is cyclic. $\square$

Example 34.16. Let F be a field and $f = t^n - 1 \in F[t]$. Then the set of roots of f in F is a cyclic group.

We end this section with a short discussion of irreducible polynomials over a field F. Clearly, every *linear* polynomial, i.e., polynomial of degree one, in $F[t]$ is irreducible.

Definition 34.17. The following conditions on a field F are equivalent:

(i) Every non-constant polynomial in $F[t]$ has a root in F.

(ii) The only irreducible polynomials in $F[t]$ are linear.

(iii) Every non-constant polynomial in $F[t]$ factors into a product of linear polynomials.

A field F satisfying these equivalent conditions is called *algebraically closed*.

When we study field theory, we shall prove (cf. 57.12 below):

Theorem 34.18. (Fundamental Theorem of Algebra) *The field of complex numbers is algebraically closed.*

We shall, however, assume here that it has already been established.

Remark 34.19. The Fundamental Theorem of Algebra allows us to compute all irreducible polynomials over the reals. They are

(i) Linear polynomials in $\mathbb{R}[t]$.

(ii) *Quadratic polynomomials* (i.e., polynomials of degree two) of the form $at^2 + bt + c$ in $\mathbb{R}[t]$ satisfying $a \neq 0$ and $b^2 - 4ac < 0$.

In particular, any non-constant polynomial in $\mathbb{R}[t]$ factors into a product of linear and irreducible quadratic polynomials. For example, we have the factorization of the polynomial $t^4 + 1 = (t^2 + \sqrt{2}t + 1)(t^2 - \sqrt{2}t + 1)$. Over any field F, the polynomial ring $F[t]$ contains infinitely many non-associative irreducible polynomials. [Can you prove this?] Finding all irreducible polynomials in $F[t]$, with F or $F[\sqrt{-1}]$ (which is in fact always a field) not algebraically closed is usually very hard, if not impossible, e.g., if $F = \mathbb{Q}$.

One of our goals will be to show that given any field F, there exists an algebraically closed field K containing F and if $F \neq K$, then no field properly between F and K is algebraically closed. Such a field will be called an *algebraic closure* of F. The proof will need Zorn's Lemma. We shall prove this in Section 51 below. In particular, $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$.

Exercises 34.20.

1. Let R be a commutative ring. Show that a polynomial $f = a_0 + a_1 t + \cdots + a_n t^n$ in $R[t]$ is a unit in $R[t]$ if and only if a_0 is a unit in R and a_i is nilpotent for every $i > 0$.
2. Prove Remark 34.10(4).

3. Let R be a nontrivial commutative ring and f a zero divisor in $R[t]$. Show that there exists a nonzero element b in R so that $bf = 0$.
4. Let R be a nontrivial commutative ring. If $f = a_0 + a_1 t + \cdots + a_n t^n$ is a polynomial in $R[t]$, define the *formal derivative* f' of f to be $f = a_1 + 2a_2 t + \cdots + na_n t^{n-1}$.
   (i) Show that the usual rules of differentiation hold.
   (ii) Suppose R is a field of characteristic zero. Show that a polynomial $f \in R[t]$ is divisible by the square of a non-constant polynomial in $R[t]$ if and only if f and f' are not relatively prime.
5. Let R be a ring and G a group (or monoid). Define
$$R[G] := \{\sum_G a_g g \mid a_g \in R \text{ and almost all } a_g = 0\}$$
(where *almost all zero* means that only finitely many are nonzero) with $+$ and $\cdot$ defined by
$$\sum_G a_g g + \sum_G b_g g = \sum_G (a_g + b_g) g \text{ and}$$
$$\sum_G a_g g \cdot \sum_G b_g g = \sum_G c_g g \text{ where } c_g = \sum_{g=hl} a_h b_l$$
for all a_g, b_g in R. Show this is a ring. It is called the *group* (respectively, *monoid*) ring of R by G. If $\mathbb{N} := \mathbb{Z}^+ \cup \{0\}$, show that $R[\mathbb{N}^n]$ is isomorphic to $R[t_1, \cdots, t_n]$. Describe $R[\mathbb{Z}^n]$.
6. Let F be a subfield of the complex numbers $\mathbb{C}$. Let $f \in F[t]$ be an irreducible polynomial. Show that f has no *multiple root* in $\mathbb{C}$, i.e., a root α of f satisfying $(t-\alpha)^n \mid f$ in $F[t]$ with $n > 1$.
7. Prove Proposition 34.14.
8. Prove that the conditions in Definition 34.17 are equivalent.
9. Show that the irreducible polynomials in $\mathbb{R}[t]$ are those stated in Remark 34.19.
10. Show that over any field F, there exist infinitely many monic irreducible polynomials in $F[t]$. Also show that if F is algebraically closed, then F must have infinitely many elements.
11. Let $F = \mathbb{Z}/p\mathbb{Z}$ with p a prime. Show that $t^4 + 1 \in F[t]$ is reducible.

35. Polynomial Rings over a UFD

In this section, we prove a theorem of Gauss that a polynomial ring over a UFD R is itself a UFD. The idea is to take a polynomial over $R[t]$, with R a UFD, and view it over the UFD $K[t]$ where K is the quotient field of R. To make this useful, we must derive a way of pulling information back to $R[t]$.

We have defined the gcd of two nonzero elements in a domain R. We can extend this in the obvious way to a finite number of elements, viz., if $a_1, \ldots, a_n$ are elements in R not all zero, then d in R is a *greatest common divisor* or *gcd* of $a_1, \ldots, a_n$ if it satisfies both of the following:

(i) $d \mid a_i$ for $i = 1, \ldots, n$.

(*ii*) If $e \mid a_i$ for some e in R, $i = 1, \dots, n$, then $e \mid d$.

[We let the gcd of nonzero a in R and 0 be a.]

We have the following lemma that we leave as an exercise.

Lemma 35.1. *Let R be a UFD and $a_1, \dots, a_n$ elements in R, not all zero. Then a gcd of $a_1, \dots, a_n$ exists and is unique up to units.*

Let R be a UFD and $f = a_n t^n + \cdots + a_0$ a nonzero polynomial in $R[t]$. A gcd of $a_0, \dots, a_n$ is called a *content* of f. It is unique up to units. We call f *primitive* if 1 is a content of f. [So a primitive constant polynomial is just a unit in R.] For notational convenience, we often use the notation $C(f)$ for a choice of a content of f.

Remarks 35.2. Let R be a UFD, b a nonzero element of R and f a nonzero polynomial in $R[t]$.

1. If x is an irreducible element in R, then it remains irreducible when considered as an element of $R[t]$. (Look at degrees if it would factor.)
2. $C(bf) \approx bC(f)$, where as before $\approx$ means is an associate of.
3. There exists a primitive polynomial f_1 in $R[t]$ satisfying $f = C(f)f_1$. (As $R[t]$ is a domain, $\deg f = \deg f_1$.) This follows easily as we can factor out a content.
4. If $\deg f > 0$ and f is irreducible in $R[t]$, then f is primitive. Indeed, we can write $f = C(f)f_1$ for some primitive polynomial f_1 in $R[t]$. If $C(f) \not\approx 1$, then it factors into irreducibles in R hence in $R[t]$. It follows that f cannot be irreducible as $\deg f_1 > 0$, so f is not a unit.
5. $t^2 - 1$ is a primitive polynomial, but not irreducible, so the converse to the previous remark is false.

If R is a UFD, to show $R[t]$ is a UFD, we must show that any nonzero nonunit polynomial in $R[t]$ factors into a product of irreducible polynomials in $R[t]$ and that this factorization is essentially unique. The existence is more elementary and will follow from the following lemma and its corollary.

Lemma 35.3. *Let R be a UFD, K the quotient field of R, and f a nonzero polynomial in $K[t]$. Then there exist a primitive polynomial $f_1 \in R[t]$ (of the same degree as f) and an element α in K satisfying $f = \alpha f_1$. Moreover, f_1 and α are unique up to units in R, i.e, up to elements in $R^\times = R[t]^\times$.*

Proof. **Existence**: Write $f = \displaystyle\sum_{i=0}^{n} \frac{a_i}{b_i} t^i$ with a_i, b_i in R and $b_i \neq 0$, $i = 0, \dots, n$. We clear denominators. Let $b = b_0 \cdots b_n \neq 0$ in the domain R. Then $bf \in R[t]$. Let $c = C(bf)$. Then there exists a primitive polynomial f_1 in $R[t]$ such that $bf = cf_1$, so $f = \frac{c}{b} f_1$ in $K[t]$ as needed.

Uniqueness Suppose that $\frac{c}{b} f_1 = f = \frac{d}{e} f_2$ in $K[t]$, with nonzero $c, b, d, e \in R$ and f_1, f_2 primitive in $R[t]$. Then in $R[t]$, we have $cef_1 = bdf_2$. Therefore, $ce \approx C(cef_1) \approx C(bdf_2) \approx bd$ in R. Consequently, there exists a unit u in $R^\times$ satisfying $ce = ubd$ in R, hence $\frac{c}{b} = u\frac{d}{e}$ in K. Thus $\frac{d}{e} f_2 = \frac{c}{b} f_1 = u\frac{d}{e} f_1$ in the domain $K[t]$. It follows that $f_2 = uf_1$. $\square$

Corollary 35.4. *Let R be a UFD, f, g primitive polynomials in $R[t]$, h a polynomial in $R[t]$, and K the quotient field of R.*

(1) If $g = sf$ in $K[t]$ for some s in K, then s is a unit in R.
(2) If $g = ah$ in $R[t]$ for some a in R, then a is a unit in R and h is primitive in $R[t]$.

In particular, if g is non-constant primitive polynomial in $R[t]$ but not irreducible, then $g = q_1q_2$ for some q_1, q_2 in $R[t]$ satisfying $0 < \deg q_i < \deg g$, for $i = 1, 2$.

Proof. (1): We have $1 \cdot g = s \cdot f$ in $K[t]$ with both f and g primitive, so $s = u \cdot 1$ for some unit u in R by Lemma 35.3, hence s is a unit in R.
(2): Write $h = C(h)h_1$ with h_1 primitive in $R[t]$. Then $1 \cdot g = a \cdot C(h)h_1$, so $aC(h) \approx 1$, hence a is a unit in R and h is primitive.
[Note if x is a nonzero element in R then x/x is a unit but x is not necessarily a unit.] $\square$

The key to showing the uniqueness statement for factorization in $R[t]$ when R is a UFD is the following:

Lemma 35.5. (Gauss' Lemma) *Let R be a UFD and f, g non-constant polynomials. Then $C(fg) \approx C(f)C(g)$. In particular, the product of primitive polynomials in $R[t]$ is primitive.*

Proof. Write $f = C(f)f_1$ and $g = C(g)g_1$ with f_1 and g_1 primitive polynomials in $R[t]$. Then
$$C(fg) \approx C(C(f)f_1C(g)g_1) \approx C(f)C(g)C(f_1g_1)$$
and
$$C(f)C(g) \approx C(f)C(f_1)C(g)C(g_1) \approx C(f)C(g)C(f_1)C(g_1),$$
so it suffices to show the last statement, i.e., we may assume that f and g are primitive in $R[t]$ and must show fg is primitive in $R[t]$. Suppose this is false. Then there exists an irreducible element p such that $p \mid C(fg)$ in R. As R is a UFD, p is a prime element, hence $\overline{R} = R/(p)$ is a domain. Let $^{-} : R[t] \to \big(R/(p)\big)[t]$ be the ring epimorphism given by $\sum a_it^i \mapsto \sum \overline{a_i}t^i$. By assumption, $\overline{0} = \overline{fg} = \overline{f}\,\overline{g}$ in the domain $\overline{R}[t]$. As f and g are primitive in $R[t]$, the prime p does not divide some coefficient of f and some coefficient of g, i.e., $\overline{f} \neq 0$ and $\overline{g} \neq 0$ in $\overline{R}[t]$, contradicting the fact that $\overline{R}[t]$ is a domain. $\square$

Corollary 35.6. *Suppose that R is a UFD with quotient field K. Let f, g be non-constant primitive polynomials in $R[t]$ and h a non-constant polynomial in $K[t]$. If $f = gh$ in $K[t]$, then h lies in $R[t]$ and is primitive.*

Proof. By Lemma 35.3, we can write $h = \alpha h_1$ with $\alpha \in K$ and h_1 a primitive polynomial in $R[t]$, so $f = \alpha gh_1$. By Gauss' Lemma, gh_1 is primitive, so α lies in $R^{\times}$ by Corollary 35.4. It follows that h lies in $R[t]$ and is primitive. $\square$

A similar proof yields:

Lemma 35.7. *Let R be a UFD with quotient field K and f a non-constant irreducible polynomial in $R[t]$. Then f remains irreducible in $K[t]$.*

PROOF. As f is a non-constant irreducible polynomial in $R[t]$, it is primitive. Suppose that f factors as $f = g_1g_2$ in $K[t]$, with $g_i \in K[t]$ and $0 < \deg g_i < \deg f$, for $i = 1, 2$. By Lemma 35.3, we can write $g_i = \alpha_i h_i$ with $\alpha_i \in K$ and h_i in $R[t]$ primitive (and $\deg g_i = \deg h_i$), for $i = 1, 2$. Therefore, we have $f = (\alpha_1\alpha_2)h_1h_2$. By Gauss' Lemma, h_1h_2 is primitive, so by Corollary 35.4, we have $\alpha = \alpha_1\alpha_2$ is a unit in R and $f = \alpha h_1h_2$ in $R[t]$. It follows that f is reducible in $R[t]$, a contradiction. □

We can now prove our theorem.

Theorem 35.8. *Let R be a UFD. Then $R[t]$ is a UFD.*

PROOF. **Existence**: Every nonzero element in R is a unit or factors into irreducibles that remain irreducible in $R[t]$, so it suffices to factor a non-constant polynomial in $R[t]$. Let f be such a polynomial. As $f = C(f)f_1$ with f_1 a primitive polynomial in $R[t]$, we may also assume that f is primitive. If f is not irreducible, then $f = g_1g_2$ for some polynomials $g_i \in R[t]$ satisfying $0 < \deg g_i < \deg f$ for $i = 1, 2$ by Corollary 35.4. By induction on $\deg f$, we conclude that g_1 and g_2 hence f factor into irreducible polynomials in $R[t]$.
Uniqueness: We know that any nonzero element in R has a unique factorization (up to associates and order), so the content of a polynomial factors uniquely up to units. It follows by Lemma 35.3 that it suffices to show any non-constant primitive polynomial f has a unique factorization (up to associates and order). Suppose that

$$f_1 \cdots f_r = f = g_1 \cdots g_s$$

with all the f_i and g_j irreducible in $R[t]$. By Lemma 35.7, all the f_i and g_j remain irreducible in $K[t]$, where K is the quotient field of R. As $K[t]$ is a UFD, $r = s$, and after relabeling, we may assume that $f_i = c_ig_i$ in $K[t]$ for some $c_i \in K^\times$ for each i. But the f_i and g_j are primitive in $R[t]$ as they are irreducible, so $c_i \in R^\times$ for all i by Corollary 35.4. We conclude that $f_i \approx g_i$ in $R[t]$ for all i, and the proof is complete. □

Corollary 35.9. *If R is a UFD, then so is $R[t_1, \dots, t_n]$.*

Examples 35.10. Let R be a UFD and F a field.

1. $\mathbb{Z}[t]$ and $F[t_1, t_2]$ are UFDs but not PIDs. In fact, $R[t]$ is a PID if and only if R is a field.
2. A polynomial ring in any number of variables [even infinitely many — definition of such a polynomial ring?] over R is a UFD.
3. Let $\varphi : R \to S$ be a ring epimorphism of domains. Then S is not necessarily a UFD (even if a domain), e.g. $e_{\sqrt{-5}} : \mathbb{Z}[t] \to \mathbb{Z}[\sqrt{-5}]$ is such an example.

We end this discussion with a few remarks about irreducible elements in $R[t]$, with R a domain.

Remarks 35.11. 1. (Eisenstein's Criterion) Let R be a UFD and K its quotient field. Let $f = \sum_{i=0}^{n} a_it^i$ be a non-constant polynomial in $R[t]$ of degree n. Suppose that there exists an irreducible element p in R (hence a prime) satisfying:

(i) $p \nmid a_n$.

   (ii) $p \mid a_i$, for $i = 0, \dots, n-1$.
   (iii) $p^2 \not| a_0$.
   Then f is irreducible in $K[t]$. In particular, if f is primitive in $R[t]$, then it is irreducible in $R[t]$.
   We leave this as an exercise.
2. If R is a domain, f a non-constant polynomial in $R[t]$, let $y = t + a$, for some a in R, and set $g(y) = f(y - a)$. Then f is irreducible in $R[t]$ if and only if g is irreducible in $R[y]$ (as $R[t] \to R[y]$ defined by $\sum a_i t^i \to \sum a_i y^i$ is a ring isomomorphism).
3. Let $p \in \mathbb{Z}^+$ be a prime. Then the polynomial $t^{p-1} + t^{p-2} + \cdots + t + 1$ is irreducible in $\mathbb{Z}[t]$ and $t^p - 1 = (t-1)(t^{p-1} + t^{p-2} + \cdots + t + 1)$ is a factorization into irreducibles in $\mathbb{Z}[t]$ (and $\mathbb{Q}[t]$):
   Let $y = t - 1$. Then
$$\begin{aligned} f &= t^{p-1} + t^{p-2} + \cdots + t + 1 = \frac{t^p - 1}{t - 1} \\ &= \frac{(y+1)^p - 1}{y} = y^{p-1} + py^{p-2} + \binom{p}{2} y^{p-3} + \cdots + p. \end{aligned}$$
   As $p \mid \binom{p}{i}$ in $\mathbb{Z}$ for $i = 1, \dots, p-1$, we conclude that f is irreducible in $\mathbb{Z}[t]$ by Eisenstein's Criterion and the previous remark.
4. Let R be a commmutative ring, $\mathfrak{A} < R$ an ideal, $^{-} : R \to R/\mathfrak{A}$ the canonical epimorphism. If $f = g_1 \cdot g_2$ in $R[t]$, then $\overline{f} = \overline{g_1} \cdot \overline{g_2}$ in $\overline{R}[t]$. In particular, if R is a domain, $\mathfrak{A}$ a prime ideal, f monic, then $\overline{f}$ irreducible in $\overline{R}[t]$ implies f is irreducible in $R[t]$.

Exercises 35.12.

1. Prove Lemma 35.1.
2. Let R be a domain that is not a field. Show that $R[t]$ is not a PID.
3. Let R be a UFD, K its quotient field. Let f and g be non-constant polynomials in $R[t]$. Write $f = C(f) f_1$ and $g = C(g) g_1$ with f_1 and g_1 primitive polynomials in $R[t]$. Show
   (i) If $f|g$ in $K[t]$ then $f_1|g_1$ in $R[t]$. In particular, if f and g are primitive, then $f|g$ in $K[t]$ if and only if $f|g$ in $R[t]$.
   (ii) Suppose that f and g are primitive. Then f and g have a common factor over $K[t]$ if and only if they have a common factor over $R[t]$.
4. Let R be a commutative ring, and $\{t_1, \mid i \in I\}$ indeterminants. Define the polynomial ring $R[t_i]_{i \in I}$ in the variables $\{t_1, \mid i \in I\}$. Show that it is a UFD if R is.
5. Let $f = \sum_{i=0}^{n} a_i t^i$ be a polynomial in $\mathbb{Z}[t]$ with $a_n \neq 0$. Let $r = a/b$ with b nonzero and a and b relatively prime integers. If r is a root of f, show that $b \mid a_n$ and if a is also nonzero, then $a \mid a_0$. In particular, if f is monic, any rational roof of f, if any, is an integer.
6. Prove Eisenstein's Criterion (Remark 35.11(1)).
7. Let $y = t + a$. Show that the map $R[t] \to R[y]$ given by $\sum a_i t^i \mapsto \sum a_i y^i$ is a ring isomorphism.

8. Prove the identity of binomial coefficients
$$\binom{n+1}{n-j} = \sum_{i=j}^{n} \binom{i}{i-j}$$
and use it to prove that the polynomial
$$t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$$
is an irreducible polynomial in $\mathbb{Z}[t]$ (hence $\mathbb{Q}[t]$) for every (positive) prime p in $\mathbb{Z}$.

36. Addendum: Polynomial Rings over a Field

In this addendum, we make some motivational remarks that we shall come back to when studying modules. If R is a commutative ring, we have defined the ring $R[t_1, \ldots, t_n]$. We view R as a subring of $R[t_1, \ldots, t_n]$ of constant polynomials. We look at the case, when R is a field.

Let F be a field and S be a nontrivial commutative ring. We have seen that any ring homomorphism $\varphi : F \to S$ is monic. Let $\mathfrak{A} < F[t_1, \ldots, t_n]$ be an ideal. We apply this to $S = F[t_1, \ldots, t_m]/\mathfrak{A}$. Let $^{-} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$ be the canonical epimorphism. Then the composition

$$(*) \qquad F \subset F[t_1, \ldots, t_n] \xrightarrow{-} F[t_1, \ldots, t_n]/\mathfrak{A}$$

is also monic. As before, we view this as an embedding, i.e., identify a in F with $\overline{a}$ in $F[t_1, \ldots, t_n]/\mathfrak{A}$. Under this identification the map $^{-} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$ becomes $f \mapsto \overline{f} = f(\overline{t_1}, \ldots, \overline{t_n})$, so it is the evaluation map $e_{(\overline{t_1}, \ldots, \overline{t_n})} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$. Observe that if $\mathfrak{A} \subset \mathfrak{B} < F[t_1, \ldots, t_n]$ are ideals, then we have a commutative diagram

$$\begin{tikzcd}
F[t_1, \ldots, t_n] \arrow[rr] \arrow[dr] & & F[t_1, \ldots, t_n]/\mathfrak{B} \\
& F[t_1, \ldots, t_n]/\mathfrak{A} \arrow[ur] &
\end{tikzcd}$$

with all the maps the obvious ring epimorphisms. The most interesting case is when $\mathfrak{B}$ is a maximal ideal, hence $F[t_1, \ldots, t_n]/\mathfrak{B}$ is a field. We look at a special case of this.

Let $\underline{x} = (x_1, \ldots, x_n)$. For each $\underline{x} = (x_1, \ldots, x_n)$ in F^n set
$$\mathfrak{m}_{\underline{x}} = (t_1 - x_1, \ldots, t_n - x_n),$$
an ideal in $F[t_1, \ldots, t_n]$. Let $\mathfrak{A} = \mathfrak{m}_{\underline{x}}$ in (*). For each $i = 1, \ldots, n$, we have $\overline{t_i} = x_i$ lies in F, so we can view $F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}} \subset F$. By (*), we conclude that $F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}} = F$. In particular, $\mathfrak{m}_{\underline{x}}$ is a maximal ideal in $F[t_1, \ldots, t_n]$. We conclude that the canonical epimorphism $^{-} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}}$ takes $f \in F[t_1, \ldots, t_n]$ to $\overline{f} = f(\underline{t}) = f(\underline{x}) = e_{\underline{x}}(f)$ in F. So we have
$$f(\underline{x}) = 0 \text{ in } F \text{ if and only if } \overline{f} = 0 \text{ in } F \text{ if and only if } f \in \mathfrak{m}_{\underline{x}}.$$

The question of interest is to solve the following:

Problem 36.1. *Let $f_1, \ldots, f_r$ be polynomials in $F[t_1, \ldots, t_n]$. Does there exist an $\underline{x}$ in F^n such that $f_i(\underline{x}) = 0$ for $i = 1, \ldots r$, i.e., a common point lying on all the hypersurfaces $f_1 = 0, \ldots, f_r = 0$ in F^n?*

Let $\mathfrak{A} = (f_1, \ldots, f_r)$, the ideal generated by the f_i in $F[t_1, \ldots, t_n]$ and $\underline{x} \in F^n$. Then we have $f_i(\underline{x}) = 0$ for $i = 1, \ldots, r$ if and only if $f(\underline{x}) = 0$ for all $f \in \mathfrak{A}$. We set up a useful notation. If $\mathfrak{B}$ is an ideal in $F[t_1, \ldots, t_n]$, let

$$Z_F(\mathfrak{B}) := \{\underline{x} \in F^n \mid f(\underline{x}) = 0 \text{ for all } f \in \mathfrak{B}\}.$$

We call $Z_F(\mathfrak{B})$ the *affine variety* defined by $\mathfrak{B}$ in F^n. If $\mathfrak{B} = (g_1, \ldots, g_s)$, we write $Z_F(g_1, \ldots, g_s)$ for $Z_F(\mathfrak{B})$.

So our problem is: Is $Z_F(\mathfrak{A})$ nonempty?

If $\mathfrak{A}$ is the unit ideal, then $Z_F(\mathfrak{A})$ is empty as $f(\underline{x}) = 1$ if f is the constant function 1. So a necessary condition is that $\mathfrak{A} < F[t_1, \ldots, t_n]$. The equation $t_1^2 + t_2^2 = -1$ has no solution in $\mathbb{R}^2$, so $Z_{\mathbb{R}}(t_1^2 + t_2^2 + 1)$ is empty. In particular, $(t_1^2 + t_2^2 + 1)$ lies in no $\mathfrak{m}_{\underline{x}}$ with $\underline{x}$ in $\mathbb{R}^n$. Therefore, in general, the answer is no, and we would need some stronger condition on the field F.

Next suppose that $\mathfrak{A}$ is an ideal in $F[t_1, \ldots, t_n]$ such that there exists an $\underline{x} \in F^n$ with $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$. In particular, if $f \in \mathfrak{A}$, then $f \in \mathfrak{m}_{\underline{x}}$, i.e., if $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$, then $\underline{x} \in Z_F(\mathfrak{A})$. Conversely, suppose that $\underline{x} \in Z_F(\mathfrak{A})$. If $f \in \mathfrak{A}$, then we have $f(\underline{x}) = 0$, hence f lies in the kernel of the map $e_{\underline{x}} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}}$, i.e., $f \in \mathfrak{m}_{\underline{x}}$. Therefore,

$$\mathfrak{A} \subset \mathfrak{m}_{\underline{x}} \text{ if and only if } \underline{x} \in Z_F(\mathfrak{A}). \tag{36.2}$$

This says that the answer to our problem will be yes for every ideal $\mathfrak{A} < F[t_1, \ldots, t_n]$, if for each ideal $\mathfrak{A}$, there exists an $\underline{x} \in F^n$ such that $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$. In particular, this will be true if and only if whenever $\mathfrak{m}$ is a maximal ideal in $F[t_1, \ldots, t_n]$, there exists an $\underline{x} \in F^n$ satisfying $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$. (Cf. Exercise 26.21(13).) [Recall we have shown (using Zorn's Lemma) that every nonunit ideal in a nontrivial commutative ring lies in a maximal ideal (for a Noetherian ring this would follow by the Maximal Principle).]

Hilbert proved the following:

Theorem 36.3. (Hilbert Basis Theorem) *Let R be a Noetherian ring. Then $R[t_1, \ldots, t_n]$ is Noetherian. In particular, if F is a field then $F[t_1, \ldots t_n]$ is Noetherian.*

This means that if $\mathfrak{A}$ is ideal in $F[t_1, \ldots, t_n]$ with F a field, then $Z_F(\mathfrak{A}) = Z_F(f_1, \ldots, f_r)$, for some $f_1, \ldots, f_r$ in $F[t_1, \ldots, t_r]$. He also proved the following wonderful theorem:

Theorem 36.4. (Hilbert Nullstellensatz) *Let F be an algebraically closed field. Then every maximal ideal in $F[t_1;, \ldots, t_n]$ is of the form $\mathfrak{m}_{\underline{x}}$ for some $\underline{x}$ in F^n. In particular, if $\mathfrak{A} < F[t_1, \ldots, t_n]$ is an ideal, then $Z_F(\mathfrak{A})$ is not empty.*

The above is also called the *Weak Hilbert Nullstellensatz*. It says, under the additional requirement that F be algebraically closed, the map

$$F^n \longrightarrow \{\mathfrak{m} \mid \mathfrak{m} < F[t_1, \ldots, t_n] \text{ a maximal ideal } \} \text{ given by } \underline{x} \mapsto \mathfrak{m}_{\underline{x}}$$

is a bijection. Actually, Hilbert proved more. In the above, let

$$\sqrt{\mathfrak{A}} = \{f \in F[t_1, \dots, t_n] \mid f^n \in \mathfrak{A} \text{ for some } n \in \mathbb{Z}^+\},$$

the *radical* of the ideal $\mathfrak{A}$. Then the *Hilbert Strong Nullstellensatz* says, if F is algebraically closed, $\mathfrak{A} < F[t_, \dots t_n]$ an ideal, then

$$\sqrt{\mathfrak{A}} = \bigcap_{\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}} \mathfrak{m}_{\underline{x}}.$$

We will return to this later and prove these assertions.

Exercises 36.5. 1. Let $R = F[t_1, \dots, t_n]$, F a field, and $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{A}_i, i \in I$, be ideals in R. Show the following:

(i) If $\mathfrak{A} \subset \mathfrak{B}$, then $Z_F(\mathfrak{B}) \subset Z_F(\mathfrak{A})$.
(ii) $Z_F(\emptyset) = F^n$
(iii) $Z_F(R) = \emptyset$.
(iv) $Z_F(\sum_I \mathfrak{A}_i) = \bigcap_I Z(\mathfrak{A}_i)$.
(v) $Z_F(\mathfrak{A}\mathfrak{B}) = Z_F(\mathfrak{A} \cap \mathfrak{B}) = Z_F(\mathfrak{A}) \cup Z_F(\mathfrak{B})$.
(vi) $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$.

2. Let $f, g \in \mathbb{C}[t_1, t_2] \setminus \mathbb{C}$. Suppose that

$$Z(f) \cap Z(g) = \{(x_i, y_i) \mid i = 1, \dots n\}.$$

Using $Z(f) \cap Z(g)$, show that there exists a ring epimormphism $\varphi : \mathbb{C}[t_1, t_2]/(f, g) \to \bigtimes_{i=1}^n \mathbb{C}$ (with coordinate operations) which induces an isomorphism $\mathbb{C}[t_1, t_2]/(f) \cap (g) \to \bigtimes_{i=1}^n \mathbb{C}$.

37. Addendum: Algebraic Weierstraß Preparation Theorem

If R is a commutative ring, we have defined the ring of formal power series $R[[t]]$. Inductively, let $R[[t_1, \dots, t_n]] := (R[[t_1, \dots, t_{n-1}]])[[t_n]]$, the formal power series in the variables $t_1, \dots, t_n$.

Previous exercises imply that $R[[t_1, \dots, t_n]]^\times = R^\times + (t_1, \dots, t_n)$, i.e., formal power series with unit constant term, (cf. Exercise 26.21(2)) and if R is a domain, then so is $R[[t_1, \dots, t_n]]$ (cf. Exercise 26.21(1)) Moreover, if R is Noetherian, so is $R[[t_1, \dots, t_n]]$ by Example 30.13(2). Further, as with polynomials, if $a_1, \dots, a_n$ lie in R, then the evaluation map $e_{a_1, \dots, a_n} : R[[t_1, \dots, t_n]] \to R$ by $f \mapsto f(a_1, \dots, a_n)$ is a ring epimorphism.

In this section we shall show that if F is a field then $F[[t_1, \dots, t_n]]$ is a UFD. By Exercise 37.11(2), it is also a *local ring*, i.e., commutative ring with a unique maximal ideal. One nice thing about formal power series is that we do not have to worry about convergence. This means that proving the algebraic analogue of the Weierstraß Preparation Theorem in complex analysis becomes a formal proof, hence much easier.

Definition 37.1. Let F be a field and f an element of $F[[t_1, \dots, t_n]]$. We say f is *regular* in t_n if $f(0, \dots, 0, t_n) \neq 0$. So if f is regular in t_n, the element f has a term ct_n^i for some nonzero element c of F. If f is regular, write

$$f = g_0 + g_1 t_n + g_2 t^2 + \cdots \quad \text{with } g_i \in F[[t_1, \dots, t_{n-1}]] \text{ for all } i$$

and satisfying

$$0 \neq f(0,\ldots,0,t_n) = g_0(0,\ldots,0) + g_1(0,\ldots,0)t_n + g_2(0,\ldots,0)t_n^2 + \cdots$$

If s is the first i such that $g_i(0,\ldots,0)$ is nonzero, we say s is the *order of* t_n in $f(0,\ldots 0,t_n)$ and write $s = \operatorname{ord}_{t_n} f$.

We need two lemmas.

Lemma 37.2. *Let g be a nonzero element in $F[[t_1,t_2]]$. Then there exists a positive integer r satisfying $g(t_2^r,t_2)$ is nonzero.*

PROOF. Order the nonzero monomials $a_{r_1r_2}t_1^{r_1}t_2^{r_2}$ of g lexicographically, i.e., $(r_1,r_2) \geq (s_1,s_2)$ if $r_1 > s_1$ or $r_1 = s_1$ and $r_2 \geq s_2$. Let $a_{s_1s_2}$ be the smallest monomial in g relative to this ordering. Choose $r > s_2$ and let $t_1 \mapsto t_2^r$. If $(r_1,r_r) > (s_1,s_2)$, then $rr_1 > rs_1 + s_2$ since either $r_1 > s_1$ (and $r > s_2$) or $r_1 = s_1$ and $r_2 > s_2$. The only term of order $rs_1 + s_2$ in $g(t_2^r,t_2)$ is $a_{s_1s_2}t_2^{rs_1+s_2}$. $\square$

The next lemma shows that the notion of regularity is rather mild.

Lemma 37.3. *Let $f_1,\ldots,f_m$ be nonzero elements in $F[[t_1,\ldots,t_n]]$. Then there exists a ring automorphism σ of $F[[t_1,\ldots,t_n]]$ fixing $F[[t_n]]$ with $\sigma(f_i)$ regular in t_n for $i = 1,\ldots,m$.*

PROOF. As $F[[t_1,\ldots,t_n]]$ is a domain, $f = f_1\cdots f_m$ is nonzero. So we are done if we find a σ that works for f, i.e., we may assume that $f = f_1$. By the previous lemma, there exist $r_1,\ldots,r_{n-1}$ such that $f(t_n^{r_1},t_2,\ldots t_n) \neq 0$, $f(t_n^{r_1},t_n^{r_2},t_3,\ldots t_n) \neq 0$, $\ldots$, $f(t_n^{r_1},\ldots,t_n^{r_n-1},t_n) \neq 0$. Let σ be the ring automorphism of $F[[t_1,\ldots,t_n]]$ fixing $F[[t_n]]$ induced by $t_1 \mapsto t_i + t_n^{r_i}$ for $i = 1,\ldots,n-1$. [Why does it exist? Its inverse is $t_1 \mapsto t_i - t_n^{r_i}$.] Then $\sigma\big(f(t_1,\ldots,t_n)\big) = f(t_1 + t_n^{r_1},\ldots,t_{n-1} + t_n^{r_{n-1}},t_n)$. Consequently,

$$e_{(0,\ldots,0,t_n)}\big(\sigma(f)\big) = f(t_n^{r_1},t_n^{r_2},\ldots,t_n^{r_{n-1}},t_n) \neq 0. \qquad \square$$

Theorem 37.4. (Algebraic Weierstraß Preparation Theorem) *Let $R = F[[t_1,\ldots,t_n]]$ with F a field and $f \in R$ regular in t_n. Suppose that $s = \operatorname{ord}_{t_n} f$ and g lies in R. Then there exist unique elements h and r in R satisfying $g = hf + r$ with $r \in F[[t_1,\ldots,t_{n-1}]][t_n]$ and* $\deg_{t_n} r < s$.

PROOF. We induct on n.

Suppose that $n = 1$ and $t = t_1$: Then

$$f = c_st^s + \sum_{j>s} c_jt^j \quad \text{with} \quad c_s \neq 0$$

$$g = \sum_{i=0}^{s-1} a_it^i + \sum_{i\geq s} a_it^i = \sum_{i=0}^{s-1} a_it^i + t^s\big(\sum_{i\geq s} a_it^{i-s}\big).$$

Let

$$r = \sum_{i=0}^{s-1} a_it^i \quad \text{and} \quad h = \big(\sum_{i\geq s} a_it^{i-s}\big)\big(\sum_{j\geq s} c_jt^{j-s}\big)^{-1}.$$

[Note that $c_s \neq 0$, so the second power series in h is a unit by Exercise 26.21(2).] Then r, h satisfy $g = hf + r$. Since $hf = 0$ or $\mathrm{ord}_t(hf) \geq s$, we see that $r = \big(\sum_{i=0}^{s-1} a_i t^i\big)$ is uniquely determined and hence so is h.

Suppose that $n > 1$: The result holds for $R_0 = F[[t_2, \dots, t_n]]$ by induction. Write

$$f = \sum_{i=0}^{\infty} f_i t_1^i \quad \text{and} \quad g = \sum_{i=0}^{\infty} g_i t_1^i$$

with f_i, g_i in R_0 for all i. Set $h = \sum_{i=0}^{\infty} h_i t_1^i$ and $r = \sum_{i=0}^{\infty} r_i t_1^i$ with all the h_i and r_i in R_0 to be determined by the equation

$$\sum_{i=0}^{\infty} r_i t_1^i = \sum_{i=0}^{\infty} g_i t_1^i - \big(\sum_{i=0}^{\infty} h_i t_1^i\big)\big(\sum_{i=0}^{\infty} f_i t_1^i\big),$$

i.e.,

$$(*) \qquad \begin{aligned} r_0 &= g_0 - h_0 f_0 \\ r_1 &= g_1 - (h_0 f_1 + h_1 f_0) \\ &\vdots \\ r_i &= g_i - (h_0 f_i + \cdots + h_i g_0) \\ &\vdots \end{aligned}$$

Since $f(0, \dots, 0, t_n) = f_0(0, \dots, t_n)$ and f_0 in R_0 is regular in t_n with $s = \mathrm{ord}_{t_n}$, by induction there exist unique h_0, r_0 with the appropriate properties, hence unique $r_1, h_1, \dots$, etc. working from the top of (*) down with $\deg_{t_n} r_i < s$ for each i, where we have replaced f by f_0 and g by $g_0 - h_0 f_1$, etc. As all the h_i, r_i are unique, so are r and h. □

We wish to prove that $F[[t_1, \dots, t_n]]$ is a UFD when F is a field by reducing to the polynomial case where we can use the fact that a polynomial ring over a UFD is a UFD. We begin with the following definition.

Definition 37.5. Let $f \in F[[t_1, \dots, t_n]]$ with F a field. Then f is called a *pseudo-polynomial* of *degree* s, if $f = t_n^s + a_1 t_n^{s-1} + \cdots + a_s$ with a_i in $F[[t_1, \dots, t_{n-1}]]$ and satisfying $a_i(0, \dots, 0) = 0$ for $1 \leq i \leq s$.

In this language, the Weierstraß Preparation Theorem implies:

Corollary 37.6. *Let $R = F[[t_1, \dots, t_n]]$ with F a field and f an element of R regular in t_n with $s = \mathrm{ord}_{t_n} f$. Then there exists a unique pseudo-polynomial of degree s that is an associate of f in $F[[t_1, \dots, t_n]]$.*

Proof. Apply the Weierstraß Preparation Theorem to f and $g = t_n^s$, to obtain unique elements $h \in R$ and $r \in F[[t_1, \dots, t_{n-1}]][t_n]$ with $\deg_{t_n} r < s$ satisfying $t_n^s = hf - r$. Write $r = \sum_{i=0}^{s-1} r_i t^i$ with r_i in $F[[t_1, \dots, t_{n-1}]]$ for all i.

Claim. $f^* = t_n^s + r$ works.

Since $f^* = hf$ with h and r unique, it suffices to show that $r_i(0, \dots, 0) = 0$ for all i and $h \in R^\times$. For the latter, by Exercise 26.21(2), it suffices to show that $h(0, \dots, 0)$ is nonzero. But if it is, then hf has no monomial of the form at_n^s with $a \in F^\times$, contradicting $f^* = t_n^s + r$. Therefore, h is a unit. Write $f = c_s t^s + \sum_{j>s} f_j t_n^j$ with c_s nonzero in F and each $f_j \in F[[t_1, \ldots, t_{n-1}]]$. Then

$$h(0, \ldots, 0, t_n)\big(c_s t_n^s + \sum_{j>s} f_j(0, \ldots, 0) t_n^j\big) = t_n^s + \sum_{i=0}^{s-1} r_i(0, \ldots, 0) t_n^i.$$

Comparing t_n^s terms shows that $r_i(0, \ldots 0) = 0$ for all i. □

Notation 37.7. We denote the unique pseudo-polynomial of f in Corollary 37.6 by f^*.

Corollary 37.8. *Let $R = F[[t_1, \ldots, t_n]]$ with F a field and f and g elements of R both regular in t_n. Then $(fg)^* = f^* g^*$.*

Proof. Note that fg is regular at t_n as f and g are. By the previous result, there exist unique units $u, v, w \in R^\times$ satisfying $f^* = uf$, $g^* = vg$, and $(fg)^* = wfg$ pseudo-polynomials of the appropriate degree. Thus $f^* g^* = uvfg$ and uniqueness shows that $(fg)^* = f^* g^*$. □

Theorem 37.9. *Let F be a field. Then $F[[t_1, \ldots, t_n]]$ is a UFD.*

Proof. We induct on n, the case $n = 0$ being immediate. So we may assume that $n \geq 1$. Let $R_0 = F[[t_1, \ldots, t_{n-1}]]$ and $R = R_0[[t_n]] = F[[t_1, \ldots, t_n]]$. By induction R_0 is a UFD, hence so is the polynomial ring $R_0[t_n]$ (by Theorem 35.8). As R is Noetherian by Example 30.13(2), every nonzero nonunit in R is a product of irreducibles by Theorem 30.14. Let $f \in R$ be irreducible. Therefore, it suffices to show that f is a prime element of R. Suppose that $f \mid rs$ in R with r and s in R. Write $fg = rs$ with $g \in R$. By Lemma 37.3, there exists a ring automorphism σ of R fixing $R_0[[t_n]]$ and satisfying $\sigma(f)$, $\sigma(g)$, $\sigma(r)$, and $\sigma(s)$ are all regular in t_n. Therefore, we may assume that f, g, r, s are all regular in t_n. By Corollary 37.6, we have $f^* g^* = r^* s^*$, so we may further assume that f, g, r, s are all pseudo-polynomials in $R_0[t_n]$. Since $R_0[t_n]$ is a UFD, we are reduced to showing:

Claim. f is irreducible in $R_0[t_n]$.

If $f \in R_0[t_n]$ is reducible, then $f = f_1 f_2$ with $f_1, f_2 \in R_0[t_n]$ satisfying $\deg_{t_n} f_i < \deg_{t_n} f$ for $i = 1, 2$. (Why?) But f is irreducible in R, so either f_1 or f_2 lies in $R^\times$, say f_1. Then

$$0 = f_1^{-1} f - f_2$$
$$0 = 0 \cdot f - 0$$

in R. By the uniqueness statement in the Weierstraß Preparation Theorem, $f_1^{-1} = 0$, which is impossible. This proves the claim and finishes the proof of the theorem. □

Remark 37.10. In general, if R is a UFD, it does not follow that $R[[t]]$ is a UFD. For example, it can be shown that $R = F[t_1, t_2, t_3]/(t_1^2 + t_2^3 + t_3^7)$ is a UFD but $R[[t]]$ is not.

Exercises 37.11.

1. Prove that the map σ in the proof of Lemma 37.3 is well-defined.
2. Let $R = F[[t_1, \ldots, t_n]]$ with F a field. Show that $\mathfrak{m} = (t_1, \ldots, t_n)$ is the unique maximal ideal in R and $\mathfrak{m}/\mathfrak{m}^2$ is a vector space over F of dimension n.

Part 4

Module Theory

CHAPTER VIII

Modules

In this chapter, we study the basic result about modules. A module is a generalization of abelian groups and vector spaces. It is an object satisfying all the axioms of a vector space, except that the scalars are allowed to come from any fixed ring instead of from a field. The basic problem in module theory is given a ring with nice properties, classify modules over this ring up to isomorphism. Of course, this meets with relatively little success for all modules, so one looks at classes of modules over the ring.

In the first section, after giving examples of modules over a given ring, we establish the analogue of the isomorphism theorems and correspondence principle. Modules are nicer, in that we do not need a stronger property than submodules to state and prove these results compared to those needing normal subgroups and ideals in group and ring theory respectively. In particular, the quotient module of any given module and a submodule always exists.

We also study the special class of free modules. This class of modules generalizes the notion of vector space. More precisely free module are those that have bases, i.e., linearly independent spanning sets. This leads to the fact that dimension (called the rank) of a free module over a commutative ring makes sense, i.e., we can attach a unique integer to a finitely generated free module over a commutative ring. [Cf. the dimension of a finite dimensional vector space.]

38. Basic Properties of Modules

In this section, we introduce the concept of a module over a ring. As mentioned above, this generalizes the definition of vector spaces, by allowing scalars to come from an arbitrary ring rather than just a field. This difference, however, makes much that one expects in vector space theory to be false in this general case. One of the main reasons for this is that the concept of linear independence becomes a rare phenomenon. We begin with the formal definition.

Definition 38.1. Let R be a ring. A (*left*) *R-module* is an additive group M together with a map, $\cdot : R \times M \to M$ called *scalar multiplication*, which we write as $r \cdot x$ for $\cdot(r, x)$, satisfying for all $r, s \in R$ and $x, y \in M$:

(*i*) $r \cdot (x + y) = r \cdot x + r \cdot y$.
(*ii*) $(r + s) \cdot x = r \cdot x + s \cdot x$.
(*iii*) $r \cdot (s \cdot x) = (r \cdot s) \cdot x$.
(*iv*) $1 \cdot x = x$.

We call the map $\cdot : R \times M \to M$ an *R-action*. Note that there are two $\cdot$'s in the properties above, the R-action of R on M and the ring multiplication on R. Note that (*i*) and (*ii*) are the analogues of the distributive law and (iii) the analogue for the associative law for rings. Condition (iv) says the identity of the ring acts like an identity. You should also compare this to G-sets where G is an additive group (or a monoid). We drop the $\cdot$ of the R-action whenever it is clear.

A subgroup N in M is called a *submodule* if it is an R-module under $\cdot|_{R\times N}$.

Remarks 38.2. 1. One defines *right R-modules* in the obvious way.

2. As usual, a submodule of an R-module M should be a subset of M that is an R-module under the inclusion map. It will be so after we define the appropriate notion of homomorphism. Of course, you should already know what it is, as we are generalizing vector space theory.
3. Some authors call M an R-module without assuming (iv) and call an M also satisfying (iv) a *unitary* R-module. (If R is a rng, we will not assume (iv) in our definition.)
4. Let M be an R-module, x an element in M, and r an element in R. Then
   (a) $0_R x = 0_M$. (We write $0x = 0$.)
   (b) $r0_M = 0_M$. (We write $r0 = 0$.)
   (c) $(-r)x = r(-x) = -(rx)$.
   (d) $m(rx) = r(mx)$ for all integers m.
   We leave this an an exercise.

As usual we start with a number of examples.

Examples 38.3. Let R be a ring and M an R-module.

1. If R is a field (or a division ring), then an R-module is the same as a vector space over R.
2. If R is the ring of integers, then a $\mathbb{Z}$-module is just an additive group. In particular, module theory generalizes both vector space theory and abelian group theory.
3. $0 := \{0\}$ and M are submodules of M.
4. R is an R-module under the addition in R with the R-action given by the multiplication in R.
5. Let $\mathfrak{A}$ be a left ideal in R. Then $\mathfrak{A} + \mathfrak{A} \subset \mathfrak{A}$ and $R\mathfrak{A} \subset \mathfrak{A}$, so $\mathfrak{A}$ is a submodule of R. Indeed, a submodule of R is the same thing as a left ideal. So the concept of an R-module also generalizes the notion of left ideal.
6. $R[t]$ is an R-module. More generally, if R is a subring of S then S is an R-module by restricting the ring multiplication on S to scalar multiplication by R, i.e., restricting $\cdot|_{S\times S}$ to $\cdot|_{R\times S}$.
7. Let S be a nonempty set and X an R-module. Set
$$M := \{f : S \to X \mid f \text{ a map}\},$$
then M is an R-module under the following operations: If f and g lie in M and r in R, then for all s in S define
$$(f+g)(s) := f(s) + g(s)$$
$$r(f(s)) : rf(s)$$

and 0_M defined by $0_M(s) = 0_X$.

8. Let N be a submodule of M. Then the factor group M/N ($N \triangleleft M$ is automatic as M is abelian) becomes an R-module by
$$\cdot : R \times M/N \to M/N \text{ defined by } r(m+N) = rm + N$$
for all $r \in R$ and for all $m \in M$. It is called the *quotient* or *factor module* of M by N.
9. Let N be an abelian group. Write it additively, i.e., N is a $\mathbb{Z}$-module. Then
$$\mathrm{End}(N) = \mathrm{End}_{\mathbb{Z}}(N) := \{\varphi : N \to N \mid \varphi \text{ a group homomorphism}\}$$
is a ring via
(*i*) $(\varphi+\psi)(x) = \varphi(x)+\psi(x)$ (the usual addition of functions into an additive group)
(*ii*) $\varphi \circ \psi(x) = \varphi\big(\psi(x)\big)$ (composition of functions).
for all φ and ψ in $\mathrm{End}(N)$ and for all x in N (as $nx = \underbrace{x + \cdots + x}_{n}$ for all positive integers n); and N is an $\mathrm{End}(N)$-module by *evaluation*, i.e., $\varphi \cdot x := \varphi(x)$ for all $\varphi \in \mathrm{End}(N)$ and $x \in N$.
10. Let F be a field, V a vector space over F. Then
$$\mathrm{End}_F(V) = \{T : V \to V \mid T \text{ a linear operator}\}$$
is a subring of $\mathrm{End}_{\mathbb{Z}}(V)$ and V is an $\mathrm{End}_F(V)$-module via *evaluation*, i.e., $T \cdot x = T(x)$ for all $T \in \mathrm{End}_F(V)$ and $x \in V$.
[Can you define the appropriate notion of homomorphism $f : M \to N$ between modules, then replace $\mathrm{End}_F(V)$ by the appropriate $\mathrm{End}_R(M)$? The ring $\mathrm{End}_R(M)$ is called the *endomorphism ring* of M and elements in it are called *R-endomorphisms*. Then M becomes an $\mathrm{End}_R(M)$-module under evaluation, $f \cdot x := f(x)$ for all $x \in M$ and $f \in \mathrm{End}_R(M)$.]
11. Let $\varphi : R \to S$ be a ring homomorphism and N an S-module. Then N becomes an R-module via the *pullback* defined by
$$\cdot : R \times N \to N \text{ given by } r \cdot x = \varphi(r) \cdot x$$
for all r in R and x in N.
12. Let V be a vector space over the field F and $T \in \mathrm{End}_F(V)$ a fixed linear operator ($=$ endomorphism). Define a ring homomorphism
$$\varphi : F[t] \to \mathrm{End}_F(V) \text{ given by the evaluation } f \to f(T).$$
So we have $\sum a_i t^i \mapsto \sum a_i T^i$. [What is $\ker \varphi$? What can you say if V is a finite dimensional vector space over F?] Then V is an $F[t]$-module via the pullback $f \cdot v := f(T)(v)$ for all $f \in F[t]$ and for all $v \in V$. Let W be a subspace of V. Then W is a submodule of the $F[t]$-module V if and only if W is a *T-invariant subspace*, i.e., $T(W) \subset W$. This is an example that will be used in Section 45.
13. Let $\mathfrak{A} \subset R$ be a left ideal. Then
$$\mathfrak{A}M := \{\sum_{x \in M} a_x x \mid a_x \in \mathfrak{A} \text{ for all } x \in M\} \subset M$$

is an R-module. Note that an infinite sum of elements in an additive group does not make sense, so when we write $\sum_{x\in M} a_x x$, we must have a_x is zero *for almost all* x, i.e., $a_x \neq 0$ for only finitely many x.

14. Let M be an R-module and M_i submodules of M for $i \in I$. Then
    (*i*) $\bigcap_I M_i$ and
    (*ii*) $\sum_I M_i := \{\sum m_i \mid m_i \in M_i,\ m_i = 0 \text{ for almost all } i \in I\}$
    are submodules of M. If the M_i also satisfy $M_i \cap \sum_{I\setminus\{i\}} M_j = 0$ for all $i \in I$, we call $\sum_I M_i$ the *internal direct sum* of the M_i and write it as $\bigoplus_I M_i$.
15. Let M_i, $i \in I$, be R-modules. Recall $(m_i)_I$ is the notation for elements in the cartesian product $\bigtimes_I M_i$. Set
    (*i*) $\coprod_I M_i := \{(m_i)_I \mid m_i \in M_i,\ m_i = 0 \text{ for almost all } i \in I\}$, an R-module under componentwise operations. It is called the *(external) direct sum* or *coproduct* of the M_i.
    (*ii*) $\prod_I M_i := \{(m_i)_I \mid m_i \in M_i\}$, an R-module under componentwise operations. It is called the *(external) direct product* (or just *product*) of the M_i.
    We have $\coprod_I M_i$ is a submodule of $\prod_I M_i$ with $\coprod_I M_i = \prod_I M_i$ if and only if I is finite or almost all $M_i = 0$.
16. Let X be a subset of M. Define
    $$\langle X\rangle := \sum_X Rx,$$
    called the submodule *generated* by X. Note that, as for vector spaces, this is equivalent to $\langle X\rangle = \bigcap_{X\subset N\subset M} N$, where N runs over all submodules of M containing X. If there exists a finite subset Y in M such that $M = \langle Y\rangle$ we say that M is *finitely generated* or simply *fg*. If $Y = \{y_1, \dots, y_n\}$, we write $\langle y_1, \dots, y_n\rangle$ for $\langle Y\rangle$. If there exists a $y \in M$ such that $M = \langle y\rangle$, we say that M is R-*cyclic*. (So a cyclic group is the same thing as a cyclic $\mathbb{Z}$-module.)

Definition 38.4. A map $f : M \to N$ of R-modules is called an R-*homomorphism* (respectively, R-*monomorphism,* R-*epimorphism,* R-*isomorphism*) if f is R-*linear*, i.e.,

$$f(rx + y) = rf(x) + f(y)$$

for all $r \in R$ and $x, y \in M$ (respectively, and injective, and surjective, and bijective with inverse R-linear). As usual we also use the abbreviations of monic and epic. Also, as one would surmise, f is an R-isomorphism if and only if it is a bijective R-homomorphism.

We say that two R-modules M and N are R-*isomorphic* and write $M \cong N$ if there exists an R-isomorphism $g : M \to N$, which we also write as $g : M \xrightarrow{\sim} N$.

Remark 38.5. As is true with groups and rings an R-homomorphism $f : M \to N$ is an R-monomorphism if and only if given any R-homomorphisms $g_1, g_2 : L \to M$ with compositions satisfying $f \circ g_1 = f \circ g_2$, then $g_1 = g_2$; and as with groups (but not rings) f is an R-epimorphism if and only if given any R-homomorphisms $h_1, h_2 : N \to L$ with compositions satisfying $h_1 \circ f = h_2 \circ f$, then $h_1 = h_2$. (Cf. Exercise 1.13(7) and (8).)

Examples 38.6. 1. A $\mathbb{Z}$-homomorphism is the same thing as an abelian group homomorphism.

2. A linear transformation of F-vector spaces is the same thing as an F-homomorphism of F-modules.
3. Let M and N be R-modules with $N \subset M$. Then N is a submodule of M if and only if the inclusion is an R-homomorphism.
4. Let R be a commutative ring and r an element of R. Then
$$\lambda_r : R \to R \text{ given by } x \mapsto rx$$
is an R-homomorphism (but not a ring homomorphism). More generally, if M is an R-module, then
$$\lambda_r : M \to M \text{ given by } x \mapsto rx$$
is an R-homomorphism.
5. Let N be a submodule of M. Then the canonical map
$$^{-} : M \to M/N \text{ given by } x \mapsto \overline{x} = x + N$$
is an R-epimorphism.
6. Let R be a ring and $\mathfrak{A} < R$ be a (2-sided) ideal, $^{-} : R \to R/\mathfrak{A}$ the canonical ring epimorphism. If M is an R-module then the R-module $M/\mathfrak{A}M$ is also an $R/\mathfrak{A}$-module by $\overline{r}(x + \mathfrak{A}M) = rx + \mathfrak{A}M$ for all $r \in R$ and $x \in M$, so the R- and $(R/\mathfrak{A})$-actions on $M/\mathfrak{A}M$ are *compatible.*, i.e., $r(x + \mathfrak{A}M) = \overline{r}(x + \mathfrak{A}M)$.
7. Let $f : M \to N$ be an R-homomorphism of R-modules. Then the additive subgroups $\ker f \subset M$ and $\operatorname{im} f \subset N$ are submodules. More generally, if N' is a submodule of N, then $\ker f \subset f^{-1}(N') \subset M$ is a submodule and if M' is a submodule of M, then $f(M')$ is a submodule of N. This makes the study of R-modules and R-homomorphisms very nice. We do not need to worry about some of the special properties needed with groups and rings.
8. If M is an R-module, then
$$\operatorname{End}_R(M) := \{f : M \to M \mid f \text{ an } R\text{-homomorphism}\}$$
is a ring under the $+$ of functions and composition (as asked for before), called the *endomorphism ring* of the R-module M. Elements of $\operatorname{End}_R(M)$ are called R-*endomorphisms*. Its unit group is $\operatorname{Aut}_F(M)$, the group of R-*automorphisms of* M, where an R-*automorphism* is an R-isomorphism $M \to M$. The ring $\operatorname{End}(M) = \operatorname{End}_{\mathbb{Z}}(M)$ contains $\operatorname{End}_R(M)$ as a subring. Note that if R is a field, then a linear operator $T : V \to V$ of a vector space V over R is the same thing as an R-endomorphism of V.
9. If R is a commutative ring and M, N are R-modules, then
$$\operatorname{Hom}_R(M, N) := \{f : M \to N \mid f \text{ an } R\text{-homomorphism}\}$$
is an R-module with the usual $+$ for functions and the R-action $\cdot$ given by $r \cdot f : x \mapsto rf(x)$. [If R is not commutative, $\operatorname{Hom}_R(M, N)$ is not an R-module and we usually just view it as an abelian group. Is it an S-module for some subrings S of R?]

We have analogous isomorphism theorems and a correspondence theorem for modules as we had before with the same proof. We state them for convenience.

Theorem 38.7. (First Isomorphism Theorem) *Let $f : M \to N$ be an R-homomorphism. Then we have a commutative diagram of R-modules and R-homomorphisms*

$$\begin{tikzcd} M \arrow[r, "f"] \arrow[d, "-"'] & N \\ M/\ker f \arrow[r, "\overline{f}"'] & \operatorname{im} f \arrow[u, "\text{inc}"'] \end{tikzcd}$$

with $^{-}$, the canonical map, an R-epimorphism, $\overline{f}$, the map given by $\overline{x} \mapsto f(x)$, an R-isomorphism between $M/\ker f$ and $\operatorname{im} f$, and the inclusion map inc *an R-monomorphism.*

Theorem 38.8. (Second Isomorphism Theorem) *Let M be an R-module and A and N submodules of M. Then $A/(A \cap N) \cong (A + N)/N$.*

Theorem 38.9. (Third Isomorphism Theorem) *Let M be an R-module with submodules A and N satisfying $A \subset N$. Then we have $M/N \cong (M/A)/(N/A)$.*

Theorem 38.10. (Correspondence Principle) *Let $f : M \to N$ be an R-epimorphism of R-modules. Then*

$$\{A \mid \ker f \subset A \subset M \text{ a submodule}\} \longrightarrow \{B \mid B \subset N \text{ a submodule}\}$$

given by $A \to f(A)$ is an order preserving bijection.

Next we introduce the generalization of a zero divisor in a ring.

Definition 38.11. Let M be an R-module and m an element of M. The *annihilator* of m is defined to be the set

$$\operatorname{ann}_R m := \{r \in R \mid rm = 0\}.$$

More generally, we shall let

$$\operatorname{ann}_R(M) := \{r \in R \mid rm = 0 \text{ for all } m \in M\}.$$

denote the *annihilator of M*.

The annihilator has the following properties:

Lemma 38.12. *Let M be an R-module and m, m' elements in M. Then*

(1) $\operatorname{ann}_R m$ *is a left ideal.*
(2) $\operatorname{ann}_R(M) \subset R$ *is an ideal.*
(3) $\rho_m : R \to M$ *given by* $r \mapsto rm$ *is an R-homomorphism and satisfies* $\ker \rho_m = \operatorname{ann}_R m$.
(4) If ρ_m is the R-homomorphism in (3), then ρ_m induces an R-isomorphism $\overline{\rho_m} : R/\operatorname{ann}_R m \xrightarrow{\sim} Rm$.
(5) If R is a commutative ring and $Rm \subset Rm'$, then $\operatorname{ann}_R m \supset \operatorname{ann}_R m'$. In particular, $\operatorname{ann}_R m$ is independent of the generator for Rm.

PROOF. (1) and (2) follow immediately.
(3) follows from the First Isomorphism Theorem for modules.
(4): Suppose that we know that $m = am'$ for some $a \in R$. If $rm' = 0$, then $ram' = arm' = 0$. $\square$

We can now characterize cyclic R-modules.

Corollary 38.13. *Let M be an R-module. Then M is a cyclic R-module if and only if there exists a left ideal $\mathfrak{A}$ in R satisfying $M \cong R/\mathfrak{A}$.*

PROOF. $R/\mathfrak{A} = \langle 1 + \mathfrak{A} \rangle = R(1 + \mathfrak{A})$ is cyclic, so this follows by the lemma. $\square$

If R is a commutative ring, we can say more.

Proposition 38.14. *Let R be a commutative ring, M a cyclic R-module. Then there exists a unique ideal $\mathfrak{A}$ in R such that $M \cong R/\mathfrak{A}$. Moreover, if $M = Rm$, then $\mathfrak{A} = \operatorname{ann}_R m$. In particular, if R is a PID, there exists an element a in R, unique up to units, satisfying $M \cong R/(a)$, and, if in addition, $M = Rm$, then $(a) = \operatorname{ann}_R m$.*

PROOF. We already know if $M = Rm$, then $M \cong R/\operatorname{ann}_R m$. Suppose that we have an R-isomorphism $f : R/\mathfrak{A} \to M$ for some ideal $\mathfrak{A}$ of R. Let $m = f(1 + \mathfrak{A})$. We have

$$f(r + \mathfrak{A}) = f\big(r(1 + \mathfrak{A})\big) = rf(1 + \mathfrak{A}) = rm$$

for all $r \in R$. In particular, $M = Rm$. Let $\rho_m : R \to M$ be the R-epimorphism given by $r \mapsto rm$. Then we have a commutative diagram

$$\begin{tikzcd} R \arrow[r, "\rho_m"] \arrow[d, "-"'] & M. \\ R/\mathfrak{A}. \arrow[ur, "f"'] & \end{tikzcd}$$

Thus

$$\operatorname{ann}_R m = \ker \rho_m = \ker(f \circ {}^-) = \ker {}^- = \mathfrak{A},$$

as f is monic and $f(\overline{r}) = \rho_m(r)$, i.e., $f = \overline{\rho_m}$. By Lemma 38.12(4), $\operatorname{ann}_R m$ is independent of the generator of M when R is commutative. If follows that $\mathfrak{A}$ is unique. Lemma 38.12(4) also implies the statements about PIDs. $\square$

We now set up a lot of useful notation. We write maps $f : A \to B$ as $A \xrightarrow{f} B$ as we have done in commutative diagrams.

Definition 38.15. Let $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ be R-homomorphisms of R-modules. We say that

$$A \xrightarrow{f} B \xrightarrow{g} C$$

is a *zero sequence* if $g \circ f = 0$, equivalently, $\operatorname{im} f \subset \ker g$ and is an *exact sequence* if $\operatorname{im} f = \ker g$. A longer *sequence*

$$\cdots \xrightarrow{f_{n+2}} A_{n+1} \xrightarrow{f_{n+1}} A_n \xrightarrow{f_n} A_{n-1} \xrightarrow{f_{n-1}} \cdots$$

of R-modules and R-homomorphisms is called a *zero sequence* or a *chain complex* (respectively, an *exact sequence* or an *acyclic complex*) if $f_n f_{n+1} = 0$ (respectively, $\operatorname{im} f_{n+1} =$

$\ker f_n$) for all $n, n+1$ occurring. The special case of an exact sequence of R-modules and R-homomorphisms of the form

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is called a *short exact sequence*.

Examples 38.16. Let $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ be R-homomorphisms of R-modules.

1. The R-homomorphism f is monic if and only if $0 \to A \xrightarrow{f} B$ is exact. [Note that $0 \to A$ must take $0 \mapsto 0_A$ as it is an R-homomorphism.]
2. The R-homomorphism g is epic if and only if $B \xrightarrow{g} C \to 0$ is exact.
3. If $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is a short exact exact sequence, then f is monic, $\operatorname{im} f = \ker g$, and g is epic.
4. The sequence $0 \to \ker f \to A \xrightarrow{f} B$ is exact (where the second map is the inclusion) and $0 \to \ker f \to A \xrightarrow{f} \operatorname{im} f \to 0$ is a short exact sequence (writing the same f after changing the target).
5. The R-homomorphism f induces an R-monomorphism $A/\ker f \xrightarrow{\overline{f}} B$ by the First Isomorphism Theorem. Therefore, the First Isomorphism Theorem is just that we have a commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & \ker f \arrow[r] \arrow[d, equal] & A \arrow[r, "-"] \arrow[d, equal] & A/\ker f \arrow[r] \arrow[d, "\overline{f}"] & 0 \\ 0 \arrow[r] & \ker f \arrow[r] & A \arrow[r, "f"] & \operatorname{im} f \arrow[r] & 0. \end{tikzcd}$$

   with exact rows and with the right hand vertical arrow $\overline{f}$ an R-isomorphism. Can you write the Third Isomorphism Theorem using such sequences?
6. As the factor R-module of M by N exists for any submodule N of an R-module M, besides the image and kernel of the R-homomorphism $B \xrightarrow{g} C$, we can define another R-module, the *cokernel* of g defined by $\operatorname{coker} g := C/\operatorname{im} g$. Then we have a short exact sequence

$$0 \to \operatorname{im} g \to C \xrightarrow{-} \operatorname{coker} g \to 0,$$

   where the second map is the inclusion. So we have
   (a) g is injective if and only if $\ker g = 0$.
   (b) g is surjective if and only if $\operatorname{coker} g = 0$.
7. Let A and B be R-modules. Then the sequence

$$0 \to A \xrightarrow{\iota_A} A \coprod B \xrightarrow{\pi_B} B \to 0$$

   with $\iota_A(a) = (a, 0)$ and $\pi_B(a, b) = b$ is a short exact sequence.
8. Let A_i, B_i, and C_i be R-modules for $i \in I$. If

$$0 \to A_i \xrightarrow{f_i} B_i \xrightarrow{g_i} C_i \to 0$$

is a short exact sequence for all $i \in I$, then

$$0 \to \coprod_I A_i \xrightarrow{\coprod_I f_i} \coprod_I B_i \xrightarrow{\coprod_I g_i} \coprod_I C_i \to 0$$

and

$$0 \to \prod_I A_i \xrightarrow{\prod_I f_i} \prod_I B_i \xrightarrow{\prod_I g_i} \prod_I C_i \to 0$$

are exact. What are the maps $\coprod_I f_i$, $\coprod_I g_i$, and $\prod_I f_i$, $\prod_I g_i$?

Lemma 38.17. (Five Lemma) *Suppose the following is a commutative diagram of R-modules and R-homomorphisms with exact rows:*

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "h_A"'] & B \arrow[r, "g"] \arrow[d, "h_B"'] & C \arrow[r] \arrow[d, "h_C"'] & 0 \\
0 \arrow[r] & A' \arrow[r, "f'"'] & B' \arrow[r, "g'"'] & C' \arrow[r] & 0.
\end{tikzcd}$$

If two of h_A, h_B, h_C are R-isomorphisms, then they all are.

The proof of this lemma is called *diagram chasing*, and we leave it as an exercise.

Exercises 38.18.

1. Prove Remark 38.2(4)
2. Prove Remark 38.5.
3. Show an R-homomorphism $f : M \to N$ is an isomorphism if and only if f is bijective.
4. Let M be a *simple* R-module, i.e., M has no proper nonzero submodules. Prove that $\operatorname{End}_R(M)$ is a division ring.
5. Let M be an R-module and M_i submodules of M for $i \in I$. Show that there is always an R-homomorphism $f : \coprod_I M_i \to \sum_I M_i$ and this homomorphism is an R-isomorphism if and only if $\sum_I M_i$ is an internal direct sum.
6. Let M be an R-module and $M_1, \ldots, M_n$ be R-submodules of M. Show that $M = \bigoplus_{i=1}^n M_i$ if and only if there exist R-homomorphisms $\iota_i : M_i \to M$ and $\pi_i : M \to M_i$, $i = 1, \ldots, n$, satisfying all of the following:
   (*i*) $\pi_i \iota_i = 1_{M_i}$ for $i = 1, \ldots, n$.
   (*ii*) $\pi_j \iota_i = 0$ for $i \neq j$.
   (*iii*) $\iota_1 \pi_1 + \cdots + \iota_n \pi_n = 1_M$.
7. Let m be a positive integer. Determine the abelian groups ($\mathbb{Z}$-modules) $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/m\mathbb{Z})$ and $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z})$ up to isomorphism.
8. Let m and n be positive integers with greatest common divisor d. Show that we have an isomorphism $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z}$.
9. (Universal Property of Direct Sums) Let M_i, $i \in I$, be R-modules. Show that the (external) direct sum $\coprod_I M_i$ satisfies the following universal property relative to the R-homomorphisms $\iota_j : M_j \to \coprod_I M_i$ defined by $m \mapsto \{\delta_{ij} m\}_{i \in I}$ for all $j \in I$: Given an R-module M and R-homomorphisms $f_j : M_j \to M$ for all $j \in I$, there exist a unique R-homomorphism $g : \coprod_I M_i \to M$ satisfying $f_j = g \circ \iota_j$ for all $j \in I$.

10. (Universal Property of Direct Products) Let M_i, $i \in I$, be R-modules. Show that the (external) direct product $\prod_I M_i$ satisfies the following universal property relative to the R-homomorphisms $\pi_j : \prod_I M_i \to M_j$ defined by $\{m_i\}_I \mapsto m_j$ for all $j \in I$: Given an R-module M and R-homomorphisms $g_j : M \to M_j$ for all $j \in I$, there exist a unique R-homomorphism $h : M \to \prod_I M$ satisfying $g_j = \pi_j \circ h$ for all $j \in I$.
11. Write the Third Isomorphism Theorem using exact sequences as in Example 38.16(5).
12. A short exact sequence of R-modules

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is called *split* if one of the following three equivalent conditions holds:

(i) There exists an R-homomorphism $f' : B \to A$ such that $f'f = 1_A$. We say that f is a *split monomorphism*.
(ii) $\operatorname{im} f$ is a *direct summand* of B, i.e., $B = \operatorname{im} f \oplus D$ some R-module D.
(iii) There exists an R-homomorphism $g' : C \to B$ such that $gg' = 1_C$. We say that g is a *split epimorphism*.

Prove that these conditions are equivalent.
13. Prove the Five Lemma 38.17.
14. Prove the full version of the Five Lemma: Suppose that

$$\begin{array}{ccccccccc}
A & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & C & \xrightarrow{\gamma} & D & \xrightarrow{\delta} & E \\
{\scriptstyle h_A}\downarrow & & {\scriptstyle h_B}\downarrow & & {\scriptstyle h_C}\downarrow & & {\scriptstyle h_D}\downarrow & & {\scriptstyle h_E}\downarrow \\
A' & \xrightarrow{\alpha'} & B' & \xrightarrow{\beta'} & C & \xrightarrow{\gamma'} & D & \xrightarrow{\delta'} & E.
\end{array}$$

is a commutative diagram of R-modules and R-homomorphisms with exact rows. Then show the following:

(i) If h_A is an epimorphism and h_B and h_D are monomorphisms, then h_C is a monomorphism.

(ii) If h_E is a monomorphism and h_B and h_D are epimorphisms, then h_C is an epimorphism.

is a commutative diagram of R-modules and R-homomorphisms with exact rows.
15. Let R be a commutative ring and M, N R-modules. Recall that $\operatorname{Hom}_R(M, N)$ is an R-module by Example 38.6(9). If $h : A \to B$ is an R-homomorphism of R-modules, define

$$h_* : \operatorname{Hom}_R(N, A) \to \operatorname{Hom}_R(N, B) \text{ by } f \mapsto h \circ f \text{ and}$$
$$h^* : \operatorname{Hom}_R(B, N) \to \operatorname{Hom}_R(A, N) \text{ by } f \mapsto f \circ h.$$

Show that these are R-homomorphisms and if

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(N, A) \xrightarrow{f_*} \mathrm{Hom}_R(N, B) \xrightarrow{g_*} \mathrm{Hom}_R(N, C) \quad \text{and}$$
$$0 \to \mathrm{Hom}_R(C, N) \xrightarrow{g^*} \mathrm{Hom}_R(B, N) \xrightarrow{f^*} \mathrm{Hom}_R(A, N)$$

are exact. (Note the missing 0 on the right.) [If R is not commutative, then your proof will only show that $h*$ and h^* are abelian group homomorphisms and the sequences are exact sequences of abelian groups.]

16. Let R be a commutative ring and $M_i, i \in I$, and N be R-modules. Using the Universal Property of Direct Sums (cf. Exercise 38.18(9)), show if $M_i, i \in I$, and N are R-modules, then there exists an R-isomorphism $\mathrm{Hom}_R(\coprod_I M_i, N) \cong \prod_I \mathrm{Hom}_R(M_i, N)$. [If R is not commutative, then the isomorphism is only an abelian group isomorphism.]
17. Let R be a commutative ring and M and $N_i, i \in I$ be R-modules. Using the Universal Property of Direct Products (cf. Exercise 38.18(10)), show if $M_i, i \in I$ and N are R-modules, then there exists an R-isomorphism $\mathrm{Hom}_R(M_i, \prod_I N_i) \cong \prod_I \mathrm{Hom}_R(M_i, N)$. (Remember that if I is finite, then the coproduct and product are equal. [If R is not commutative, then the isomorphism is only an abelian group isomorphism.]
18. Let R be a (commutative) ring, M and N R-modules. Then $\mathrm{Hom}_R(M, \ \ )$ and $\mathrm{Hom}_R(\ \ , N)$ take split exact sequences to split exact sequences.
19. Let Q be an R-module. Then Q is called *R-injective* if given any R-monomorphism $f : A \to B$ and R-homomorphism $g : A \to Q$, there exists an R-homomorphism $h : B \to Q$ such that the diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, "h"] \\ Q & \end{tikzcd}$$

commutes. Show that Q is an R-injective if and only if, whenever

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(C, Q) \xrightarrow{g^*} \mathrm{Hom}_R(B, Q) \xrightarrow{f^*} \mathrm{Hom}_R(A, Q) \to 0$$

is exact. (Cf. Exercise (123).)
20. Let Q_i, $i \in I$ be R-modules. Show that $\prod_I Q_i$ is R-injective if and only Q_i is R-injective for all $i \in I$. In particular, if I is finite, then $\coprod_I Q_i$ is R-injective if and only Q_i is R-injective for all $i \in I$ (since then the coproduct and product of the Q_I's are the same).
21. Let A be an R-module. Show that A is R-injective if and only if any short exact sequence of R-modules of the form $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ splits.

22. (Baer Criterion) Let Q be an R-module. Show that Q is an R-injective if and only if given any ideal $\mathfrak{A}$ in R and an R-homomorphism $g : \mathfrak{A} \to Q$, there exists an R-homomorphism $h : R \to Q$ such that the diagram

$$\begin{tikzcd} \mathfrak{A} \arrow[r, hook, "\text{inc}"] \arrow[d, "g"'] & R \arrow[dl, "h"] \\ Q & \end{tikzcd}$$

commutes where inc is the inclusion.

23. Let R be a domain with F its quotient field. Use the Baer Criterion to show that F is R-injective.
24. Show that a divisible abelian group is an injective $\mathbb{Z}$-module.

39. Free Modules

Vector spaces have nice properties because of the concept of linear independence and the fact that one can divide by any nonzero scalar. In general, R-modules have neither of these two properties. There is little we can do about the lack of units in an arbitrary ring, but we can look at R-modules that are spanned by linearly independent sets. Such modules are called *free* R-modules. Though sparse, these modules are quite nice, in fact, just as a vector space is isomorphic to a external direct sum of copies of the underlying field, the analogous statement is true for free R-modules. The first result one would then expect is that the notion of dimension makes sense, i.e., any two linearly independent spanning set of a free module should have the same cardinality. Unfortunately, this is false in general. The problem does vanish, however, if the ring R is commutative. We shall leave the proof of this fact as an exercise. This is fortuitous as these are the rings we are studying in detail.

Definition 39.1. A nonzero R-module M is called a *free* R-module (or R-*free*) if there exists a *basis* for M, i.e., a subset $\mathcal{B}$ of M satisfying:

(i) $M = \langle \mathcal{B} \rangle$, i.e., $\mathcal{B}$ *generates* or *spans* M.

(ii) $\mathcal{B}$ is *linearly independent*, i.e., if $0 = \sum_{\mathcal{B}} r_x x$ (so $r_x = 0$ for almost all $x \in \mathcal{B}$), then $r_x = 0$ for all $x \in \mathcal{B}$. Equivalently, $M = \bigoplus_{\mathcal{B}} Rx$ and if $rx = 0$ with $x \in \mathcal{B}$, $r \in R$, then $r = 0$.

We shall also call the trivial module a free R-module.

Of course, a set $\mathcal{D}$ in M is called *linearly dependent* if it is not linearly independent. Equivalently, there exist $x_1, \dots, x_n \in \mathcal{D}$, some n, and $r_1, \dots, r_n$ in R not all zero such that $r_1 x_1 + \cdots + r_n x_n = 0$.

If M is a free R-module on basis $\mathcal{B}$, the definition says that $\mathcal{B}$ generates M and there are no nontrivial relations on these generators. Contrast this with a finite (abelian) group say of order n which says that every element x satisfies the relation $x^n = e$. In particular, this is true of any element in any generating set.

Examples 39.2. Let R be a ring and M be a nontrivial R-module.

1. R is R-free on basis $\{u\}$ with u a unit in R.
2. Suppose that M is R-cyclic. Then M is R-free if there exists an $x \in M$ satisfying $M = Rx$ and $\operatorname{ann}_R x = 0$. It follows that M is R-free if and only if $M \cong R$.
3. If M is a finitely generated R-module, then M is R-free if and only if there exists a finite basis for M if and only if
$$M \cong \underbrace{R \coprod \cdots \coprod R}_{n} = R^n \text{ for some positive integer } n. \text{ (Why?)}$$
4. If R is a field (or a division ring), then every R-module is free and any two bases have the same cardinality.
5. If R is commutative and M is R-cyclic, then M is R-free if and only if $\operatorname{ann}_R m = 0$ whenever $M = Rm$, i.e., for any generator of M.

   To show this one needs to use the first two parts of Exercise 39.12(7) (which establishes that all bases for a finitely generated free module over a commutative ring have the same cardinality). If we know this and M is R-free, then say every basis for M has say n elements. Let $\mathcal{B}$ be a basis for M and $\mathfrak{m}$ be a maximal ideal in R. Then this exercise would allow us to prove the following sequence of isomorphisms:
$$\begin{aligned} Rm/\mathfrak{m}m = M/\mathfrak{m}M &= (\bigoplus_{\mathfrak{B}} Rx)/\big(\mathfrak{m}(\bigoplus_{\mathfrak{B}} Rx)\big) \\ &\cong R^n/\mathfrak{m}R^n \cong (R/\mathfrak{m})^n \end{aligned}$$
   as R- and as $R/\mathfrak{m}$-modules. As $R/\mathfrak{m}$ is a field, this is an isomorphism of vector spaces over $R/\mathfrak{m}$. As $m + \mathfrak{m}m$ generates $Rm/\mathfrak{m}m$, by (3), we must have $n = 1$.
6.
7. If R is not commutative, it can happen that $R^m \cong R^n$ for different positive integers, i.e., different bases can have different cardinalities. For example, let M be a free R-module on a countable (not finite) basis (e.g., $M = R[t]$ (cf (9) below). Let S be the endomorphism ring $\operatorname{End}_R(M)$. Then it can be shown that S^n is S-free and $S^n \cong S^m$ for all positive integers n and m.
8. Suppose that M is R-free with basis $\mathcal{B}$ and $x \in \mathcal{B}$. Then the map $\rho_x : R \to M$ given by $r \mapsto rx$ is an R-monomorphism. In particular, the cardinality of M is at least that of R. This gives another proof that no nontrivial finite abelian group is $\mathbb{Z}$-free.
9. Let M_i be free R-modules for $i \in I$. Then $\coprod_I M_i$ is R-free. [If $\mathcal{B}_i$ is a basis for M_i for each $i \in I$, what is a basis for $\coprod_I M_i$?] For example $\coprod_{i=1}^{\infty} \mathbb{Z}$ is $\mathbb{Z}$-free on the *standard basis*
$$\mathcal{S} = \{e_i = (0, \dots, \underset{i}{1}, 0, \dots) \mid i \geq 1\}.$$
   [Warning: It turns out that $\prod_{i=0}^{\infty} \mathbb{Z}$ is not $\mathbb{Z}$-free.]
10. $\mathbb{Q}$ is not $\mathbb{Z}$-free.
11. $R[t]$ is R-free on basis $\mathcal{B} = \{t^i \mid i \geq 0\}$. One can now define the ring $R[t]$ to be the free R-module on basis $\mathcal{B}$ with multiplication induced by $t^i \cdot t^j := t^{i+j}$ and $rt = tr$ for all $r \in R$ (and extend linearly). More generally, $R[t_1, \dots, t_n]$ is the free R-module on basis $\{t_1^{i_1} \cdots t_n^{i_n} \mid i_j \geq 0 \text{ for all } i_j\}$.

12. Let R_0 be a ring and $R = R_0[t_1, t_2]$. The R-module $M = Rt_1 + Rt_2$ is a submodule of R as it is a left ideal. Then M is not R-free. [Proof?]
[This shows that a submodule of a free R-module may not be free.]
13. Let $n \mid m$ in $\mathbb{Z}^+$ and $^- : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ be the ring epimorphism given by $r + m\mathbb{Z} \mapsto r + n\mathbb{Z}$. Then $\mathbb{Z}/n\mathbb{Z}$ is a $(\mathbb{Z}/m\mathbb{Z})$-module via the pullback, e.g., $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are $(\mathbb{Z}/6\mathbb{Z})$-modules. By element count neither $\mathbb{Z}/2\mathbb{Z}$ nor $\mathbb{Z}/3\mathbb{Z}$ are $\mathbb{Z}/6\mathbb{Z}$-free, but $\mathbb{Z}/2\mathbb{Z} \amalg \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$ as $(\mathbb{Z}/6\mathbb{Z})$-modules (why?), so a *direct summand* of a free module may not be free. (A submodule N is a *direct summand* of an R-module M if $M = N \oplus M'$ for some submodule M'.)
14. If $M = \bigoplus_I Rx_i$ is a free R-module on basis $\mathcal{B} = \{x_i \mid i \in I\}$ and m is an element in M, then there exist unique $r_i \in R$, $i \in I$, almost all zero such that $m = \sum_I r_i x_i$. The r_i are called the *coordinates* of m *relative to the basis* $\mathcal{B}$ and r_i is called the *coordinate* of m on x_i.
15. The set $\{2\}$ is linearly independent in the free $\mathbb{Z}$-module $\mathbb{Z}$, but it is not a basis as $2\mathbb{Z} < \mathbb{Z}$. Indeed any two distinct elements in $\mathbb{Z}$ form a linearly dependent set. This shows that, in general, not every linearly independent set in a free R-module need be part of a basis.
16. Let R be a domain and M a free R-module. If z is a nonzero element in M satisfying $rz = 0$ for some $r \in R$, then $r = 0$ (for if $\mathcal{B}$ is a basis, then $z = \sum_{\mathcal{B}} r_x x$).

Universal properties usually produce maps satisfying certain properties together with a uniqueness statement. This is very useful, as one of the more difficult problems is to show the existence of a map, or to prove that a potential explicit map is well-defined. Usually the problem arises as we know generators but cannot show that the putative map respects the relations among these generators. As a basis for a free module does not satisfy any nontrivial relations, this causes no problems and free modules satisfy the following universal property:

Theorem 39.3. (Universal Property of Free Modules) *Let $\mathcal{B} = \{x_i\}_I$ be a basis for a free R-module M. If N is an R-module and y_i, $i \in I$, elements in N (not necessarily distinct), then there exists a unique R-homomorphism $f : M \to N$ such that $x_i \mapsto y_i$ for all $i \in I$.*

PROOF. If $z \in M$, there exist unique $r_i \in R$, almost all $r_i = 0$, such that $z = \sum_I r_i x_i$. In particular, the uniqueness of the r_i implies that $f : M \to N$ given by $z \mapsto \sum_I r_i y_i$ is well-defined. Clearly, f is uniquely determined by $x_i \mapsto y_i$ and f is an R-homomorphism. $\square$

In the notation of the theorem, we say that $x_i \mapsto y_i$ *extends linearly* to a homomorphism $f : M \to N$. The theorem says that any R-homomorphism from a free module M to another R-module is completely determined by where a basis for M is sent. Note, of course, that the variant of this statement (and proof) is one that you should know about vector spaces, which are just free modules over a field.

Remark 39.4. A better way of writing the universal property of free modules is the following: Let M be a free R-module on basis $\mathcal{B}$ and N an R-module. Given any set map

$g : \mathcal{B} \to N$ there exists a unique R-homomorphism $f : M \to N$ such that the diagram

$$\begin{tikzcd} \mathcal{B} \arrow[r, "inc"] \arrow[dr, "g"'] & M \arrow[d, "f"] \\ & N. \end{tikzcd}$$

commutes where *inc* is the inclusion map.

Corollary 39.5. *Let M and N be free R-modules on bases $\mathcal{B}$ and $\mathcal{C}$ respectively. If there exists a bijection $g : \mathcal{B} \to \mathcal{C}$, i.e., $|\mathcal{B}| = |\mathcal{C}|$, then $M \cong N$.*

PROOF. The maps g and g^{-1} of sets induce inverse R-isomorphisms $M \to N$ and $N \to M$. $\square$

Corollary 39.6. *Let M be an R-module (respectively, a finitely generated R-module). Then there exists a free R-module (respectively, a finitely generated free R-module) P and an R-epimorphism $g : P \to M$. In particular, we have a short exact sequence*

$$0 \to \ker g \longrightarrow P \xrightarrow{g} M \to 0.$$

PROOF. Let $Y = \{y_i\}_I$ generate M. If M is finitely generated, we may assume that I is finite. Let $P = \coprod_I R$ and $\mathcal{S}_I$ the *standard basis* for P. (So $\mathcal{S}_I := \{e_i \mid e_i = (\delta_{ij})_{j \in I})\}_I$, i.e., e_i has 1 in the ith component, 0 elsewhere.) Then the set map $\mathcal{S}_I \to M$ given by $e_i \mapsto y_i$ extends linearly to an R-homomorphism $g : P \to M$. As Y generates M, the map is surjective. $\square$

Unfortunately, the kernel of g in the corollary may not be nice, in particular, it may not be free. If R is a PID, it will be free. We shall show this later under the assumption that M is finitely generated.

Example 39.7. Let R be a commutative ring and M a free R-module on basis $\mathcal{B}$. As R is commutative, $M^* := \operatorname{Hom}_R(M, R)$ is an R-module, called the *dual module* of M. [Warning: It is usually not R-free.] For each $x \in \mathcal{B}$ define the *xth coordinate function* $f_x : M \to R$ by $\sum_{\mathcal{B}} r_x x \mapsto r_x$. It is an R-homomorphism. Set $\mathcal{B}^* := \{f_x \mid x \in \mathcal{B}\}$. The map $\mathcal{B} \mapsto \mathcal{B}^*$ by $x \mapsto f_x$ extends linearly to an R-homomorphism $f : M \mapsto M^*$. This map is monic and its image is R-free on basis $\mathcal{B}^*$. We call $\mathcal{B}^*$ the *dual basis* of $\mathcal{B}$. Unfortunately, this map depends on the basis $\mathcal{B}$, so it is not "natural". However, if we let $M^{**} := \operatorname{Hom}_R(M^*, R)$ and $L : M \to M^{**}$ the evaluation map at each x in M, i.e., the map L is given by $x \mapsto L_x : f \mapsto f(x)$, then this map is an R-monomorphism and is "natural" as it does not depend on the choice of a basis. If $\mathcal{B}$ is finite, then f is also surjective, hence $M \cong M^*$ (and $M \cong M^{**}$).

Remark 39.8. It follows by the Schroeder-Bernstein Theorem A.13, that any two bases of a free R-module that is not finitely generated have the same cardinality. Unfortunately, this is not true in general for finitely genenerated free R-modules, i.e., in general, we have no notion of dimension for finitely generated free R-modules (cf. Example 39.2(7)).

We end this section addressing an important case when any two bases of a finitely generated free R-module always have the same cardinality. The proofs of the following results are left as important exercises.

Definition 39.9. We say that a ring R satisfies the *invariant dimension property* or *IDP* if for every finitely generated free R-module P, every basis for P has the same cardinality (finite by Example 39.2(3). If R satisfies IDP, and P is a finitely generated free R-module, then the cardinality of a basis for P is called the *rank* of P and written $\operatorname{rank} P$.

We know that fields satisfy IDP. In fact, we have shown this (cf. Remark 17.13). It is a main ingredient in the proof of the following theorem that we leave as an exercise (cf. Exercise 39.12(7)).

Theorem 39.10. *Let R be a commutative ring. Then R satisfies IDP, i.e., if M is a finitely generated free R-module then all bases of M have the same cardinality.*

We also leave the proof of the following corollary to this theorem as an exercise (Exercise 39.12(8)).

Corollary 39.11. *Let R be a commutative ring and M and N be finitely generated free R-modules. Then* $\operatorname{rank} M \coprod N = \operatorname{rank} M + \operatorname{rank} N$.

Exercises 39.12.

1. Show if M is a nonzero finitely generated R-module, then M is R-free if and only if there exists a finite basis for M if and only if $M \cong R^n$ for some positive integer n.
2. Show that $\mathbb{Q}$ is not $\mathbb{Z}$-free. (Hint: Show any two distinct elements of $\mathbb{Q}$ are $\mathbb{Z}$- linearly dependent.)
3. Let R_0 be a commutative ring and $R = R_0[t_1, t_2]$. Let $M = Rt_1 + Rt_2$ Show the R-module M is not R-free.
4. Let R be a nontrivial commutative ring. Show that R is a field if and only if every finitely generated R-module is free.
5. Suppose that
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
is a short exact sequence of R-modules with C a free R-module. Show the sequence splits. (Cf. Exercise 38.18(12) for the definition of a split exact sequence.)
6. Let M be a free R-module. Suppose that $\mathcal{B}$ is a set and $\iota : \mathcal{B} \to M$ is set map (not necessarily one-to-one). Then $\iota : \mathcal{B} \to M$ satisfies the following universal property: Given any set map $j : \mathcal{B} \to N$, there exists a unique R-homomorphism $f : M \to N$ such that the diagram
$$\begin{tikzcd} \mathcal{B} \arrow[r, "\iota"] \arrow[dr, "g"'] & M \arrow[d, "j"] \\ & N. \end{tikzcd}$$
commutes. [Cf. Remark 39.4.]
7. Let P be a free R-module on basis $\mathcal{B} = \{x_i\}_{i \in I}$. Let $\mathfrak{A} < R$ be a (2-sided) ideal. Show all of the following:
   (i) $P/\mathfrak{A}P \cong \coprod_I Rx_i/\mathfrak{A}x_i \cong \coprod_I R/\mathfrak{A}$.

(*ii*) Let $^{-} : R \to R/\mathfrak{A}$ be the canonical ring epimorphism. Set $\overline{\mathcal{B}} = \{\overline{x}_i := x_i + \mathfrak{A}P \mid i \in I\}$. Then $P/\mathfrak{A}P$ is a free $\overline{R}$-module on basis $\overline{\mathcal{B}}$ and $|\overline{\mathcal{B}}| = |\mathcal{B}|$.

(*iii*) Let $\varphi : R \to S$ be a ring epimorphism with $S \neq 0$. If S satisfies IDP so does R.

(*iv*) Any commutative ring satisfies IDP.

8. Prove Corollary 39.11.
9. Show Example 39.2(7) has the desired properties.
[Hint: If $\mathcal{B} = \{e_i \mid i \in \mathbb{Z}^+\}$ is a basis for P, show that $\{f_1, f_2\}$ is a basis for S, where for all n, we have $f_1(e_{2n}) = e_n$, $f_1(e_{2n+1}) = 0$ and $f_2(e_{2n}) = 0$, $f_2(e_{2n+1}) = e_n$.]

Projective Modules:

10. Let P be an R-module. Then P is called *R-projective* if given any R-epimorphism $f : B \to C$ and R-homomorphism $g : P \to C$, there exists an R-homomorphism $h : P \to B$ such that the diagram

$$\begin{tikzcd} & P \arrow[dl, "h"'] \arrow[d, "g"] \\ B \arrow[r, "f"'] & C \end{tikzcd}$$

commutes. Show that any free R-module is projective.

11. Show that a direct summand of an R-free module is projective and a direct sum of R-modules is projective if and only if each direct summand of it is R-projective. (Projective modules are nicer than free modules in some ways, but submodules of projective modules may not be projective, e.g., $\mathbb{Z}/2\mathbb{Z}$ is a projective $\mathbb{Z}/6\mathbb{Z}$-module as $\mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$) as $\mathbb{Z}/6\mathbb{Z}$-modules).
12. Let P be an R-module. Show P is a projective R-module if and only if whenever $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is a short exact sequence, then it is split exact.
13. Let P be an R-module. Show that P is a projective R-module if and only if, whenever

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0 \tag{*}$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \operatorname{Hom}_R(P, A) \xrightarrow{f_*} \operatorname{Hom}_R(P, B) \xrightarrow{g_*} \operatorname{Hom}_R(P, C) \to 0$$

is exact. In particular, if C is R-projective, then (*) is split exact. (Cf. Exercises 38.18(123) and 38.18(19).)

14. Let G_i, $i \in I$, be groups. Show that $\bigtimes_I G_i$ satisfies the following property relative to the group homomorphisms $\pi_j : \bigtimes_I G_i \to G_j$ defined by $\{g_i\}_I \mapsto g_j$ for all $j \in I$: Given a group G and group homomorphisms $\phi_j : G \to G_j$, there exist a unique group homomorphism $\psi : G \to \bigtimes_I G_i$ satisfying $\phi_j = \pi_j \circ \psi$ for all $j \in I$.

Tensor Products:

15. Let R be a commutative ring and M, N two R-modules. Let P be the free R-module on basis $\{(m, n) \mid m \in M,\ n \in N\}$ and X the submodule of P generated by the elements

(*i*) $(m_1 + m_2, n) - (m_1, n) - (m_2, n)$ for all $m_1, m_2 \in M$ and $n \in N$.

(*ii*) $(m, n_1 + n_2) - (m, n_1) - (m, n_2)$ for all $m \in M$ and $n_1, n_2 \in N$.

(*iii*) $(rm,n) - r(m,n)$ for all $m \in M$, $n \in N$, and $r \in R$.
(*iv*) $(m,rn) - r(m,n)$ for all $m \in M$, $n \in N$, and $r \in R$.

Let $f : M \times N \to P/X$ be the R-bilinear map induced by $(m,n) \mapsto (m,n) + X$, i.e, an R homomorphism in each variable. Show that $f : M \times N \to P/X$ satisfies the following universal property: If $g : M \times N \to Q$ is an R-bilinear map, then there exists a unique R-homomorphism $h : P/X \mapsto Q$ such that

$$\begin{tikzcd} M \times N \arrow[r, "f"] \arrow[dr, "g"'] & P/X \arrow[d, "h"] \\ & Q \end{tikzcd}$$

commutes. The R-module P/X is called the *tensor product* of M and N and denoted by $M \otimes_R N$ and the elements $(m,n) + X$ are denoted by $m \otimes n$.

16. Let R be a commutative ring and M and N be R-modules. Show that $M \otimes_R N \cong N \otimes_R M$.
17. Let R be a commutative ring and M an R-module. Show that $M \cong R \otimes_R M$.
18. Let m and n be positive integers with greatest common divisor d. Show $\mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$.
19. Let R be a commutative ring and M_i, $i \in I$, and N be R-modules. Show $(\coprod_I M_i) \otimes_R N \cong \coprod_I (M_i \otimes_R N)$. In particular, show if M and N are free R-modules (respectively, and finitely generated), then $M \otimes_R N$ is a free R-module (respectively, of rank $M \cdot$ rank N).
20. Let R be a commutative ring and $f_1 : M_1 \to N_1$ and $f_2 : M_2 \to N_2$ be R-homomorphisms. Show these induce a unique R-homomorphism $f_1 \otimes f_2 : M_1 \otimes_R M_2 \to N_1 \otimes_R N_2$ satisfying $m_1 \otimes m_2 \mapsto f_1(m_1) \otimes f_2(m_2)$ for all $m_1 \in M_1$, $m_2 \in M_2$.
21. Let R be a commutative ring and
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
an exact sequence of R-modules and R-homomorphisms. Show for every R-module M, the sequence
$$A \otimes_R M \xrightarrow{1_M \otimes f} B \otimes_R M \xrightarrow{1_M \otimes g} C \otimes_R M \to 0$$
is exact.
22. Let M be a free module and $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to C \to 0$ an exact sequence of R-modules with R commutative. Show that
$$0 \to M \otimes_R A \xrightarrow{1_R \otimes f} M \otimes_R B \xrightarrow{1_R \otimes g} M \otimes_R C \to 0$$
is exact. Show the same holds if F is projective.
23. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. We know that S becomes an R-module via the pull back $rs := \varphi(r)s$. Let M be an R-module. Show that the R-module $S \otimes_R M$ becomes an S-module via $s_2(s_1 \otimes m) = s_1 s_2 \otimes m$ for all $s_1, s_2 \in S$ and $m \in M$. In particular, if M is a free R-module on basis $\mathcal{B}$, then $S \otimes_R M$ is a free S-module on basis $\{1 \otimes m \mid m \in \mathfrak{B}\}$.

24. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. Show if P is a projective R-module, then $S \otimes_R P$ is a projective S-module.
25. Let R be a commutative ring and A, B, C be R-modules. Show that
$$\operatorname{Hom}_R(A, \operatorname{Hom}_R(B, C)) \cong \operatorname{Hom}_R(A \otimes_R B, C).$$

CHAPTER IX

Noetherian Rings and Modules

Although this chapter is relatively short, it includes some of the most important basic results in commutative algebra, viz. the theorems of Hilbert introduced in Section 36. It is based upon the concept of a Noetherian ring, previously introduced, and its generalization to modules. Hilbert's theorems are necessary to begin the study of algebraic geometry. In particular, they show any set of polynomials not generating the unit ideal in $F[t_1, \dots, t_n]$, with F an algebraically closed field, have a common zero. A brief introduction to affine plane curves is also discussed.

40. Noetherian Modules

Proposition 40.1. *Let R be a ring and M an R-module. Then the following are equivalent:*

(1) Every submodule of M is finitely generated.
(2) M satisfies the ascending chain condition, *i.e., if $M_i \subset M$ are submodules and*

$$M_1 \subset M_2 \subset \cdots \subset M_n \subset \cdots,$$

then there exists a positive integer N such that $M_N = M_{N+i}$ for all $i \geq 0$. We say every ascending chain of submodules of M stabilizes. *Equivalently, there exists no infinite chain*

$$M_1 < M_2 < \cdots < M_n < \cdots.$$

(3) M satisfies the Maximum Principle, *i.e., if S is a nonempty set of submodules of M, then S contains a maximal element, that is a module $M_0 \in S$ such that if $M_0 \subset N$ with $N \in S$, then $N = M_0$.*

An R-module M satisfying any (hence all) of these equivalent conditions is called a *Noetherian R-module* or *R-Noetherian*.

Proof. (1) $\Rightarrow$ (2): Let

$$\mathcal{C}: \quad M_1 \subset M_2 \subset \cdots \subset M_n \subset \cdots$$

be a chain of submodules of M. It follows that the subset $M' := \bigcup_{i=1}^{\infty} M_i \subset M$ is a submodule. By (1), it is finitely generated, so we can write $M' = \sum_{i=1}^{n} Rx_i$ for some $x_i \in M'$. By definition, $x_i \in M_{j_i}$ some j_i. Let s be the maximum of the finitely many j_i's. Then $M' = M_s$. It follows that $M_s = M' = M_{s+i}$ for all $i \geq 0$.

(2) $\Rightarrow$ (3): Suppose that S is a nonempty set of submodules of M. Let M_1 lie in S. If M_1 is not maximal, there exists an $M_2 \in S$ with $M_1 < M_2$. Inductively, if M_i is not

maximal, there exists an M_{i+1} in S with $M_i < M_{i+1}$. By the ascending chain condition, the sequence

$$M_1 < M_2 < \cdots < M_i < \cdots$$

must terminate.

[To prove this statement completely, note that we made a choice of M_2 to contain M_1. This really needs the Axiom of Choice (Appendix A (A.8)). (Cf. the note in Exercise 30.22(11).)]

(3) $\Rightarrow$ (1): Let $N \subset M$ be a submodule and set

$$S := \{M_i \mid M_i \subset N \text{ is a finitely generated submodule }\}.$$

Then $(0) \in S$ so $S \neq \emptyset$. By assumption, there exists a maximal element $M' \in S$. If $N \neq M'$, then there exists $x \in N \setminus M'$. But M' finitely generated means that $M' + Rx \subset N$ is also finitely generated, so the submodule $M' + Rx$ of N lies in S. This contradicts the maximality of M'. Hence $N = M'$ is finitely generated. $\square$

Remark 40.2. To use Zorn's Lemma instead of the Axiom of Choice, we proceed as follows.

(1) $\Rightarrow$ (3): Let S is a nonempty set of submodules of M ordered by $\subset$. If $\mathcal{C}$ is a chain in S, then $\bigcup_{\mathcal{C}} N$ is a submodule of M so finitely generated by (1). It follows that there exists an $N_0 \in \mathcal{C}$ such that $\bigcup_{\mathcal{C}} N \subset N_0$. As $N' \subset \bigcup_{\mathcal{C}} N$ for all $N' \in \mathcal{C}$, $\bigcup_{\mathcal{C}} N = N_0$ lies in $\mathcal{C}$ and is an upper bound for $\mathcal{C}$. Zorn's Lemma now gives a maximal element in S. (Cf. Exercise 30.22(12).)

(2) $\Rightarrow$ (1): Let M be generated by $\{x_1, x_2, \dots, \}$. We may assume that with $Rx_1 < Rx_1 + Rx_2 < \cdots$. As this must be a finite chain by (2), the module M must be finitely generated.

(1) $\Rightarrow$ (2) and (3) $\Rightarrow$ (1) are as before.

Remark 40.3. Let $R = F[t_1, \dots, t_n, \dots]$ (infinitely many t_i). Let $M = R$ as an R-module. Then M is finitely generated since cyclic but the ideal $(t_1, \dots, t_n, \dots)$ is clearly not finitely generated, so R is not a Noetherian R-module. Thus, in general, submodules of finitely generated modules need not be finitely generated.

Definition 40.4. Let R be a commutative ring. We say that R is a *Noetherian ring* if R is a Noetherian R-module.

Note that Definition 40.4 agrees with our previous definition of a Noetherian ring.

We need another Noetherian R-module result. It is a crucial property about Noetherian modules.

Proposition 40.5. *Let M be an R-module and N a submodule of M. Then M is R-Noetherian if and only if N and M/N are R-Noetherian. In particular, if*

$$0 \to M' \to M \to M'' \to 0$$

is an exact sequence of R-modules with two of the modules M, M', M'' being R-Noetherian, then they all are R-Noetherian.

Proof. ⇒: Since $N_0 \subset N$ is a submodule, $N_0 \subset M$ is a submodule hence finitely generated – or any ascending chain in N is an ascending chain in M. Thus N is R-Noetherian. By the Correspondence Principle, a (countable) chain of submodules in M/N has the form $M_1/N \subset M_2/N \subset \cdots$ where $N \subset M_1 \subset M_2 \subset \cdots$ is a chain of submodules of M. Thus there exists an r such that $M_r = M_{r+j}$ for all $j \geq 0$ and hence $M_r/N = M_{r+j}/N$ for all $j \geq 0$.
⇐ is left as an exercise (cf. the analogous solvable groups result). □

Corollary 40.6. *If M, N are Noetherian R-modules, so is $M \amalg N$.*

Proof. $(M \amalg N)/N \cong M$ and N are Noetherian. □

The following result shows that the collection of Noetherian R-modules is a good generalization of finite dimensional vector spaces.

Theorem 40.7. *Let R be a Noetherian ring. If M is a finitely generated R-module, then M is R-Noetherian.*

Proof. Suppose $M = \sum_{i=1}^{n} Rx_i$. Let $f : R^n \to M$ be the R-epimorphism given by $e_i \mapsto x_i$, where $\{e_1, \dots, e_n\}$ is the standard basis for R^n. Since R is R-Noetherian so is R^n by Corollary 40.6, and hence so is $M \cong R^n/\ker f$ by Proposition 40.5. □

Corollary 40.8. *Let R be a Noetherian ring and M a finitely generated R-module. Then there exists positive integers m and n and an exact sequence*

$$R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0. \tag{40.9}$$

Proof. As is the proof of the theorem there exists an integer n and an R-epimorphism $f : R^n \to M$. Since R^n is finitely generated hence R-Noetherian, $\ker f$ is also finitely generated. Therefore, there exists an R-epimorphism $h : R^m \to \ker f$. Setting g to be the composition of h and the inclusion yields the result. □

Remark 40.10. If M is an R-module for which a sequence (40.9) exists, we say that M is a *finitely presented R-module*. The corollary says that any finitely generated module over a Noetherian ring is finitely presented. This is very useful. A commutative ring in which every finitely generated module is finitely presented is called a *coherent* ring. Therefore, every Noetherian ring is coherent.

Proposition 40.11. *Suppose that $f : R \to S$ is a ring epimorphism of commutative rings. If R is a Noetherian ring, then S is also a Noetherian ring.*

Proof. Let $\mathfrak{A} \subset S$ be an ideal. Then $f^{-1}(\mathfrak{A}) \subset R$ is an ideal hence finitely generated. Thus $\mathfrak{A} = f\big(f^{-1}(\mathfrak{A})\big)$ is finitely generated. □

Exercises 40.12.

1. Complete the proof of Proposition 40.5.
2. Let M be a Noetherian R-module. Show that any surjective R-endomorphism $f : M \to M$ is an isomorphism.
3. Let M be an R-module. Prove the following are equivalent:

(i) M satisfies DCC (the *descending chain condition*), i.e., if $M_i \subset M$ are submodules and
$$M_1 \supset M_2 \supset \cdots \supset M_n \supset \cdots$$
then there exists a positive integer N such that $M_N = M_{N+i}$ for all $i \geq 0$.

(ii) M satisfies the *Minimum Principle*, i.e., if S is a nonempty set of submodules of M, then S contains a minimal element, that is a module $M_0 \in S$ such that if $M_0 \supset N$ with $N \in S$, then $N = M_0$.

If M satisfies one of these two equivalent conditions, we say that M is an *Artinian R-module*.

[One needs the Axiom of Choice to show DCC implies the Minimum Principle.]

Also show that if R is a field then an R-module V is Artinian if and only if V is finite dimensional.

4. Let
$$0 \to M' \to M \to M'' \to 0$$
be an exact sequence of R-modules. Show that if two of the modules M, M', M'' are R-Artinian, then they all are R-Artinian.

5. Let R be a Noetherian ring, $\mathfrak{A}$ an ideal in R, A a finitely generated R-module, and B a submodule of A. Suppose that C is a submodule of A that contains $\mathfrak{A}B$ and is maximal with respect to the property that $C \cap B = \mathfrak{A}B$. Let x be an element of $\mathfrak{A}$. Show all of the following:
   (i) The chain of submodules of A, $D_m = \{a \in A \mid x^m a \in C \text{ for all } a \in A\}$, $m \in \mathbb{Z}^+$, stabilizes.
   (ii) There exists an integer n such that $(x^n A + C) \cap B = \mathfrak{A}B$.
   (iii) $\mathfrak{A}^n A \subset C$ for some n.
   (iv) (Krull Intersection Theorem) If $B = \bigcap_{i=0}^{\infty} \mathfrak{A}^n A$, then $\mathfrak{A}B = B$.

6. Let R be a commutative ring, $\mathfrak{A}$ an ideal in R, M an R-module generated by n elements, and x an element of R satisfying $xM \subset \mathfrak{A}M$. Show that $(x^n + y)M = 0$ for some y in $\mathfrak{A}$. In particular, if $\mathfrak{A}M = M$, then $(1 + y)M = 0$ for some $y \in \mathfrak{A}$.

7. Let R be a Noetherian ring. Using the previous two exercises, show the following:
   (i) Suppose that R is a domain and $\mathfrak{A} < R$ be an ideal. Let M be a finitely generated R-module satisfying $\operatorname{ann}_R(m) = 0$ for all $m \in M$. (We say that M is *R-torsion-free*.) Then $\bigcap_{i=0}^{\infty} \mathfrak{A}^n M = 0$.
   (ii) Let R be a Noetherian ring $\mathfrak{A} = \bigcap_{\operatorname{Max}(R)} \mathfrak{m}$, the *Jacobson radical* of R, (cf. Exercise 28.19(12)), and M a finitely generated R-module. Then $\bigcap_{i=0}^{\infty} \mathfrak{A}^n M = 0$.

41. Hilbert's Theorems

In this section, we prove the theorems that we discussed in Section 36. These theorems form the foundation of affine algebraic geometry.

Theorem 41.1. (Hilbert Basis Theorem) *If R is a Noetherian ring so is the ring $R[t_1, \ldots, t_n]$.*

Proof. By induction on n, it suffices to show that $R[t]$ is Noetherian. Let $0 \neq \mathfrak{B} \subset R[t]$ be an ideal. We must show that $\mathfrak{B}$ is finitely generated. Let

$$\mathfrak{A} = \{r \in R \mid r = \operatorname{lead} f,\ f \in \mathfrak{B}\}.$$

(Recall $\operatorname{lead} f$ is the leading coefficient of f.)

We first show that $\mathfrak{A}$ is an ideal. (Note that $0 \in \mathfrak{A}$ as $0 \in \mathfrak{B}$.) Let $a, b \in \mathfrak{A}$ and $r \in R$. To show $ra + b$ lies in $\mathfrak{A}$. We may assume that a and b are nonzero. Choose $f, g \in \mathfrak{B}$ say of degrees m and n respectively satisfying $\operatorname{lead} f = a$ and $\operatorname{lead} g = b$. Set $h = rt^n f + t^m g$ in $\mathfrak{B}$. Then $ra + b = \operatorname{lead} h$ proving $\mathfrak{A}$ is an ideal. As R is Noetherian, $\mathfrak{A} = (a_1, \ldots, a_n)$ for some nonzero $a_1, \ldots, a_n \in \mathfrak{A}$ with $n \in \mathbb{Z}^+$. Choose f_{d_i} in $\mathfrak{B}$ such that $a_i = \operatorname{lead} f_{d_i}$ and $\deg f_{d_i} = d_i$ for $i = 1, \ldots, n$. Let $\mathfrak{B}_0 = (f_{d_1}, \ldots, f_{d_n})$, an ideal in $R[t]$, and $N = \max\{d_1, \ldots, d_n\}$.

Let $f \in \mathfrak{B}$ with $\operatorname{lead} f = a$ and $\deg f = d$. Suppose that $d > N$. There exist $r_i \in R$ satisfying $a = \sum_{i=1}^n r_i a_i$, hence $f - \sum_{i=1}^n r_i t^{d-d_i} f_{d_i}$ lies in $\mathfrak{B}$ and has degree less than d. It follows by induction that there exists a $g \in \mathfrak{B}_0$ such that $f - g$ lies in $\mathfrak{B}$ with $\deg(f-g) \leq N$. As the R-module $M := \sum_{i=0}^N Rt^i$ is finitely generated and R is Noetherian, M is a Noetherian R-module. In particular, the submodule $M_0 = \{f \in \mathfrak{B} \mid \deg f \leq N\}$ is finitely generated. If $M_0 = \sum_{i=0}^m R[t]g_i$, then $\mathfrak{B} = \mathfrak{B}_0 + \sum_{i=0}^m R[t]g_i$ is finitely generated. □

It is worth noting the similarity of this proof and that for the general division algorithm. Of course, the division algorithm is a much stronger property, but the idea of a proof often generalizes.

Remarks 41.2. 1. Noetherian domains need not be UFD's, as $\mathbb{Z}[\sqrt{-5}]$ is an example of such a domain.

2. If F is a field, $F[t_1, \ldots, t_n, \ldots]$ is a UFD but is not Noetherian.

Definition 41.3. Let $R \subset S$ be commutative rings. We say that S is a *finitely generated commutative R-algebra* (or an *affine R-algebra* when R is a field) if there exist $x_1, \ldots, x_n$ in S satisfying $S = R[x_1, \ldots, x_n]$ as rings. (Recall that $R[x_1, \ldots, x_n]$ is the image of the evaluation map $e_{x_1,\ldots,x_n} : R[t_1, \ldots t_n] \to S$.)

Remarks 41.4. Let R be a commutative ring.

1. $R[t]$ is a finitely generated commutative R-algebra but definitely not a finitely generated R-module. (Why?)
2. Let S be a commutative ring with R a subring. If S is a finitely generated R-module, then S is a finitely generated commutative R-algebra.
3. If S is a finitely generated commutative R-algebra, it is not in general a finitely generated R-module, a much stronger condition. (Cf. (1).)
4. If T is a finitely generated commutative S-algebra and S is a finitely generated commutative R-algebra, then T is a finitely generated commutative R-algebra.
5. If T is a commutative ring with S a subring such that T is a finitely generated S-module and S is a finitely generated commutative R-algebra, then T is a finitely generated commutative R-algebra.

Proposition 41.5. *Let R be a commutative ring and S a finitely generated commutative R-algebra. If R is Noetherian so is S.*

PROOF. Let $S = R[x_1, \dots, x_n]$. Since $R[t_1, \dots, t_n]$ is Noetherian and we have a ring epimorphism $R[t_1, \dots, t_n] \to R[x_1, \dots, x_n]$ via $f(t_1, \dots, t_n) \mapsto f(x_1, \dots, x_n)$, all ideals of S are finitely generated by the Correspondence Principle. $\square$

Remark 41.6. Let R be a commutative ring. The proof (or definition) shows that S is a finitely generated commutative R-algebra if and only if there exists an ring epimorphism $R[t_1, \dots, t_n] \to S$ fixing R. Our definition of a finitely generated commutative R-algebra S is not the usual one. In general, one does not assume that $R \subset S$, hence only that there exists a surjective ring homomorphism $R[t_1, \dots, t_n] \to S$ for some n. If R is a field, we can always assume that $R \subset S$, as the above map must be monic when restricted to R if R is a field. This is the case of interest here, except for the following lemma.

Lemma 41.7. (Artin-Tate) *Let T be a commutative ring and $R \subset S$ be subrings of T. Suppose that R is Noetherian and T is a finitely generated commutative R-algebra. Suppose that as an S-module T is finitely generated. Then S is a finitely generated commutative R-algebra. In particular, S is a Noetherian ring.*

PROOF. Let $T = R[x_1, \dots, x_n]$ as a commutative R-algebra for some $x_i \in T$, some n and $T = \sum_{i=1}^{m} S y_i$ as an S-module for some $y_i \in T$, some m. Then for all $i = 1, \dots, n$ and $p, q = 1, \dots, m$, we have equations:

(i) $x_i = \sum_{j=1}^{m} a_{ij} y_j$ for some $a_{ij} \in S$ and

(ii) $y_p y_q = \sum_{k=1}^{m} b_{pqk} y_k$ for some $b_{pqk} \in S$ (since T is a ring).

Let $S_0 = R[a_{ij}, b_{pqk} \mid i = 1, \dots, n \text{ and } j, p, q, k = 1, \dots, m]$, a finitely generated R-algebra. Then $R \subset S_0 \subset S \subset T$.

Claim. T is a finitely generated S_0-module.

Let $f \in T$, so $f = \sum c_{i_1, \dots, i_n} x_1^{i_1} \cdots x_n^{i_n}$ for some $c_{i_1, \dots, i_n} \in R$. Applying properties (i) and (ii) repeatedly shows that $f \in \sum_{i=1}^{m} S_0 y_i$. Consequently, $T = S_0 y_1 + \cdots + S_0 y_m$ as claimed.

Since S_0 is a finitely generated commutative R-algebra, it is Noetherian by Corollary 41.5. Thus T, being a finitely generated S_0-module, is a Noetherian S_0-module. Consequently, $S \subset T$ is a finitely generated S_0-module. It follows immediately that S is a finitely generated commutative R-algebra. $\square$

Let R be an affine F-algebra, say $R = F[x_1, \dots, x_n]$. If R is a domain, we write the quotient field of R by $F(x_1, \dots, x_n)$. We shall need the following two lemmas about fields that we leave as easy exercises.

Lemma 41.8. *Suppose that $L \subset K \subset M$ are fields. Then M is a finite dimensional vector space over L if and only if both M is a finite dimensional vector space over K and K is a finite dimensional vector space over L.*

Lemma 41.9. *Let $L \subset M$ be fields. Suppose that $x \in M$ has the property that $L[x]$ is not a finite dimensional vector space over L. Then $L[x] \cong L[t]$ as rings and $L(x) \cong L(t)$ as fields.*

The clever Artin-Tate Lemma allows us to give an elementary proof of the following result that is essentially a form of the Hilbert Nullstellensatz.

Lemma 41.10. (Zariski's Lemma) *Let F be a field and E a field containing F such that E is an affine F-algebra. Then E is a finite dimensional vector space over F (i.e., a finitely generated F-module).*

PROOF. Suppose that $E = F[x_1, \dots, x_m]$. Since E is a field, $E = F(x_1, \dots, x_m)$. Suppose that E is not a finite dimensional vector space over F. Using Lemma 41.8, we see that $F(x_i)$ is an infinite dimensional vector space over F for some i. Relabeling, we may assume $i = 1$. Continuing in this way, we see that after relabeling the x_i, we may assume that $F(x_1, \dots, x_i)$ is not a finite dimensional vector space over $F(x_1, \dots x_{i-1})$ for $1 \le i \le r$, some r, and E is a finite dimensional fvector space over $F(x_1, \dots, x_r)$.

Let $K = F(x_1, \dots, x_r) \subset E$. We have $E = K(x_{r+1}, \dots, x_m)$ is a finitely generated K-module and F is a Noetherian ring (since a field). Thus the Artin-Tate Lemma implies that K is a finitely generated F-algebra. Write $K = F[y_1, \dots, y_n]$ for some $y_i \in K$. By Lemma 41.9, it follows that and $F[x_1, \dots, x_r] \cong F[t_1, \dots, t_r]$ and $K \cong F(t_1, \dots, t_r)$ as rings. Thus we can write

$$y_i = \frac{f_i(x_1, ..., x_r)}{g_i(x_1, ..., x_r)}$$

for some $f_i, g_i \in F[t_1, \dots, t_r]$, $g_i \neq 0$, $1 \le i \le n$. Let $g = g_1 \cdots g_n$ in $F[t_1, \dots, t_r]$. Thus $g(x_1, ..., x_r) \in F[x_1, \dots, x_r]$. We know the UFD $F[t_1, \dots, t_r]$ contains infinitely many non-associative irreducibles hence prime elements by Exercise 34.20(10). In particular, there exists an irreducible $f \in F[t_1, \dots, t_r]$ such that $f \not| g$. Thus $f(x_1, ..., x_r) \not| g(x_1, ..., x_r)$ in $F[x_1, \dots, x_r]$. As K is a field,

$$\frac{1}{f(x_1, ..., x_r)} \in K = F[y_1, \dots, y_n] = F\Big[\frac{f_1(x_1, ..., x_r)}{g_1(x_1, ..., x_r)}, \dots, \frac{f_n(x_1, ..., x_r)}{g_n(x_1, ..., x_r)}\Big].$$

This leads to an equation for $1/f(x_1, ..., x_r)$. Choosing an appropriate $N \ge 0$ to clear denominators, we see that we have an equation

$$\frac{g(x_1, ..., x_r)^N}{f(x_1, ..., x_r)} \in F[f_1(x_1, ..., x_r), \dots, f_n(x_1, ..., x_r)].$$

Then $\dfrac{g(x_1, ..., x_r)^N}{f(x_1, ..., x_r)}$ lies in $F[x_1, \dots, x_r]$, i.e., $f | g^N$ in $F[t_1, \dots, t_r]$, a contradiction. $\square$

Recall from §36, if $R = F[t_1, \dots, t_n]$ and $\mathfrak{A} \subset R$ is an ideal, the *affine variety* of $\mathfrak{A}$ in F^n is defined by

$$Z_F(\mathfrak{A}) = \{\underline{a} = (a_1, ..., a_n) \in F^n \mid f(\underline{a}) = 0 \text{ for all } f \in \mathfrak{A}\}.$$

By the Hilbert Basis Theorem, there exist $f_1, \dots, f_r$ in $F[t_1, \dots, t_n]$ satisfying $\mathfrak{A} = (f_1, \dots, f_r)$. We then also write $Z_F(f_1, \dots, f_r)$ for $Z_F(\mathfrak{A})$. In particular, $\underline{a}$ lies in $Z_F(\mathfrak{A})$ if and only if $f_i(\underline{a}) = 0$ for $i = 1, \dots, r$, as any element in $\mathfrak{A}$ is an R-linear combination of the f_i's.

Theorem 41.11. (Hilbert Nullstellensatz) (Weak Form) *Suppose that F is an algebraically closed field, $R = F[t_1, \dots, t_n]$, and $\mathfrak{A} = (f_1, \dots, f_r)$ is an ideal in R. Then $Z_F(\mathfrak{A})$ is the empty set if and only if $\mathfrak{A}$ is the unit ideal. Moreover, if $\mathfrak{A} < R$, then there exists a point $\underline{a} \in F^n$ satisfying $f_1(\underline{a}) = 0, ..., f_r(\underline{a}) = 0$.*

Proof. Certainly if $\mathfrak{A} = R$, then there exists no $\underline{a} \in F^n$ such that the element 1 in R evaluated at $\underline{a}$ takes the value zero, so we need only show if $\mathfrak{A} < R$, then $Z_F(\mathfrak{A})$ is not empty.

So suppose that $\mathfrak{A} < R$. Then there exists a maximal ideal $\mathfrak{m} < R$ satisfying $\mathfrak{A} \subset \mathfrak{m}$. Let $^- : R \to R/\mathfrak{m}$ be the canonical ring epimorphism. Set $E = R/\mathfrak{m} = F[\bar{t}_1, \dots, \bar{t}_n]$. As R is an affine F-algebra, so is E. By Zariski's Lemma, E is a finite dimensional F-vector space.

Claim. $E = F$:

Indeed let $x \in E$. Then $F[x]$ is a finite dimensional F-vector space, so $\{1, x, x^2, \dots, x^N\}$ must be linear dependent over F for some N, i.e., x is a root of some non-zero polynomial $f \in F[t]$. But any such f factors completely over F, as F is algebraically closed. Therefore, $E = F$.

Consequently the point $\underline{t} = (\bar{t}_1, \dots, \bar{t}_n)$ in $E^n = F^n$ lies in $Z_F(\mathfrak{A})$, since $\mathfrak{A} \subset \mathfrak{m}$, i.e., $\overline{\mathfrak{A}} = \overline{\mathfrak{m}} = 0$ □

Recall if F is a field and $\mathfrak{m}_{\underline{x}}$ is the ideal $(t_1 - x_1, \dots, t_n - x_n)$ with $\underline{x} = (x_1, \dots, x_n)$ in F^n, then $\mathfrak{m}_{\underline{x}}$ is a maximal ideal and equation (36.2) says

$$(*) \qquad \mathfrak{A} \subset \mathfrak{m}_{\underline{x}} \text{ if and only if } \underline{x} \in Z_F(\mathfrak{A}).$$

for an ideal $\mathfrak{A}$ in $F[t_1, \dots, t_n]$.

The Weak Form of the Hilbert Nullstellensatz says that the analogue of Exercise 26.21(13) that the maximal ideals in the ring of continuous real-valued functions on a finite closed (or compact) set are in one to one correspondence with the points of the set is valid over an algebraically closed field.

Corollary 41.12. *Let F be an algebraically closed field and $\mathfrak{m}$ a maximal ideal in $F[t_1, \dots, t_n]$. Then there exists an element $\underline{x}$ in F^n such that $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$.*

Proof. By the Weak Nullstellensatz, there exists a point $\underline{x}$ in $Z_F(\mathfrak{m})$. Let $^- : F[t_1, \dots, t_n] \to F[t_1, \dots, t_n]/\mathfrak{m}_{\underline{x}}$ be the canonical epimorphism. It takes $t_i \to \overline{t_i} = x_i$ for each i. By (*), we have $\mathfrak{m} \subset \mathfrak{m}_{\underline{x}}$. So $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$ as $\mathfrak{m}$ is a maximal ideal. □

Note that the corollary says if F is algebraically closed, then there is a bijection between points in F^n and maximal ideals in $F[t_1, \dots, t_n]$ and the points in $Z_F(\mathfrak{A})$ bijective with the maximal ideals containing $\mathfrak{A}$ in $F[t_1, \dots, t_n]$. (Cf. Exercise 26.21(13).)

The geometric interpretation of the Weak Hilbert Nullstellensatz is that the intersection of the "hyperplanes"

$$f_1 = 0, \dots, f_r = 0 \quad \text{in} \quad F^n.$$

always contains a common point over an algebraically closed field F, unless the f_i generate the unit ideal in $F[t_1, \dots, t_n]$.

Recall also that F being algebraically closed is essential. Indeed $f(t_1, ..., t_n) = t_1^2 + \cdots + t_n^2 + 1$ has no solution in $\mathbb{R}^n$ yet $(f) < \mathbb{R}[t_1, \ldots, t_n]$. Can you state what the above argument shows when F is not algebraically closed?

The Weak Hilbert Nullstellensatz says that $f_1, \ldots, f_r$ determine $Z_F(f_1, \ldots, f_r)$ when F is algebraically closed. The natural question is what polynomials in $F[t_1, \ldots, t_n]$ does $Z_F(f_1, \ldots, f_r)$ determine when F is algebraically closed, i.e., what polynomials in $F[t_1, \ldots, t_n]$ have zeros at every point in $Z_F(f_1, \ldots, f_r)$. Certainly elements in $\mathfrak{A} = (f_1, \ldots, f_r)$ have this property. However, a point in F^n is a zero of a polynomial f in the domain $F[t_1, \ldots t_n]$ if and only if it is a zero of f^m for any $m > 0$ if and only if it is a zero of f^m for some $m > 0$. This means that $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$ for any ideal $\mathfrak{A}$, i.e., at a minimum, every element in the *radical ideal* $\sqrt{\mathfrak{A}} := \{x \mid x^n \in F[t_1, \ldots t_n] \text{ for some positive integer } n\}$ of $\mathfrak{A}$ has every point in $Z_F(\mathfrak{A})$ a zero. The Strong Hilbert Nullstellensatz says, when F is algebraically closed, this is the totality of such polynomials.

Theorem 41.13. (Hilbert Nullstellensatz) (Strong Form) *Suppose that F be an algebraically closed field and $R = F[t_1, \ldots, t_n]$. Let $f, f_1, ..., f_r$ be elements in R and $\mathfrak{A} = (f_1, ..., f_r) \subset R$. Suppose that $f(a) = 0$ for all $a \in Z_F(\mathfrak{A})$. Then there exists an integer m such that $f^m \in \mathfrak{A}$, i.e., $f \in \sqrt{\mathfrak{A}}$. In particular, if $\mathfrak{A}$ is a prime ideal, then $f \in \mathfrak{A}$.*

PROOF. (Rabinowitch Trick). We may assume that f is nonzero. Let $S = R[t]$. Define the ideal $\mathfrak{B}$ in S by $\mathfrak{B} = (f_1, ..., f_r, 1 - tf) \subset S$. If $\mathfrak{B} < S$, then there exists a point $(a_1, ..., a_{n+1}) \in Z_F(\mathfrak{B})$. Thus $f_i(a_1, ..., a_n) = 0$ for all i and $1 - a_{n+1}f(a_1, ..., a_n) = 0$. In particular, $(a_1, ..., a_n)$ lies in $Z_F(\mathfrak{A})$. By hypothesis, this means that $f(a_1, ..., a_n) = 0$ which in turn implies that $1 = 0$, a contradiction. Thus $\mathfrak{B} = S$. So we can write

$$1 = \sum_{i=1}^{r} g_i f_i + g \cdot (1 - tf)$$

for some $g, g_i \in S$. Substituting $1/f$ for t and clearing denominators yields the result. □

If R is a commutative ring, and $\mathfrak{A}$ an ideal in R, then applying the Correspondence Principle to the canonical epimorphism $\bar{\ } : R \to R/\mathfrak{A}$ shows that $\sqrt{\mathfrak{A}}$ is the intersection of all prime ideals containing $\mathfrak{A}$. The strong form of the Nullstellensatz has an algebraic version that says for any field F, if R is an affine F-algebra and $\mathfrak{A}$ an ideal in R, then

$$\sqrt{\mathfrak{A}} = \bigcap_{\substack{\mathfrak{A} \subset \mathfrak{m} \\ \mathfrak{m} \text{ a maximal ideal}}} \mathfrak{m}.$$

We do not prove this here [The proof can be found below in Theorem 97.11 using localization techniques.] Instead we indicate what is going on (also without proof) assuming the reader knows what a topology is.

Let R be a nonzero commutative ring and set

$$\mathrm{Spec}(R) := \{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal}\},$$

called the *Spectrum* of R. If T is a subset of R, define

$$V_R(T) := \{\mathfrak{p} \mid \mathfrak{p} \in \mathrm{Spec}(R) \text{ with } T \subset \mathfrak{p}\}$$

called a *variety*. . Note that $V_R(T) = V_R(\langle T \rangle)$. We leave the following as an exercise.

Lemma 41.14. *Let R be a commutative ring, $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{A}_i, i \in I$, be ideals in R. Then*

(1) *If $\mathfrak{A} \subset \mathfrak{B}$, then $V_R(\mathfrak{B}) \subset V_R(\mathfrak{A})$.*
(2) *$V_R(\emptyset) = \mathrm{Spec}(R)$*
(3) *$V_R(R) = \emptyset$.*
(4) *$V_R(\sum \mathfrak{A}_i) = \bigcap_I V_R(\mathfrak{A}_i)$.*
(5) *$V_R(\mathfrak{A}\mathfrak{B}) = V_R(\mathfrak{A} \cap \mathfrak{B}) = V_R(\mathfrak{A}) \cup V_R(\mathfrak{B})$.*
(6) *$V_R(\mathfrak{A}) = V_R(\sqrt{\mathfrak{A}})$.*

Lemma (2) – (4) means that the collection

$$\mathcal{C} := \{V_R(T) \mid T \subset R\}$$

forms a system of *closed sets* for a topology on $\mathrm{Spec}(R)$ called the *Zariski topology*.

One shows that the set

$$\mathrm{Max}(R) := \{\mathfrak{m} \mid \mathfrak{m} < R \text{ a maximal ideal}\}$$

of *maximal ideals* in R is precisely the set of *closed points* in $\mathrm{Spec}(R)$, i.e., a closed set with precisely one element. If F is a field, then the collection

$$\mathcal{Z} := \{Z_F(T) \mid T \subset F[t_1, \dots, t_n]\}$$

also forms a system of closed sets for a topology of F^n by Exercise 36.5(1) called the *geometric Zariski topology* for F^n. If F is algebraically closed, the weak form of the Nullstellensatz says if $\mathrm{Max}(R)$ is given the induced topology, then

$$\mathrm{Max}(F[t_1, \dots, t_n]) \to F^n \text{ given by } \mathfrak{m}_{\underline{x}} \mapsto \underline{x}$$

is a homeomorphism and the strong form of the Nullstellensatz says that $\mathrm{Max}(F[t_1, \dots, t_n])$ is dense in $\mathrm{Spec}(F[t_1, \dots, t_n])$, i.e., every nonempty open set in $\mathrm{Spec}(R)$ intersects $\mathrm{Max}(R)$ nontrivially. This means that the closed points determine varieties. This last statement generalizes to $\mathrm{Max}(R)$ is dense in $\mathrm{Spec}(R)$ for any affine F-algebra R. (Cf. with (*) above.)

Exercises 41.15.

1. Prove that if R is a Noetherian ring so is the formal power series $R[[t]]$.
2. Prove Lemma 41.8. If M is a finite dimensional vector space over L, then $\dim_L M = \dim_k M \dim_L K$? [Compare this to the generalized version of Lagrange's Theorem for groups.]
3. Prove Lemma 41.9.
4. Let F be a field and $\varphi : A \to B$ be a ring homomorphism of affine F-algebras such that $\varphi|_F = 1_A$. Prove if $\mathfrak{m}$ is a maximal ideal in B, then $\varphi^{-1}(\mathfrak{m})$ is a maximal ideal in A.
5. Prove if R is a commutative ring, then $\sqrt{\mathfrak{A}}$ is the intersection of all prime ideals. containing $\mathfrak{A}$.
6. Prove Lemma 41.14.
7. Let R be a commutative ring. Prove that $\mathrm{Spec}(R)$ contains minimal elements under $\subset$.

42. Addendum: Affine Plane Curves

Let F be a field and $\mathfrak{A}$ an ideal in $F[t_1, \ldots, t_n]$. Then the affine variety $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$ (cf. Exercise 36.5). Note if $\mathfrak{p}$ is a prime ideal then $\mathfrak{p} = \sqrt{\mathfrak{p}}$. One can show:

Fact 42.1. *Let F be a field and $\mathfrak{A} < F[t_1, \ldots, t_n]$ an ideal. Then*

$$Z_F(\mathfrak{A}) = Z_F(\mathfrak{p}_1) \cup \cdots \cup Z_F(\mathfrak{p}_n) \tag{*}$$

where $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are the finitely many prime ideals in $F[t_1, \ldots, t_n]$ that minimally contain $\mathfrak{A}$ (i.e., if $\mathfrak{P}$ is a prime ideal of $F[t_1, \ldots, t_n]$ satisfying $\mathfrak{A} \subset \mathfrak{P}$, then there exists an i such that $\mathfrak{A} \subset \mathfrak{p}_i \subset \mathfrak{P}$).

We know that $F[t_1, \ldots, t_n]$ is a Noetherian ring. To show that there are finitely many such $\mathfrak{p}_i$ in $F[t_1, \ldots, t_n]$ in Fact 42.1 uses this. (cf. Exercise 30.22(21)). This reduces the study of $Z_F(\mathfrak{A})$ to $Z_F(\mathfrak{p})$ with $\mathfrak{p}$ a prime ideal in $F[t_1, \ldots, t_n]$. Such a variety is called *irreducible*. The decomposition (*) for $Z_F(\mathfrak{A})$ is called the *irreducible decomposition* of $\mathbb{Z}_F(\mathfrak{A})$ (cf. Exercise 30.22(21)) and the $Z_F(\mathfrak{p}_i)$ are called the *irreducible components* of $Z_F(\mathfrak{A})$. As $F[t_1, \ldots, t_n]$, is a UFD, it also turns out that every irreducible affine variety of "codimension one" in F^n is $Z_F(f)$, for some prime element f. [This essentially follows from the fact that every nonzero prime ideal in a UFD contains a prime element together with Exercise 31.7(3)) and "dimension theory", that we have not and will not develop that shows how this defines codimension algebraically.]

We shall need to use Exercise 35.12(3ii) in the development below. As we are interested in affine varieties in F^2, it is convenient to use X, Y as our variables instead of t_1, t_2.

Example 42.2. Let F be a field and $\mathfrak{A} = (XY)$ in $F[X, Y]$. Then $Z_F(\mathfrak{A}) = Z_F(X) \cup Z_F(Y)$ is an irreducible decomposition. Here $Z_F(X)$ is the Y-axis defined by $X = 0$ and $Z_F(Y)$ is the X-axis defined by $Y = 0$. The origin is the only point in $Z_F(X, Y)$. As $Z_F(X, Y) \subset Z_F(XY)$ (cf. Exercise 36.5), we see that irreducible components can intersect nontrivially. If $f \in F[X, Y]$ is a non-constant polynomial then $Z_F(f)$ is called an *affine (plane) curve* and *irreducible* if f is an irreducible polynomial as $Z_F(f)$ is the locus of $f = 0$ in F^2.

Proposition 42.3. *Let R be a UFD and f and g non-constant polynomials in $R[X, Y]$ having no non-constant common factor in $R[X, Y]$. Then*

$$Z_R(f) \cap Z_R(g) := \{(a, b) \in R^2 \mid f(a, b) = 0 = g(a, b)\}$$

is a finite set in R^2.

Proof. Let F be the quotient field of R and K the quotient field of $F[X]$. We view f and g in $K[Y]$. By Exercise 35.12(3ii), f and g have no common factor in $K[Y]$, i.e., they are relatively prime in the PID $K[Y]$, so there exists an equation, $1 = rf + sg$, with r and s polynomials in $K[Y]$. Clearing the denominators of the K-coefficients leads to an equation $0 \neq h = wf + zg$ in $R[X, Y]$, with w and z in $R[X, Y]$ and h in $R[X]$. Any common solution (a, b) in $Z_F(g) \cap Z_F(g)$ then satisfies $h(a) = h(a, b) = 0$ in R^2. Since h has finitely many roots in the domain R, there exist finitely many a satisfying $(a, b) \in Z_F(g) \cap Z_F(g)$. Applying the same argument with the variables reversed shows that $Z_F(g) \cap Z_F(g)$ is a finite set. $\square$

Remark 42.4. (Cf. Exercise 36.5). Let F be a field, $\mathfrak{A}$ and $\mathfrak{B}$ ideals in $F[X,Y]$. Then

$$Z_F(\mathfrak{A}\cap\mathfrak{B}) = Z_F(\mathfrak{A}\mathfrak{B}) = Z_F(\mathfrak{A})\cup Z_F(\mathfrak{B}).$$

The remark, reduces our study to ideals $\mathfrak{A} = (f_1,\dots,f_r)$ in $F[X,Y]$ with a gcd of the $f_1,\dots,f_r$ equal to one, i.e., with no common factors.

If h is a polynomial in $R[X,Y]$, R a ring, define the *total degree* $\deg h$ of h to be

$$\deg h := \max\{i+j \mid aX^iY^j \text{ a monomial in } h \text{ with } a \text{ nonzero}\}.$$

The proposition can be improved to

Facts 42.5. *(Bezout's Lemma) Let F be a field and f and g two non-constant polynomials in $F[X,Y]$ having no non-constant common factor in $F[X,Y]$. Then $|Z_F(f)\cap Z_F(g)| \le \deg f \deg g$.*

Our argument does not show this or give any bounds.

The proposition, in some sense, reduces our study of affine plane curves over a field to $Z_F(f)$ with f an irreducible polynomial in $F[X,Y]$ as $Z_F(f) = Z_F(f^r)$ for any positive integer r.

If $f = f(X,Y)$ is a polynomial in two variables, we can view f as a polynomial in Y over $F[X]$ or its quotient field $F(X)$, say the degree in Y of f is n. If we evaluate X at x in F, then $f(x,Y)$ is a polynomial of degree at most n. We want to show if F is algebraically closed of characteristic zero, that for almost all such x, the polynomial in the variable Y, $f(x,Y)$ is precisely of degree n and moreover has no multiple roots in F, where a is a *multiple root* of $f(x,Y)$ in F if $(Y-a)^2 \nmid f(x,Y)$ in $F[Y]$ for any root a of $f(x,Y)$ in F.

Definition 42.6. Let f be a polynomial in $F[X,Y]$ of degree $\deg_Y f$ of f in the variable Y satisfying $\deg_Y f = n > 0$. An element x in F is called a *regular value of f* if $f(x,Y)$ has n distinct roots in F.

Remark 42.7. Let F be a field of characteristic zero and x a regular value of f in $F[X,Y]$ of degree n in Y. We can write $f = \sum_{i=0}^n q_i(X)Y^i$ with $q_i(X)$ polynomials in $F[X]$ and $q_n(X)$ nonzero. As x is a regular value, $q_n(x)$ is nonzero and $f(x,Y)$ has n distinct roots. In particular none of the roots of $f(x,Y)$ are roots of the formal derivative $\frac{\partial f}{\partial Y}(x,Y)$ by Exercise 34.20(4ii) or by Exercise 42.10(4).

Proposition 42.8. *Let F be an algebraically closed field of characteristic zero, e.g., $F = \mathbb{C}$. Suppose f is an irreducible polynomial in $F[X,Y]$ with the degree $\deg_Y f$ of f in the variable Y satisfying $\deg_Y f = n > 0$. Then*

(1) For each a in F, there exists at most n elements b in F satisfying $(a,b)\in Z_F(f)$.
(2) There exists a finite subset Δ of F with the property that for any element a in $F\setminus\Delta$, there exist exactly n elements b in F satisfying (a,b) lies in $Z_F(f)$.

Question 42.9. What is $Z_F(f)$ if f in the proposition is irreducible but $\deg_Y f = 0$?

PROOF. Write $f = \sum_{i=0}^n q_i(X)Y^i$ with $q_i(X)$ polynomials in $F[X]$, $q_n(X)$ nonzero for some $n > 0$. If a lies in F, we have $f(a,Y) = \sum_{i=0}^n q_i(a)Y^i$ has at most n roots in F unless $q_i(a) = 0$ for all i. But $X - a \mid q_i(X)$ in $F[X]$ for all i if and only if $X - a \mid f$ in

$F[X, Y]$. As f is irreducible and $f \notin F[X]$, this possibility cannot occur. The remaining problem is when $f(a, Y)$ has a multiple root in $F[Y]$. Let

$$\Delta = \{a \in F \mid q_n(a) = 0 \text{ or } f(a, Y) \text{ has a multiple root in } F\}.$$

If $a \in \Delta$, then $f(a, Y)$ has at most n roots in F, as $f(a, Y)$ is not zero and if $a \in F \setminus \Delta$ then $f(a, Y)$ is a product of n linear polynomials in $F[Y]$ as F is algebraically closed, i.e., has precisely n distinct roots. So it suffices to show that Δ is finite. Certainly, there exist finitely many a in F with $q_n(a) = 0$, so it suffices to show the following:

Claim. There exist finitely many a in F with the nonzero polynomial $f(a, Y)$ having a multiple root.

As F is algebraically closed, Remark 42.7 says that $f(a, Y)$ has a multiple root if and only if $f(a, Y)$ and $\partial f(a, Y)/\partial Y$ have a common root in F.
We know that

(*i*) $\deg_Y \partial f(X, Y)/\partial Y < \deg_Y f(X, Y)$.

(*ii*) $f(X, Y)$ is irreducible in $F[X, Y]$ (as $f(X, Y) \in F(X)[Y]$ is irreducible by Lemma 35.7, where $F(X)$ is the quotient field of $F[X]$).

(*iii*) $\partial f(X, Y)/\partial Y$ is not zero (as $0 < \deg_Y f(X, Y)$ and the characteristic of F is zero (Why?).

These conditions mean that f and $\partial f/\partial Y$ have no common factors in $F[X, Y]$. Consequently, by Proposition 42.3, we have $Z_F(f) \cap Z_F(\partial f/\partial Y)$ is a finite set. This proves the claim. $\square$

If $F = \mathbb{C}$ in the above, then the affine plane curve $Z_{\mathbb{C}}(f)$ is called a *Riemann surface.* It is called a surface as $\dim_{\mathbb{R}} \mathbb{C} = 2$ and it is viewed as a surface over $\mathbb{R}$.

Using a little bit of analysis (e.g., the Implicit Function Theorem), one can show that if f is an irreducible polynomial in $\mathbb{C}[X, Y]$ with $\deg_Y f > 0$, then $Z_{\mathbb{C}}(f)$ has the following properties: Let Δ be as in Proposition 42.9 and $Z_{\mathbb{C}}(f)$ the induced topology in $\mathbb{C}^2$. Then there exists a continuous map $\pi : Z_{\mathbb{C}}(f) \to Z_{\mathbb{C}}(Y)$ (so $Z_C(Y)$ is the X-axis in $\mathbb{C}^2$, i.e., the real plane defined by $Y = 0$ in $\mathbb{C}^2$) satisfying

1. $|\pi^{-1}(x)| = n$ for all x in $Z_{\mathbb{C}}(Y) \setminus \Delta$.
2. For every x_0 in $Z_F(Y) \setminus \Delta$, there exists an open neighborhood U of x such that $\pi^{-1}(U)$ is a union of n disconnected sets V_i (so $\pi^{-1}(U) = V_1 \cup \cdots \cup V_n$) with each $V_i \subset Z_{\mathbb{C}}(f)$ open and $\pi|_{V_i} : V_i \to U$ a homeomorphism.

We say that $Z_{\mathbb{C}}(f)$ is an *n-sheeted branched covering* of $Z_F(Y)$.

Exercises 42.10.

1. Prove Remark 42.4.
2. Show if F is a field of characteristic zero and f a non-constant polynomial in $F[t]$, the its derivative f' is never zero. Show that if the characteristic of F is not zero, this may not be true by producing counterexamples.
3. Let F be a field of characteristic zero and f a polynomial in $F[t]$. Let $f = (t - a)^r g$ in $F[t]$ for some positive integer n and polynomial g in $F[t[$ with $g(a)$ nonzero. Show that a is a multiple root of f, i.e., $r > 1$ if and only if a is a root of f'.

4. Show that the Proposition 42.8 still is valid, if we weaken the condition that f be irreducible, to f is irreducible as a polynomial in X.

CHAPTER X

Finitely Generated Modules Over a PID

This is the deepest, and most difficult chapter in this part of the text. We are interested in classifying finitely generated modules over a PID which we succeed in doing. This has applications to abelian group theory and finite dimension vector spaces.

We begin by discussing row reduction for matrices over a ring. In particular, if the ring is a PID, we shall show that there is a unique matrix up to equivalence of matrices called the *Smith Normal Form* of the matrix.. If R is a euclidean domain, we give an algorithm for computing this matrix in §D. We shall show that any finitely generated module over a PID decomposes into a direct sum of two pieces, a free submodule and a *torsion* submodule (a module in which for every element m there exists a nonzero ring element r satisfying $rm = 0$). The torsion submodule also decomposes nicely (in two different ways). There is also a uniqueness statement for each of the two forms. For finitely generated abelian groups, this says that, up to isomorphism, every finitely generated abelian group is isomorphic to $\mathbb{Z}^r \times G$ with r unique and G a unique group that is a product of finite cyclic groups, each of which decomposes by the Chinese Reminder Theorem.

The theory is also applied to linear algebra. If V is a finite dimensional vector space over F and T a linear operator on V, then V becomes and $F[t]$-module by t acting on V by T. This leads to canonical forms of matrices over a field F. If F is arbitrary, the desired canonical form is called Rational canonical form.. If the field is algebraically closed, the desired canonical form is called Jordan canonical form. Further results in linear algebra are also proven, e.g., the Cayley-Hamilton Theorem.

43. Smith Normal Form

If R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$. Compare this to Change of Basis Theorem in linear algebra of matrix representations of linear transformations when R is a field (or division ring). [Cf. Appendix C.] Certainly, this relation is an equivalence relation, so one would like to find a good system of representatives. If $m = n$, then we are usually interested in the equivalence relation given by similarity of matrices, i.e., when $Q = P^{-1}$. This last equivalence relation is especially important in linear algebra, i.e., when F is a field. [Cf. this to the Change of Basis Theorem for the matrix representation of a linear operator in one basis to another in Appendix C.] For a general field this set of representatives will turn out to be the set of matrices in rational canonical form and over an algebraically closed field this set of representatives will turn out to be the set of Jordan canonical forms. In this section, we determine a set of representatives for equivalence of matrices over a PID. As this includes $F[t]$ when F is a field, we shall see that it will aid us solving the equivalence of similarity over matrices. The system of representatives for equivalence of matices will be the set of matrices in Smith Normal Form. In Appendix D, we will give an algorithm for compute this form of any matrix when R is a euclidean ring. In this section, we do the more general case that R is a PID. Unfortunately, unlike the euclidean case, there is no algorithm to compute it.

Definition 43.1. Let R be a domain and $A \in R^{m\times n}$. We say that A is in *Smith Normal Form* if A is a diagonal matrix of the form

$$A = \operatorname{diag}(q_1, \dots, q_r, 0, \dots, 0) \text{ with } q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r \text{ in } R \text{ and } q_r \neq 0.$$

[Note that the number of diagonal entries is the minimum of n and m.]

We want to prove that every matrix in $R^{m\times n}$ is equivalent to a matrix in Smith Normal Form if R is a PID. If R is a euclidean domain, using the euclidean function provides a method to induct on the value of remainders in the division algorithm. As mentioned above, this leads to a computable algorithm, one used by computer programs to find the Jordan canonical form of square matrices over the complex numbers. We need a substitute for the euclidean function on a euclidean domain. As any PID is a UFD, we can induct on the number of irreducible factors (counted with multiplicity) for any nonzero nonunit.

Definition 43.2. Let R be a UFD and a a nonzero nonunit in R. Let $a = p_1 \cdots p_r$ be a factorization into irreducibles in R (not necessarily distinct). Define the *length*, $l(a)$ of a to be r. If a is a unit, define its length, $l(a)$ to be zero.

To prove our result, we need the following lemma. It requires R to be a PID.

Lemma 43.3. *Let R be a PID and $A = (a_{ij})$ a nonzero $m \times n$ matrix in $R^{m\times n}$. Fix i, j with $a_{ij} \neq 0$ and $j < n$. Let d in the ideal $(a_{ij}, a_{i,j+1})$ in R satisfy $(d) = (a_{ij}, a_{i,j+1})$. Then we have:*

(1) If $a_{ij} \nmid d$, then $l(d) < l(a_{ij})$. In particular, this is true if $a_{ij} \nmid a_{i,j+1}$.
(2) There exists a matrix B in $\mathrm{SL}_n(R)$ satisfying $AB = (c_{lk})$ with $c_{ij} = d$ and $c_{i,j+1} = 0$.

Remark 43.4. An analogous statement holds for two column entries in A, but in (2), the elment B in $\mathrm{SL}_n(R)$ multiplying A on the left. More generally, a similar statement holds for any two fixed elements in A, at least one not zero, in any fixed row or any fixed column. We prove the lemma as it is stated for convenience of notation.

Proof. (of the lemma). That d satisfies the first condition is immediate. We construct B in (2). For notational convenience, write $v = a_{ij}$ and $w = a_{i,j+1}$, so $(d) = (v, w)$. Then $0 \neq d = xv + yw = (xa + yb)d$ in the domain R, so $1 = xa + yb$ in R, hence

$$P = \begin{pmatrix} x & -b \\ y & a \end{pmatrix} \text{ lies in } \mathrm{SL}_2(R) \text{ [with inverse } P^{-1} = \begin{pmatrix} a & b \\ -y & x \end{pmatrix}].$$

Then

$$\begin{pmatrix} v & w \end{pmatrix} P = \begin{pmatrix} vx + wy & -bv + wa \end{pmatrix} = \begin{pmatrix} d & -bad + bda \end{pmatrix} = \begin{pmatrix} d & 0 \end{pmatrix},$$

so

$$B = \begin{array}{c@{}l} \begin{pmatrix} 1 & \cdots & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ \vdots & & & P & & & \vdots \\ & & & & 1 & & \\ & & & & & \ddots & \\ 0 & \cdots & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \\ j \\ j+1 \\ \\ \\ \end{array} \\ \begin{array}{ccccccc} \quad & \quad & \quad & i & i+1 & \quad & \quad \end{array} & \end{array}$$

works (if $j \neq m$, otherwise have P end in the mth row). $\square$

Theorem 43.5. *Let R be a PID and A an $m \times n$ matrix in $R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form.*

PROOF. We may assume that the matrix $A = (a_{ij})$ is nonzero.

We use elementary row and column operations on matrices. We call an elementary row (respectively, column) operation *Type I* if we add a multiple of one row (respectively, one column) to another and *Type II* if we permute two rows (respectively, columns). These correspond to multiplying on the right or left by invertible matrices. [For more details, see Appendix D.] In particular, these elementary row and column operations yield equivalent matrices. Using Type II elementary row and column operations, we may assume that the $(1,1)$ entry in A is a nonzero element a in R such that among all nonzero entries in A, we have $l(a)$ is minimal.

Suppose that a is a unit. Then using Type I elementary row and column operations, we can convert A to the equivalent matrix

$$(*) \qquad \begin{pmatrix} a & \cdots & 0 \\ 0 & & \\ \vdots & A_1 & \\ 0 & & \end{pmatrix}$$

with A_1 an $(m-1)\times(n-1)$ matrix.

By induction on matrix size, there exist appropriate invertible matrices P_1 and Q_1 such that $P_1 A_1 Q_1$ is in Smith Normal Form. Then

$$\begin{pmatrix} 1 & \cdots & 0 \\ 0 & & \\ \vdots & P_1 & \\ 0 & & \end{pmatrix} \begin{pmatrix} a & \cdots & 0 \\ 0 & & \\ \vdots & A_1 & \\ 0 & & \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 \\ 0 & & \\ \vdots & Q_1 & \\ 0 & & \end{pmatrix}$$

is in Smith Normal Form and is equivalent to A. This proves the result in the case that a is a unit. Note that if A is equivalent to a matrix in the form of (*) such that a divides every entry in A_1, the argument above shows that we would also be done.

So we may assume that $l(a) > 0$. If $a \nmid a_{ij}$ for some $i = 1$ or $j = 1$, using Lemma 43.3, we can convert A into an equivalent matrix

$$\begin{pmatrix} d & * \cdots & * \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix}$$

with $l(d) < l(a)$, and we are done by induction on $l(a)$. So we may assume that $a \mid a_{ij}$ if $i = 1$ or $j = 1$. Using Type I row and column operations converts A to a matrix equivalent to A of the form (*). By induction of matrix size, there exist invertible matrices of the appropriate size such that A is equivalent to a diagonal matrix $\operatorname{diag}(a, a_2, \ldots, a_r, 0, \ldots, 0)$ with $a_2 \mid a_3 \mid \cdots \mid a_r$ and $a_r \neq 0$ in R. If $a \mid a_2$, we are done; if not use an Type I elementary row operation to convert this matrix to

$$\begin{pmatrix} a & a_2 & \cdots & & & \\ 0 & a_2 & & & & \\ 0 & 0 & a_3 & & & \\ \vdots & & & \ddots & & \\ & & & & a_r & \\ & & & & & \ddots \end{pmatrix}.$$

As $a \nmid a_2$, we can apply the Lemma 43.3 to produce an equivalent matrix with $(1, 1)$ entry d satisfying $l(d) < l(a)$, and we are done by induction on $l(a)$. $\square$

To prove that the Smith Normal Form of a matrix in $R^{m \times n}$ with R a PID is essentially unique, we shall need facts about the determinant that we shall use but shall not prove here. [However, cf. Section 121 for a sophisticated proof of the determinant and its properties.]

Let R be a commutative ring. Then the *determinant* $\det : \mathbb{M}_n(R) \to R$ is a map that has the following properties:

(1) det is *n-multilinear* (or *n-linear*) as a function of the rows (respectively, columns) of matrices in $\mathbb{M}_n(R)$. This means that it is R-linear (i.e., an R-homomorphism) in each of the n entries fixing the others, i.e.,

$$\begin{aligned} &\det(\alpha_1, \ldots, r\alpha_i + \alpha_i', \ldots, \alpha_n) \\ &\qquad = r \det(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + \det(\alpha_1, \ldots, \alpha_i', \ldots, \alpha_n) \end{aligned}$$

for all r in R, and rows (respectively columns) of matrices A in $\mathbb{M}_n(R)$.

(2) det is *alternating* as a function of the rows (respectively, columns) of matrices in $\mathbb{M}_n(R)$, i.e., if A has two identical rows (respectively, columns), then $\det A = 0$.

[Note: This implies that the matrix obtained by interchanging two rows (respectively columns) of A has determinant $-\det A$.

(3) $\det I = 1$

Indeed it can be shown that det : $\mathbb{M}_n(R) \to R$ is the unique function satisfying (1), (2), and (3); and it is given by

$$\det(a_{ij}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

where

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \in A_n \\ -1 & \text{if } \sigma \notin A_n. \end{cases}$$

Let A be an $m \times n$ matrix in $R^{m\times n}$ and $1 \leq l \leq \min\{m, n\}$. An *$l$-order minor* of A is a determinant of an $l \times l$ submatrix of A, i.e., a matrix obtained from A be deleting $m - l$ rows and $n - l$ columns of A. The key to proving our uniqueness statement is the following lemma:

Lemma 43.6. *Suppose that R is a UFD and A an $m \times n$ matrix in $R^{m\times n}$. Let P and Q be invertible matrices in $\mathrm{GL}_m(R)$ and $\mathrm{GL}_n(R)$ respectively. Set $B = PAQ$. If $1 \leq l \leq \min\{m, n\}$ and*

(1) if the element a in R is a gcd of all the l-order minors of A, and
(2) if the element b in R a gcd of all the l-order minors of B,

then a and b are associates, i.e., the ideals (a) and (b) are the same.

Proof. Let $P = (p_{ij})$, $A = (a_{ij})$, $Q = (q_{ij})$ and c in R a gcd of all the l-order minors of PA. Then the kith entry of PA is $\sum_j p_{kj}a_{ji}$, so the kth row of PA is $\sum_j p_{kj}(a_{j1}\, a_{j2} \cdots a_{jn})$ (with the obvious notation). As the determinant is multilinear as a function of the rows of the square submatrices of A, we have $a \mid c$ in R. As the determinant is multilinear as a function of the columns of the square submatrices of PA, an analogous argument shows that $c \mid b$ in R, hence $a \mid b$ in R. But we also have $A = P^{-1}BQ^{-1}$, so arguing in the same way, we conclude that we also have $b \mid a$. Since R is a domain, a and b must be associates. □

Corollary 43.7. *Suppose that R is a PID and A is an $m \times n$ matrix in $R^{m\times n}$ with $B = \operatorname{diag}(d_1, \dots, d_r, 0 \dots, 0)$ in $R^{m\times n}$ satisfying $d_1 \mid \cdots \mid d_r$ and $d_r \neq 0$ in R a Smith Normal Form of A. Let*

$$\Delta_l \textit{ be a gcd of all the } l\textit{-order minors of } A \textit{ in } R \textit{ for } 1 \leq l \leq r$$

and $\Delta_0 = 1$. Then

$$\Delta_0 \mid \Delta_1 \mid \cdots \mid \Delta_r \quad \textit{and} \quad d_l \approx \frac{\Delta_l}{\Delta_{l-1}} \textit{ in } R \textit{ for all } l > 0.$$

Putting this all together, we obtain the following theorem:

Theorem 43.8. *Let R be a PID and A an $m \times n$ matrix in $R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form. Moreover, if $\operatorname{diag}(a_1, \dots, a_r, 0, \dots, 0)$ and $\operatorname{diag}(b_1, \dots, b_s, 0, \dots, 0)$ are two Smith Normal Forms for A, then $r = s$ and $a_i \approx b_i$ for $1 \leq i \leq r$. In particular, the descending sequence of ideals in R*

$$(a_1) \supset (a_2) \supset \cdots \supset (a_r)$$

completely determine a Smith Normal Form of A with any generator of (a_l) for $1 \leq l \leq r$ being an associate of Δ_l/Δ_{l-1}, where Δ_l is a gcd of all the l-order minors of A in R and $\Delta_0 = 1$.

The elements $a_1 \mid \cdots \mid a_r$ in the theorem are called the *invariants factors* of A. They are unique up to units. If $g : R^m \to R^n$ is an R-homomorphism, a matrix representation $[g]_{\mathcal{B},\mathbb{C}}$ of g in Smith Normal Form will be called *Smith Normal Form* of g and the invariant factors of $[g]_{\mathcal{B},\mathbb{C}}$ will be called the *invariant factors* of g.

Exercise 43.9. Prove Remark 43.4.

44. The Fundamental Theorem

In this section, we completely determine finitely generated modules over a PID up to an isomorphism. The basic idea is to first show, unlike in the general case, that any submodule of a finitely generated free module over a PID is itself free. This means that if we have a finitely generated R-module M with R a PID, that we have a short exact sequence

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0.$$

Therefore, M is isomorphic to $\operatorname{coker} g$. As g is a map of free R-modules, it has a matrix representation A (cf. Appendix C) relative to the appropriate standard bases. Looking at a Smith Normal Form of A produces a direct sum decomposition for $\operatorname{coker} g$, hence M decomposes into a direct sum of R-cyclic modules. The uniqueness of the Smith Normal Form will show this decomposition is unique up to isomorphism.

We begin by proving that submodules of finitely generated free modules over a PID are free.

Proposition 44.1. *Let R be a PID, N a free R-module of rank n, and M a submodule of N. Then M is a free R-module satisfying* $\operatorname{rank} M \leq \operatorname{rank} N = n$.

Proof. We know that $N \cong R^n$, so we may assume that M is a submodule of R^n; and must show that M is free of rank at most n. We show this by induction on n. If $n = 0$, this is trivial, so we may assume that $n \geq 1$. Let

$$\pi_1 : R^n \to R \text{ given by } (a_1, \ldots, a_n) \mapsto a_1$$

be the projection onto the first coordinate, an R-epimorphism with kernel

$$\ker \pi_1 = \{(0, a_2, \ldots, a_n) \mid a_i \in R\} \cong R^{n-1}$$

free of rank $n-1$. Let $\pi_1|_M : M \to R$ be the restriction. It is an R-homomorphism with $\ker \pi_1|_M \subset \ker \pi_1 \cong R^{n-1}$, so a free R-module of rank at most $n-1$ by induction. For convenience, let $\pi_1' = \pi_1|_M$. In particular, we are done if π_1' is the zero map. Hence we may assume that π_1' is not trivial, i.e., $0 < \operatorname{im} \pi_1' \subset R$ is a nonzero left ideal in the PID R. It follows that there exists a nonzero $a \in R$ satisfying $(a) = \operatorname{im} \pi_1'$. As $ra = 0$ in the domain R implies that $r = 0$, we conclude that $(a) \cong R$ as R-modules, so (a) is free of rank 1. Choose $m \in M$ such that $\pi_1'(m) = a \neq 0$. If $rm = 0$, then $0 = \pi_1'(rm) = ra$ in the domain R, so $r = 0$ and $Rm \cong R$ is a free R-module of rank one.

Claim. $M = Rm \oplus \ker \pi_1'$:

If we show the claim, we are done. Indeed if Rm is R-free of rank 1 and $\ker \pi_1'$ is R-free of rank at most $n-1$, then M is R-free of rank $\operatorname{rank} Rm + \operatorname{rank} \ker \pi_1' \leq n$ by Corollary 39.11, as needed. So we need only show the claim.

We first show $M = Rm + \ker \pi_1'$:

Let $x \in M$, so $\pi_1'(x) = ra$ for some r in R, hence lies in (a). It follows that $x - rm$ lies in $\ker \pi_1'$, hence x lies in $Rm + \ker \pi_1'$.

Next we show $Rm \cap \ker \pi_1' = \emptyset$:

Suppose that x lies in $Rm \cap \ker \pi_1'$. Then $x = rm$ for some $r \in R$ and $0 = \pi_1'(x) = \pi_1'(rm) = ra$ in the domain R. As a is nonzero, we have $r = 0$, so $x = 0$ as needed.

This establishes the claim and hence the proposition. $\square$

As mentioned before, it is in fact true that any submodule of a free module over a PID is free, finitely generated or not. The fact that a free submodule of a finitely generated free module N has rank bounded by the rank of N is in fact true over any commutative ring, but this is harder (especially the case when the ring is not a domain).

The theorem has an interesting consequence even stronger than that mentioned in the introduction of this section.

Corollary 44.2. *Let R be a PID and M a finitely generated R module. Then there exists an exact sequence of R-modules*

$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0.$$

with $m \leq n$.

Proof. As M is generated by n elements, for some positive integer n, we have an exact sequence

$$0 \to \ker f \xrightarrow{inc} R^n \xrightarrow{f} M \to 0.$$

As $\ker f \subset R^n$, it is R-free of rank at most n by the proposition. $\square$

If M is a finitely generated R module with R a PID, then the sequence

$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0.$$

says that M can be generated by n elements and the relations on these generators generates a free R-module of rank at most m, i.e., the minimal number of relations on the generators that generate all relations on these generators is at most m. It is called a *free resolution* of M. We shall also say that M has a *free presentation*. In particular, if n is the minimal number of generators of M, one can show that m is the minimal number of generating relations for M among all free resolutions of M. We can also say something about submodules of finitely generated modules over a PID.

Corollary 44.3. *Let R be a PID and M a finitely generated R-module. Suppose that M can be generated by n elements. Then any submodule of M can be generated by n elements.*

Proof. Let
$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0$$
be a free resolution of M and N a submodule of M. By the Correspondence Principle, there exists a submodule B of R^n satisfying $\ker f \subset B \subset R^n$ and $f(B) = N$, i.e., $N \cong B/\ker f$. As B is R-free of rank at most n, it can be generated by $\leq n$ elements, hence the same is true of N. □

Remarks 44.4. 1. If R is a commutative noetherian ring but not a PID, then there exists a non-principal ideal $\mathfrak{A} = (x, y)$ of R. We have an exact sequence
$$0 \to \mathfrak{A} \to R \xrightarrow{-} R/\mathfrak{A} \to 0.$$
R is a cyclic R-module but $\mathfrak{A}$ is not. Nor is $\mathfrak{A}$ R-free as $x \cdot y - y \cdot x = 0$ in R for any generators of $\mathfrak{A}$. In fact, a similar argument shows that an ideal in a ring R is free as an R-module if and only if it is principal. This shows that the last three results do not hold for commutative rings that are not PIDs.

2. The proper $\mathbb{Z}$-free submodule $2\mathbb{Z}$ of $\mathbb{Z}$ is of rank 1, the same as $\mathbb{Z}$.
3. It is, in fact, true that if R is a commutative ring, M a finitely generated free R-module of rank n, then any submodule of M that is also R-free has rank at most n. The proof is not easy (although easier for the case when R is a domain).

Observation 44.5. Let N_i be a submodule of the R-module M_i for $i = 1, 2$. Then we have an R-isomorphism
$$(M_1 \coprod M_2)/(N_1 \coprod N_2) \cong (M_1/N_1) \coprod (M_2/N_2)$$

Theorem 44.6. (Fundamental Theorem of fg Modules over a PID, Form I) *Let R be a PID and M a nontrivial finitely generated R-module. Then M is a direct sum of cyclic R-modules. More precisely, there exist a nonnegative integer r and*

(*) $\qquad m_1, \dots, m_r \in M$ *satisfying* $M = P \oplus Rm_1 \oplus \cdots \oplus Rm_r$

with P a free R-module of rank s, some $s \geq 0$, and there exist nonzero nonunits

(**) $\qquad d_i \in R$ *satisfying* $\begin{cases} \operatorname{ann}_R m_i = (d_i) & i = 1, \dots, r \\ d_1 \mid \cdots \mid d_r. \end{cases}$

Moreover, r and s are unique in (*) and the d_i in (**) are unique up to units and the order determined by $d_1 \mid \cdots \mid d_r$, i.e., any direct sum decomposition of M satisfying (*) and (**) has the R-free submodule in (*) of the same rank s, the number of non-free R-cyclic modules the same r with the corresponding descending chain of ideals
$$(d_1) \supset (d_2) \supset \cdots \supset (d_r)$$
unique.

The elements $d_1, \dots, d_r$ in the Fundamental Theorem 44.6 are called *invariant factors* of M. Note also that the Rm_i in the Fundamental Theorem 44.6 are unique up to isomorphism with $Rm_i \cong R/(d_i)$ by Proposition 38.14. Of course, if $d_i \approx d_j$, then $Rm_i \cong Rm_j$. In particular,
$$M \cong R^s \coprod R/(d_1) \coprod \cdots \coprod R/(d_r),$$

and the right hand side completely determines the isomorphism class of M, i.e., the invariant factors and s completely determine M up to isomorphism.

PROOF. **Existence**: We have shown that there exists an exact sequence

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0 \tag{44.7}$$

of R-modules with $m \leq n$, i.e, $M \cong \operatorname{coker} g = R^n / \operatorname{im} g$. Let $\mathcal{S}_m$ and $\mathcal{S}_n$ be the standard bases for R^m and R^n, respectively. By Theorem 43.8 and the Change of Basis Theorem from linear algebra (cf. Theorem C.7 in Appendix C), there exist bases $\mathcal{B}$ and $\mathcal{C}$ for R^m and R^n, respectively, such that the $n \times m$ matrix $[g]_{\mathcal{B},\mathcal{C}}$ is a Smith Normal Form of $[g]_{\mathcal{S}_m,\mathcal{S}_n}$.

Let $\mathcal{C} = \{v_1, \dots, v_n\}$ and $[g]_{\mathcal{B},\mathcal{C}} = \operatorname{diag}(d_1, \dots, d_r, 0, \dots, 0)$ with $d_1 \mid \cdots \mid d_r$ in R, $r \leq m \leq n$, and d_r nonzero. As g is injective, no column of $[g]_{\mathcal{B},\mathcal{C}}$ can be zero. Therefore, $r = m$. We have

$$V = \bigoplus_{i=1}^{n} Rv_i$$
$$\operatorname{im} g = \bigoplus_{i=1}^{m} Rd_iv_i \bigoplus \bigoplus_{i=m+1}^{n} 0.$$

Let $d_i = 0$ for $m < i \leq n$. As $R \to Rv_i$ by $r \mapsto rv_i$ is an R-isomorphism taking $(d_i) \to Rd_iv_i$ isomorphically, we have $R/(d_i) \cong Rv_i/Rd_iv_i$ for $i = 1, \dots, n$. Hence by Observation 44.5, we have

$$\begin{aligned} M \cong \operatorname{coker} g &\cong (\coprod_{i=1}^{n} Rv_i)/(\coprod_{i=1}^{n} Rd_iv_i) \\ &\cong \coprod_{i=1}^{m} (Rv_i/Rd_iv_i) \amalg \coprod_{i=m+1}^{n} Rv_i \cong \coprod_{i=1}^{m} R/(d_i) \amalg R^{n-m}. \end{aligned}$$

If d_i is a unit, then $R/(d_i) = 0$, so dropping such d_i, we have an isomorphism

$$f : M \longrightarrow \coprod_{\substack{i=1 \\ d_i \notin R^\times}}^{m} R/(d_i) \amalg R^{n-m}$$

Let $m_i = f^{-1}\big(1_R + (d_i)\big)$ for d_i not a unit or zero. Then $(d_i) = \operatorname{ann}_R(m_i)$ by Proposition 38.14. The isomorphism f^{-1} gives the desired decomposition of M with $P = f^{-1}(R^{n-m})$.

For the uniqueness, we need further observations that we leave as exercises.

Observations 44.8. Let R be a PID, e and d nonzero elements in R, g a gcd of e and d, p a prime element in R (so p is nonzero and $R/(p)$ is a field), and N an R-module. Then the following are true:

(1) $dR^m \cong R^m$.

(2) $d\big(R/(e)\big) \cong R/(\frac{e}{g})$.

(3) $R/(p)$ acts on N/pN by

$$\big(r + (p)\big)(x + pN) := rx + pN \; [= r(x + pN)]$$

for all $r \in R$ and $x \in N$, i.e., the $R/(p)$-action and the R-action on N/pN are *compatible* (have the same effect). In particular, N/pN is a vector space over $R/(p)$.

(4) As R-modules and $R/(p)$-vector spaces, we have

$$\big(R/(d)\big)/p\big(R/(d)\big) \cong \begin{cases} R/(p) & \text{if } p \mid d \\ 0 & \text{if } p \nmid d. \end{cases}$$

Using these observations, we show:

Uniqueness: Suppose that

$$R/(d_1) \coprod \cdots \coprod R/(d_r) \coprod R^s \cong M \cong R/(d'_1) \coprod \cdots \coprod R/(d'_{r'}) \coprod R^{s'}$$

with $d_1 \mid \cdots \mid d_r$, the d_i nonzero nonunits and $d'_1 \mid \cdots \mid d'_{r'}$, the $d'_{i'}$ nonzero nonunits, for some integers r, r', s, s'.

We know that a submodule of a free R-module is R-free, hence an isomorphic copy of $R/(d)$ with d a nonzero nonunit in R cannot be a submodule of an R-free. In particular, if there are no d_i or no d'_i, then M is free and there are neither any d_i nor any d'_i. As R is commutative, $s = \operatorname{rank} M = s'$ by Theorem 39.10. So we may assume both $r \geq 1$ and $r' \geq 1$. Since $d_r\big(R/(d_i)\big) = 0$ and $d'_{r'}\big(R/(d'_{i'})\big) = 0$ for all i and i', by Observation 44.8(1), we have

$$R^s \cong d_r d'_{r'} R^s \cong d_r d'_{r'} M \cong d_r d'_{r'} R^{s'} \cong R^{s'}$$

is R-free. Hence $s = s'$ by Theorem 39.10. Let $m \in M$ correspond to a generator of $R/(d'_{r'})$ under the given isomorphism. Then $Rd_r m \subset d_r M \cong R^s$ must also be R-free. As $d'_{r'} d_r m = 0$, we have $d_r m = 0$ (Why?), so $d_r \in \operatorname{ann}_R m = (d'_{r'})$. Similarly, $d'_{r'} \in (d_r)$, so $d_r \approx d'_{r'}$, i.e., $(d_r) = (d'_{r'})$.

Assume that $r \geq r' > 0$. Let p be a prime element (so nonzero) satisfying $p \mid d_1$, which exists as d_1 is a nonzero nonunit. Then by Observations 44.8(2) and (4) and Observation 44.5, we have $M/pM \cong \coprod_{i=1}^{r+s} R/(p)$ is an $R/(p)$-vector space, hence of rank $r + s$. Therefore, $\coprod_{i=1}^{r'} R/(d'_i)/p\big(R/(d'_i)\big) \amalg R^s/pR^s$ is free of rank $r + s$. But, by Observation 44.8(4), we also know that it is free of rank at most $r' + s$. It follows that $r = r'$. In particular, we are done if $p \approx d_r \approx d'_{r'}$. By Observations 44.8(1) and (2), we have

$$\coprod R/(\frac{d_i}{p}) \amalg R^s \cong pM \cong \coprod R/(\frac{d'_i}{p}) \amalg R^s.$$

By induction on the length $l(d_r)$ of d_r, we conclude that $\frac{d_i}{p} \approx \frac{d'_i}{p}$ for all i, hence $d_i \approx d'_i$ for all i. □

Note the proof of the uniqueness statement is similar to the corresponding proof of the Fundamental Theorem of Arithmetic.

Remark 44.9. It should also be noted, that to prove the existence in Theorem 44.6, we do not need to use the existence of a free presentation of M. Indeed the same proof works if we have an exact sequence

$$R^m \xrightarrow{g'} R^n \to M \to 0.$$

But such an exact sequence always exists by Corollary 40.8, since a PID is a Noetherian ring.

Remarks 44.10. Let M be a finitely generated module over a PID.

1. The proof of the uniqueness part of Theorem 44.6 shows, up to isomorphism, M depends only on the cokernel of g in the free presentation (44.7) of M. Also note that $\operatorname{im} g \cong R^m$ is of rank m. This determination of M up to isomorphism is independent of the integers m and n. Indeed given any other free presentation

$$0 \to R^{m'} \xrightarrow{g'} R^{n'} \to M \to 0,$$

we know by Proposition 44.1, that we also must have $m' \leq n'$. Moreover, if we compute the Smith Normal form associated to the map g', we must have $\operatorname{coker} g \cong \operatorname{coker} g'$ by the uniqueness argument in the proof of Theorem 44.6, thus determining M up to isomorphism, for this different n'. The only difference would be in the number of d_i that are units arising in each of the two Smith Normal Forms arising from g and g', respectively, i.e., $|n-m|$ and $|n'-m'|$, respectively; and we are dropping the trivial modules corresponding to such d_i.
2. The only remaining question is what if we have an exact sequence

$$R^m \xrightarrow{g'} R^n \to M \to 0$$

with g' not necessarily an R-monomorphism? In this case, we have exact sequences

$$0 \to \ker g' \to R^m \xrightarrow{g'} R^n \to M \to 0.$$

and

$$0 \to R^m/\ker g' \xrightarrow{\overline{g'}} R^n \to M \to 0$$

with $\overline{g'}$ the canonical epimorphism induced by First Isomorphism Theorem. The second exact sequence is a free presentation since $R^m/\ker g'$ is free of rank at most n and $\operatorname{coker} g' = \operatorname{coker} \overline{g'}$. Indeed, the only difference arising between these two will be zero columns in the Smith Normal form of g' arising from its kernel.

Examples 44.11. 1. We determine the abelian group M, up to isomorphism, generated by x_1, x_2, x_3, x_4 subject to the relations

$$\begin{aligned} 2x_1 + 3x_2 + x_4 &= 0 \\ x_1 - 2x_2 + x_3 &= 0. \end{aligned}$$

(Cf. presentation of groups.) Let $\mathcal{S}_2, \mathcal{S}_4$ be the standard bases for $\mathbb{Z}^2, \mathbb{Z}^4$ respectively. Consider the sequence

$$0 \to \mathbb{Z}^2 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0.$$

with g defined by $g(e_1) = 2e_1 + 3e_2 + e_4$ and $g(e_2) = e_1 - 2e_2 + e_3$ and h defined by $h(e_i)) = x_i$, $i = 1, 2, 3, 4$. Then the sequence is exact (noting $g(e_1)$ and $g(e_2)$ are linearly independent). We have

$$[g]_{\mathcal{S}_2,\mathcal{S}_4} = \begin{pmatrix} 2 & 1 \\ 3 & -2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and $M \cong \operatorname{coker} g$.

To find the isomorphism type of M we find a Smith Normal Form of $[g]_{\mathcal{S}_2,\mathcal{S}_4}$. Using the algorithm in Appendix D (essentially the division algorithm), we row and column reduce $[g]_{\mathcal{S}_2,\mathcal{S}_4}$ to

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence by Remark 44.10(1), M is isomorphic to $\mathbb{Z}/\mathbb{Z} \amalg \mathbb{Z}/\mathbb{Z} \amalg \mathbb{Z} \amalg \mathbb{Z} \cong \mathbb{Z}^2$. We also conclude that

$$0 \to \mathbb{Z}^2 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

is a free resolution of M.

2. We determine the abelian group M, up to isomorphism, generated by x_1, x_2, x_3, x_4 subject to the relations

$$\begin{aligned} 4x_1 + 2x_2 + 4x_3 + 3x_4 &= 0 \\ 2x_1 + 2x_2 - 2x_3 + 2x_4 &= 0 \\ -6x_1 + 6x_3 - 6x_4 &= 0 \end{aligned}$$

Using the notation of the previous example, we have an exact sequence

$$0 \to \mathbb{Z}^3 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

with

$$[g]_{\mathcal{S}_3,\mathcal{S}_4} = \begin{pmatrix} 4 & 2 & -6 \\ 2 & 2 & 0 \\ 4 & -2 & 6 \\ 3 & 2 & -6 \end{pmatrix}$$

and $h(e_i) = x_i$ for $i = 1, 2, 3, 4$.

Row and column reducing $[g]_{\mathcal{S}_3,\mathcal{S}_4}$ using the division algorithm as in Appendix D, we see that a Smith Normal Form for $[g]_{\mathcal{S}_3,\mathcal{S}_4}$ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and } h \text{ takes } e_i \mapsto x_i, \ 1 \le i \le 4,$$

Hence, by Remark 44.10(1), M is isomorphic to

$$\mathbb{Z}/\mathbb{Z} \coprod \mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/6\mathbb{Z} \coprod \mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/6\mathbb{Z} \coprod \mathbb{Z}.$$

Note that this is also isomorphic to $\mathbb{Z}/2\mathbb{Z} \amalg \mathbb{Z}/2\mathbb{Z} \amalg \mathbb{Z}/3\mathbb{Z} \amalg \mathbb{Z}$. This last isomorphism is what the second form of the Fundamental Theorem will give. We also conclude that

$$0 \to \mathbb{Z}^3 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

is a free resolution of M.

We next look at a property of modules that is, in general, weaker than being free but for finitely generated modules over a PID turn out to be equivalent.

Definition 44.12. Let R be a domain an M and R-module. An element m in M is called an R-*torsion* element if there exists a nonzero element r in R satisfying $rm = 0$, i.e., $\operatorname{ann}_R m > 0$. Set

$$M_t := \{m \in M \mid m \text{ is an } R\text{-torsion element}\}.$$

We say that M is a *torsion* R-module if $M = M_t$ and a *torsion-free* R-module if $M_t = 0$.

The proof of the following is straight-forward and left as an exercise.

Properties 44.13. Let R be a domain, M an R-module, and m a nonzero element in M. Then

(1) The element m is not an R-torsion element if and only if Rm is torsion-free if and only if Rm is R-free.
(2) M_t is a submodule of M.
(3) M/M_t is a torsion-free R-module.
(4) If M is R-free, then it is R-torsion-free.

Remarks 44.14. 1. Any finite abelian group is a $\mathbb{Z}$-torsion module.
2. If S is a domain and $R \subset S$ is a subring, then S is an R-torsion-free module.
3. $\mathbb{Q}$ is $\mathbb{Z}$-torsion-free but not $\mathbb{Z}$-free.

We shall show that over a PID finitely generated torsion-free modules are free. The following computation together with the Fundamental Theorem 44.6 will ensure this.

Example 44.15. Let R be a PID and $d_1 \mid \cdots \mid d_r$ nonzero nonunits in R. Set

$$\begin{aligned} N &= R/(d_1) \coprod \cdots \coprod R/(d_r) \coprod R^s \quad \text{and} \\ N_0 &= R/(d_1) \coprod \cdots \coprod R/(d_r), \end{aligned}$$

so $N = N_0 \amalg R^s$. We have

$$d_r N = d_r N_0 \coprod d_r R^s = d_r R^s \cong R^s$$

is R-free of rank s and $N_0 \subset N_t$. Next, suppose that $(n_0, v) \in N_t$ with $n_0 \in N_0$ and $v \in R^s$. Then there exists a nonzero element a in R satisfying $a(n_0, v) = 0$, i.e., $an_0 = 0$

and $av = 0$. As R^s is R-free, it is R-torsion-free, so $v = 0$ and we conclude that $N_0 = N_t$. As the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N_0 & \longrightarrow & N & \longrightarrow & R^s & \longrightarrow & 0 \\ & & \| & & \| & & \uparrow & & \\ 0 & \longrightarrow & N_t & \longrightarrow & N & \longrightarrow & N/N_t & \longrightarrow & 0. \end{array}$$

has exact rows and is commutative, the vertical right hand arrow exists, (why?), hence is an isomorphism, so $N/N_t \cong R^s$. In particular, N is R-free if and only if $N_0 = 0$ if and only if $N_t = 0$ if and only if N is R-torsion-free.

The example together with the Fundamental Theorem yields:

Corollary 44.16. *Let R be a PID and M a finitely generated R-module. Then*

(1) M is R-torsion-free if and only if M is R-free.

(2) $M = M_t \oplus P$ for some finitely generated free R-module $P \subset M$ with $\operatorname{rank} P$ *unique.*

Remark 44.17. $\mathbb{Q}$ is a torsion-free abelian group but not $\mathbb{Z}$-free, so the assumption in the corollary that the module be finitely generated torsion-free is necessary in general to conclude that it is free. [The abelian group $\prod_{i=1}^{\infty} \mathbb{Z}$ is another example of a torsion-free abelian group that is not $\mathbb{Z}$-free.]

We turn to establishing an alternate form of the Fundamental Theorem. We must do some preliminary work. We begin with another example.

Example 44.18. Let R be a PID and M a nontrivial cyclic R-module. Then $M = Rm \cong R/(d)$, for some $m \in M$ and unique $(d) = \operatorname{ann}_R m < R$. Assume that M is not R-free. Then $(d) > 0$ and d is not a unit. Let $d = p_1^{e_1} \cdots p_r^{e_r}$ be a factorization, with $p_1, \ldots, p_r$ non-associative irreducible, hence prime, elements and $e_1, \ldots, e_r$ positive integers. Since R is a PID, we know that the ideals $(p_i^{e_i})$, $i = 1, \ldots, r$, are all comaximal ideals. In particular, by the Chinese Remainder Theorem, $R/(d) \cong R/(p_1^{e_1}) \times \cdots \times R/(p_r^{e_r})$ as rings. It follows that $R/(d) \cong R/(p_1^{e_1}) \amalg \cdots \amalg R/(p_r^{e_r})$ as R-modules, as the R-action on $R/(a)$ is compatible with the $R/(a)$-action for any $a \in R$. (Cf. Observation 44.8(3).) Therefore, there exist $m_1, \ldots, m_r$ in M satisfying $M = Rm_1 \oplus \cdots \oplus Rm_r$ with $\operatorname{ann}_R m_i = (p_i^{e_i})$ for $i = 1, \ldots, r$.

Although this example suffices to get our alternate form, we first make some further comments. If R is a domain, p an irreducible element in R, and M an R-module, we say that M is *p-primary* if for all x in M, there exists a positive integer n (depending on x) satisfying $p^n x = 0$. If R is Noetherian, let $\mathcal{P}$ denote a system of representatives of equivalence classes of irreducible elements in R under the equivalence relation $\approx$.

Example 44.19. Let R be a Noetherian domain and $p \in \mathcal{P}$. Then $R/(p^r)$ is a p-primary R-module for all positive integers r.

Theorem 44.20. (Primary Decomposition Theorem) *Let R be a PID and M a torsion R-module. If $p \in \mathcal{P}$, let*

$$M_p := \{x \in M \mid p^r x = 0 \text{ for some positive integer } r\}.$$

Then M_p is a p-primary submodule of M and $M = \bigoplus_{\mathcal{P}} M_p$. Moreover, if M is finitely generated, then $M_p = 0$ for almost all p in $\mathcal{P}$.

Proof. Clearly, M_p is a p-primary submodule of M for all $p \in \mathcal{P}$. We may assume that M is not trivial.

$M = \sum_{\mathcal{P}} M_p$: Let x be a nonzero element of M and $(d) = \operatorname{ann}_R x$. Then d is nonzero nonunit as M is a torsion R-module. It follows by Example 44.18 that x lies in $\sum_{\mathcal{P}} M_p$. In fact, it follows that x lies in $\sum_{\substack{p \mid d \\ p \in \mathcal{P}}} M_p$.

$M = \bigoplus_{\mathcal{P}} M_p$: Suppose that $x \in M_{p_0} \cap \sum_{p_0 \neq p \in \mathcal{P}} M_p$. Let $(d) = \operatorname{ann}_R x$, so d is not zero. We have $p_0^{e_0} x = 0$ for some positive integer e_0, as $x \in M_{p_0}$, so $p_0^{e_0} \in \operatorname{ann}_R x$, i.e., $d \mid p_0^{e_0}$. As $x \in \sum_{p_0 \neq p \in \mathcal{P}} M_p$, there exist $p_1, \ldots, p_n$ in $\mathcal{P}$ and positive integers $e_1, \ldots, e_n$ satisfying $x = x_1 + \cdots + x_n$ with $p_i^{e_i} x_i = 0$ for $i = 1, \ldots, n$. It follows that $p_1^{e_1} \cdots p_n^{e_n} x = 0$, so $p_1^{e_1} \cdots p_n^{e_n} \in \operatorname{ann}_R x$, i.e., $d \mid p_1^{e_1} \cdots p_n^{e_n}$. By choice p_0 and $p_1^{e_1} \cdots p_n^{e_n}$ are relatively prime in the PID R. It follows that d is a unit hence $x = 0$.

Finally, suppose that M is finitely generated, say $M = Rx_1 + \cdots + Rx_n$ with x_i nonzero for $i = 1, \ldots, n$. Let d_i be a nonzero nonunit satisfying $(d_i) = \operatorname{ann}_R x_i$ for $i = 1, \ldots, n$ and set $d = d_1 \cdots d_n$ a nonzero element in the domain R. It follows that $M_p = 0$ for all $p \in \mathcal{P}$ satisfying $p \nmid d$. Hence $M = \bigoplus_{\substack{p \in \mathcal{P} \\ p \mid d}} M_p$, a finite sum. □

Example 44.21. Suppose that R is a PID and M is a nontrivial cyclic R-module with $\operatorname{ann}_R x = (d)$ and $d = p_1^{e_1} \cdots p_r^{e_n}$, with $p_i \in \mathcal{P}$. Then $M = \bigoplus_{i=1}^n M_{p_i}$ and $M_{p_i} \cong R/(p_i^{e_i})$ for $i = 1, \ldots, n$.

Using this, together with the Fundamental Theorem 44.6, we can reformulate it as

Theorem 44.22. (Fundamental Theorem of fg Modules over a PID, Form II) *Let R be a PID and M a nontrivial finitely generated R-module. Then there exists a finitely generated free submodule P of M of unique rank and $p_i \in \mathcal{P}$, $i = 1, \ldots, r$ some r and x_{ij} in M, $1 \leq j \leq n_i$ and $1 \leq i \leq r$ and for some n_i, satisfying*

$$M = P \oplus \left(\bigoplus_{i=1}^{r} \bigoplus_{j=1}^{n_i} Rm_{ij}\right) \quad \textit{satisfying}$$

and for all i and j, we have

(1) $Rm_{ij} \cong R/(p_i^{e_{ij}})$ and

(2) $1 \leq e_{i1} \leq e_{i2} \leq \cdots \leq e_{in_i}$.

Moreover, this decomposition is unique relative the ordering of the p_i subject to the order of the e_{ij} above and the uniqueness of the rank of P.

Generators of the ideals $(p_i^{e_{ij}})$ in the theorem are called *elementary divisors* of M. They are unique up to units, and together with the rank of P completely determine M up to isomorphism. The condition on the e_{ij}, of course, orders the set of cyclic modules that are p_i-primary submodules in the decomposition (which are unique up to isomorphism). This data completely determines the isomorphism class of M.

Of course, both forms of the Fundamental Theorem apply to finitely generated abelian groups, although in this case an easier proof will suffice. If we are given any finitely generated abelian group defined by generators and relations, the invariant factors will determine completely the group up to isomorphism. Moreover, given a positive integer n, up to isomorphism, we can write down every abelian group of order n (assuming that we can factor n). We give an example how this works.

Example 44.23.
Find all abelian groups up to isomorphism of order $2^3 \cdot 3^2 \cdot 5$. We write these groups in both forms. Given the factorization, it is easier to write down the representatives of isomorphism classes in Form II. To get Form I, we take the biggest p-primary pieces for each prime p and combine them to get one term. Then continue with what is left. This is illustrated in our example by the overlines and underlines of various p-primary pieces and what they become in Form I.

$$\begin{aligned}
\mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/2\mathbb{Z}} \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/6\mathbb{Z}} \times \underline{\mathbb{Z}/30\mathbb{Z}} \\
\mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/90\mathbb{Z}} \\
\overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/4\mathbb{Z}} \times \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \overline{\mathbb{Z}/6\mathbb{Z}} \times \underline{\mathbb{Z}/60\mathbb{Z}} \\
\overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/4\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/180\mathbb{Z}} \\
\underline{\mathbb{Z}/8\mathbb{Z}} \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/120\mathbb{Z}} \\
\underline{\mathbb{Z}/8\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} &\cong \underline{\mathbb{Z}/360\mathbb{Z}}.
\end{aligned}$$

[Note: Abelian direct sums of abelian groups are usually written as products as a finite product of abelian groups is the same as a direct sum of the same groups.]

Exercises 44.24.

1. Let R be a PID. Suppose that M is a finitely generated R-module and
$$0 \to R^m \xrightarrow{g} R^n \to M \to 0$$
is a free presentation of M with n minimal. Without using the Fundamental Theorem (or Smith Normal Forms) show if
$$0 \to R^{m'} \xrightarrow{g'} R^{n'} \to M \to 0$$
is another free presentation of M, then $m' \leq m$.
2. Prove the following generalization of Proposition 44.1 using Exercises 39.12(126.3), (126.5)): Let R be a commutative (this is not needed) ring with the property that every (left) ideal of R is R-projective. Let M be an R-module isomorphic to a submodule of R^n. Then M is a direct sum of $m \leq n$ submodules each isomorphic to a (left) ideal of R. In particular M is R-projective.
3. Prove Observation 44.5.

4. Prove all the statements in Observations 44.8.
5. Prove all the properties in 44.13.
6. Let R be a domain. Show every ideal in R is R-torsion-free, but is free if and only if it is principal. In particular, show that R is a PID if every submodule of a free module is free. [The converse is also true.]
7. Let R be a domain and M a finitely generated R-module. Prove that M is isomorphic to a free submodule of finite rank.
8. Let $f : \mathbb{Z}^n \to \mathbb{Z}^m$ be a $\mathbb{Z}$-module homomorphism. Let $\mathcal{S}_l$ be the standard basis for $\mathbb{Z}^l$. Prove that f is monic if and only if the rank of $[f]_{\mathcal{S}^n,\mathcal{S}^m}$ is n and f is epic if and only if a gcd of the mth ordered minors of $[f]_{\mathcal{S}^n,\mathcal{S}^m}$ is 1.
9. Let R be a commutative ring. Let $\mathrm{E}_n(R)$ be the subgroup of $\mathrm{GL}_n(R)$ generated by all matrices of the form $I + \lambda$ where λ is a matrix with precisely one non zero entry and this entry does not occur on the diagonal. Suppose that R is a euclidean ring. Show that $\mathrm{SL}_n(R) = \mathrm{E}_n(R)$. (Cf. Appendix D.)
10. Let A be a finite abelian group and let
$$\hat{A} := \{\chi : A \to \mathbb{C}^\times \mid \chi \text{ a group homomorphism}\}.$$
It is easily checked that $\hat{A}$ is a group via $\chi_1\chi_2(x) := \chi_1(x)\chi_2(x)$. Show
    - (i) A and $\hat{A}$ have the same order and, in fact, are isomorphic.
    - (ii) If χ is not the identity element of $\hat{A}$ then $\sum_{a\in A}\chi(a) = 0$.
11. Determine all abelian groups of order 400 up to isomorphism.

45. Canonical Forms for Matrices

In this section, we use the Fundamental Theorem of fg Modules over a PID to determine a good system of representatives for matrices over a field under the equivalence relation of similarity. Recall if V is a finite dimensional vector space over a field F with basis $\mathcal{B}$ and $T : V \to V$ is a linear operator then the *characteristic polynomial* of T is defined to be
$$f_T := \det(tI - [T]_{\mathcal{B}}) \in F[t].$$
This is independent of bases by the Change of Basis Theorem (cf. Appendix C). Roots of f_T are called *eigenvalues* of T. An element λ is a root of f_T if and only if there exists a nonzero vector v in V such that $T(v) = \lambda v$. Such a v is called a (nonzero) *eigenvector* for T.

Construction 45.1. Let V be a finite dimensional vector space of dimension n over a field F and $T : V \to V$ a linear operator (endomorphism). Define
$$\varphi : F[t] \to \mathrm{End}_F(V) \text{ be given by } f \mapsto f(T)$$
evaluation at T. This is a ring homomorphism with
$$\mathrm{im}\,\varphi = \{\sum_{1=0}^{r} a_iT^i \mid a_i \in F \text{ some } m\},$$

a commutative subring of $\operatorname{End}_F(V)$. By the First Isomorphism Theorem $F[t]/\ker\varphi \cong \operatorname{im}\varphi = F[T]$. As $F[t]$ is a vector space over F on basis $\{t^i \mid i \geq 0\}$, we have $\dim_F F[t]$ is infinite. We also know $\dim_F F[T] \leq \dim_F \operatorname{End}_F(V) = \dim_F \mathbb{M}_n(F) = n^2$ is finite. Thus φ cannot be monic. Consequently, $0 < \ker\varphi$. As φ is a ring homomorphism, $\varphi(1) = 1_V$, so φ is not the trivial map. Therefore, $F[t]$ being a PID means:

$$\begin{aligned}\text{There exists }\ &\text{a unique non-constant monic polynomial } q_T \text{ in } F[t]\\ &\text{satisfying } \ker\varphi = (q_T) > 0.\end{aligned}$$

We call q_T the *minimal polynomial* of T over F.

Note: By definition, $q_T(T) = 0$. (This polynomial need not be irreducible.)

The vector space V is an $\operatorname{End}_F(V)$-module via evaluation so becomes an $F[t]$-module via the pullback along φ, i.e., $f \cdot v := \varphi(f)(v) = f(T)(v)$ for all $f \in F[t]$ and all $v \in V$. Thus V is an $F[t]$-module via *evaluation* at T. As V is a finite dimensional vector space over F, it is a finitely generated $F[t]$-module. Moreover, $q_T \neq 0$ in $F[t]$ and $q_T \cdot v = q_T(T)(v) = 0v = 0$ for all v in V, so V is a finitely generated torsion $F[t]$-module. Therefore, the two forms of the Fundamental Theorem of fg Modules over a PID are applicable. We shall use both forms. To use them, we need to investigate cyclic submodules of the $F[t]$-module V.

We shall first apply the first form of the Fundamental Theorem 44.6 to the construction above. We shall use it for the case of an arbitrary field. We begin with some preliminaries.

Lemma 45.2. *Suppose V is a finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then*

$$(q_T) = \operatorname{ann}_{F[t]} V := \{f \in F[t] \mid fv = 0 \text{ for all } v \in V\} = \bigcap_{v \in V} \operatorname{ann}_{F[t]} v.$$

Proof. We know that $q_T v$ is zero for all v in V, so q_T lies in $\operatorname{ann}_{F[t]} V$. Let $\varphi : F[t] \to \operatorname{End}_F(V)$ be evaluation at T. Suppose that $f \in \operatorname{ann}_{F[t]} V$. Then $0 = fv = f(T)(v)$ for all $v \in V$, hence $f(T)$ is the zero endomorphism in $\operatorname{End}_F(V)$, so $f \in \ker\varphi = (q_T)$. The result follows. $\square$

We need another fact from linear algebra.

Definition 45.3. Let F be a field and $h = t^d + a_{d-1}t^{d-1} + \cdots + a_1 t + a_0$ be a monic polynomial in $F[t]$ of positive degree d. Define the *companion matrix* of h to be the matrix in $\mathbb{M}_d(F)$ given by

$$C_h := \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & \cdots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}.$$

[If $h = 1$, we let $C_h = 0$.]

Remarks 45.4. Let F be a field and g, h be monic non-constant polynomials in $F[t]$.

1. $g = h$ if and only if $C_g = C_h$,

2. $h = f_{C_h}$, the characteristic polynomial of the matrix C_h:
If $h = t^d + a_{d-1}t^{d-1} + \cdots + a_0$ with $d > 0$, then

$$f_{C_h} = \det \begin{pmatrix} t & 0 & \cdots & 0 & a_0 \\ -1 & t & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t + a_{d-1} \end{pmatrix}.$$

Using induction and expanding the determinant by minors along the top row yield

$$f_{C_h} = t(t^{d-1} + a_{d-1}t^{d-2} + \cdots + a_1) + (-1)^{d-1}a_0(-1)^{d-1} = h.$$

To apply the Fundamental Theorem 44.6 to a finite dimensional vector space over a field F and linear operator T with $F[t]$-module structure given by evaluation at T, we must compute $F[t]$-cyclic submodules. The key is to determine an appropriate F-basis, for such a submodule.

Proposition 45.5. *Suppose V is a finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be a cyclic $F[t]$-module by evaluation at T, i.e., there exists a vector v in V satisfying $V = F[t]v$. Let $\mathcal{B} = \{v, T(v), \dots, T^{d-1}(v)\}$ with d the degree of the minimal polynomial q_T of T. Then*

(1) $\mathcal{B}$ is an ordered basis for V.
(2) $\operatorname{ann}_{F[t]} v = \operatorname{ann}_{F[t]} V = (q_T)$
(3) The matrix representation $[T]_{\mathcal{B}}$ of T in the $\mathcal{B}$ basis satisfies $[T]_{\mathcal{B}} = C_{q_T}$.
(4) The minimal polynomial q_T satisfies $q_T = f_T$.

Proof. **$\mathcal{B}$ spans V as a vector space over F:** Let w be a vector in $V = F[t]v$, so there exists a polynomial h in $F[t]$ satisfying $w = hv = h(T)(v)$, As $F[t]$ is a euclidean domain with euclidean function the degree, we can write $h = q_T s + r$ for some polynomials s and r in $F[t]$ with $r = 0$ or $\deg r < \deg q_T$. Evaluating this polynomial at T yields

$$w = hv = (q_T s + r)v = q_T(T)s(T)(v) + r(T)(v) = r(T)(v)$$

which lies in the span of $\mathcal{B}$.
$\mathcal{B}$ is linearly independent (hence is a basis for V): Suppose that $\sum_{i=0}^{d-1} a_i T^i(v) = 0$ for some a_i in F. Set $g = \sum_{i=0}^{d-1} a_i t^i$ in $F[t]$. Then $gv = g(T)(v) = 0$, so $gfv = fgv = 0$ in $V = F[t]v$ for all polynomials f in $F[t]$, i.e., $gw = 0$ for all w in V. By Lemma 45.2, we have $g \in \operatorname{ann}_{F[t]} V = \operatorname{ann}_{F[t]} v = (q_T)$, so $q_T \mid g$. As $g = 0$ or $\deg g < \deg q_T$, we must have $g = 0$, i.e., $a_i = 0$ for all i. The proof of (3) is simple and (4) follows from Remarks 45.4. $\square$

We now apply the Fundamental Theorem 44.6.

Theorem 45.6. (Rational Canonical Form) *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then there exists a direct sum decomposition of $F[t]$-modules*

$$V = V_1 \oplus \cdots \oplus V_r$$

for some r with each V_i a cyclic $F[t]$-module, $1 \le i \le r$, and satisfying:

(1) Let q_i be the monic non-constant polynomial in $F[t]$ satisfying $(q_i) = \operatorname{ann}_{F[t]} V_i$ for $1 \leq i \leq r$. Then the q_i satisfy $q_1 \mid \cdots \mid q_r$ in $F[t]$ and are unique relative to this ordering and are invariant factors of V as an $F[t]$-module.
(2) The polynomial q_r in (1) is the minimal polynomial q_T of T, and $(q_T) = \operatorname{ann}_{F[t]} V$.
(3) Let q_i, $i = 1, \ldots, r$ be as in (1). Then $f_T = q_1 \cdots q_r$.
(4) There exist ordered bases $\mathcal{B}_i$ for V_i, $1 \leq i \leq r$ satisfying:
(a) $T|_{V_i}$ lies in $\operatorname{End}_F(V_i)$ for $i = 1, \ldots, r$.
(b) $[T|_{V_i}]_{\mathcal{B}_i} = C_{q_i}$, for $i = 1, \ldots, r$.
(c) $q_i = f_{T|_{V_i}} = q_{T|_{V_i}}$.
(d) Let $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is an ordered basis for V and

$$[T]_{\mathcal{B}} = \begin{pmatrix} C_{q_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{q_r} \end{pmatrix} \text{ (in block form.)}$$

This matrix is unique relative to the monic polynomials in (1) *as ordered there.*

The monic polynomials, $q_1, \ldots, q_r$ in (1) in the theorem are called *the invariant factors* of T and the matrix in (4) is called the *rational canonical form* of T.

Proof. Apply the Fundamental Theorem 44.6 and Proposition 45.5 after noting that V_i being $F[t]$-cyclic implies that $\operatorname{ann}_{F[t]} V_i = (q_i) = (q|_{V_i})$ and $T|_{V_i} : V_i \to V_i$ as V_i is an $F[t]$-module, i.e., V_i is T-invariant as $tv = T(v)$ for all $v \in V_i$ for $i = 1, \ldots, r$. □

Examples 45.7. Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator such that $T^m = 0$ for some $m > 0$ (we say that T is *nilpotent*), then the rational canonical form of T is upper triangular with diagonal entries all 0.

An immediate consequence of Theorem 45.6 is the following:

Corollary 45.8. *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then $q_T \mid f_T$ in $F[t]$. In particular, q_T and f_T have the same roots in F and q_T is a product of linear polynomials in $F[t]$ if and only if f_T is a product of linear polynomials in $F[t]$.*

Of course, the multiplicity of a root of q_T and f_T may be different. [Recall from linear algebra that the roots of f_T are the eigenvalues of T.]

An important consequence of Theorem 45.6 is the well-known:

Corollary 45.9. (Cayley-Hamilton Theorem) *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Then the characteristic polynomial, f_T in $F[t]$, satisfies $f_T(T) = 0$, i.e., $f_T \in \operatorname{ann}_{F[t]} V = (q_T)$. In particular, $q_T \mid f_T$ in $F[t]$ and q_T and f_T have the same irreducible factors (although not necessarily with the same multiplicities).*

The rational canonical forms of a matrices in $\mathbb{M}_n(F)$ give an excellent system of representatives of similarity classes of such matrices. To establish this, we need the following observation.

Observation 45.10. If R is a domain, A a matrix in $\mathbb{M}_n(R)$ with $\det A$ nonzero, and $AX = 0$ (X and 0 in $R^{n\times 1}$), then $X = 0$, i.e., A is an R-monomorphism when viewed as a map $R^{n\times 1} \to R^{n\times 1}$. [This follows as A is invertible in $\mathbb{M}_n\big(qf(R)\big)$.]

Theorem 45.11. (Classification of Similarity Classes of Matrices over a Field) *Let F be a field and A and B be two $n \times n$ matrices in $\mathbb{M}_n(F)$.*

(1) There exists a unique matrix C in $\mathbb{M}_n(F)$ in rational canonical form with A and C similar.

(2) The following are equivalent:

(a) $A \sim B$.

(b) A and B have the same rational canonical form.

(c) A and B have the same invariant factors.

(d) $tI - A$ and $tI - B$ have the same Smith Normal Form when we choose monic invariants and the nonidentity monic invariant factors are the same as those for A and B in (c).

(e) $tI - A$ and $tI - B$ are equivalent in $\mathbb{M}_n(F[t])$.

Proof. Let

$$\mathcal{S}_{n,1} = \{e_1, \dots, e_n\} \text{ with } e_i = \begin{pmatrix} 0 & 0 & \underset{i}{1} & 0 & \cdots & 0 \end{pmatrix}^t$$

be the standard basis for $F^{n\times 1}$ and also for $F[t]^{n\times 1}$ as free modules over F and $F[t]^{n\times 1}$, respectively. [Aside: $\mathcal{C} = \{t^j e_i \mid 1 \le i \le n,\ j \ge 0\}$ is an F-basis for $F[t]^{n\times 1}$. The linear operator $A : F^{n\times 1} \to F^{n\times 1}$ given by $v \mapsto Av$ satisfies $A = [A]_{\mathcal{S}_{n,1}}$, so there exists a basis $\mathcal{B}$ for $F^{n\times 1}$ such that $[A]_{\mathcal{B}}$ is in rational canonical form by the Change of Basis Theorem (cf. Appendix C). Therefore, we have (1) and the equivalence of (a), (b), and (c) in (2) follow easily.

By the Universal Property of Free Modules 39.3, there exists a unique $F[t]$-endomorphism (hence also F-linear) $g : F[t]^{n\times 1} \to F[t]^{n\times 1}$ extending $e_j \mapsto te_j - Ae_j$. So $g = tI - A$. We, therefore, have a sequence

$$0 \to F[t]^{n\times 1} \xrightarrow{g} F[t]^{n\times 1} \xrightarrow{e_A} F^{n\times 1} \to 0 \tag{45.12}$$

with e_A evaluation at A, i.e., defined by $fe_j \to f(A)e_j$ for all $f \in F[t]$ and all j. We shall show that this sequence is an exact sequence.

As $\det(tI_n - A) = f_A$, the characteristic polynomial of the matrix A, hence nonzero, we have g an $F[t]$-monomorphism by Observation 45.10 above. It is also clear that the evaluation map e_A is a surjection. So we must show exactness at the middle term, i.e., $\operatorname{im} g = \ker e_A$.

Clearly, $\operatorname{im} g \subset \ker e_A$. Therefore, $x + \operatorname{im} g \mapsto x + \ker e_A$ defines an an $F[t]$-epimorphism $F[t]^{n\times 1}/\operatorname{im} g \to F[t]^{n\times 1}/\ker e_A$. As we want to show $\operatorname{im} g = \ker e_A$, it suffices to show that this map is an isomorphism. Composing this map with $\overline{e_A} : F[t]^{n\times 1}/\ker e_A \to F^{n\times 1}$, the map induced by e_A given by the First Isomorphism Theorem, defines an $F[t]$-epimorphism $\widetilde{e_A} : \operatorname{coker} g \to F^{n\times 1}$, so it suffices to show that $\widetilde{e_A}$ is an isomorphism, i.e., a bijection.

By Remark 44.10(1), a Smith Normal Form S of the matrix representation $tI_n - A$ of g determines $\operatorname{coker} g$ as an $F[t]$-module. We may assume that diagonal matrix S was chosen with its nonzero diagonal entries monic polynomials. As $\det(tI_n - A) = f_A \neq 0$, S has no nonzero diagonal entries.

Let the non-constant monic diagonal entries of S be $d_1|\cdots|d_r$. We know that $S = P[g]_{\mathcal{S}_{n,1}}Q$ with $P, Q \in \mathrm{GL}_n(F[t])$, and an invertible matrix in $\mathrm{GL}_n(F[t])$ has determinant in $F[t]^\times = F^\times$. (This also follows from the algorithm given in Appendix D, Theorem D.2, since Type I, II, and III matrices have determinant in $F^\times$). Consequently, we have $f_A = d_1 \cdots d_r$ as f_A and all the d_i are monic. Hence

$$\begin{aligned}\dim_F \operatorname{coker} g &= \sum_{i=1}^{r} \dim_F F[t]/(d_i)\\ &= \sum_{i=1}^{r} \deg d_i = \deg f_A = n = \dim_F F^{n\times 1}\end{aligned}$$

using the isomorphism we showed when proving the Fundamental Theorem 44.6 and Proposition 34.14. It follows that the $F[t]$-epimorphism hence F-epimorphism $\widetilde{e_A}$ is an F-isomorphism, hence $\operatorname{im}(g) = \ker e_A$ and the sequence in 45.12 is exact.

It follows from the exactness of 45.12 that the non-constant invariant factors of a Smith Normal Form S determines the invariants of $F^{n\times 1}$ as an $F[t]$-module which therefore must be the same as those determined by the invariants of A. The equivalence of (c), (d), and (e) follow.

□

Remark 45.13. As pointed out in the proof above, the Smith Normal Form for a matrix $A \in \mathbb{M}_n(F)$ is determined by the Smith Normal Form of the matrix $tI - A \in \mathbb{M}_n(F[t]$. In particular, the Smith Normal form of a linear operator on a finite dimensional F-vector space can be computed using the algorithm given in Appendix D, Theorem D.2, hence the invariant factors of A.

Corollary 45.14. *Let F be a field and $A \in \mathbb{M}_n(F)$. Then A is similar to its transpose A^t.*

Proof. It follows by Theorem 43.8 (or by the algorithm given in the proof of Theorem D.2 in Appendix D) that a matrix and its transpose in $\mathbb{M}_n(F[t])$ have equivalent Smith Normal Forms. In particular, the matrices $tI - A$ and $tI - A^t$ have the same Smith Normal Form when we choose monic invariants. □

Using the proof above, we can give a summary of what is going on. Instead of giving this summary using matrices, we use linear operators.

Summary 45.15. Let V be a finite dimensional vector space over F of dimension n and $T : V \to V$ a linear operator. We view V as an $F[t]$-module by $tv := T(v)$ for all $v \in V$. We then have an exact sequence

$$0 \to F[t]^n \xrightarrow{t1_{F[t]^n} - T} F[t]^n \xrightarrow{e_T} V \to 0 \tag{45.16}$$

with the $F[t]$-homomorphism e_T defined by $\sum f_i e_i \mapsto \sum f_(T)e_i$, where e_i in the ith basis element in the (ordered) standard basis $\mathcal{S}_n$ for $F[t]^n$ (and F^n). We call the sequence 45.16 the *characteristic sequence* of T. The matrix $t1_{F[t]^n} - [T]_{\mathcal{S}_n}$ is called the *characteristic* *matrix* of $[T]_{\mathcal{S}_n}$. It has nonzero determinant f_T, the characteristic polynomial of T (and is the reason that $t1_{F[t]^n} - [T]_{\mathcal{S}_n}$ is a monomorphism). In particular, the characteristic matrix of $[T]_{\mathcal{S}_n}$ is nonsingular, so has a unique Smith Normal Form of $(tI_n - [T]_{\mathcal{S}_n})$ given by $\mathrm{diag}(1, \dots, q_1, \dots, q_r)$ has the q_i monic non-constant polynomials satisfying $q_1 \mid \cdots \mid q_r$. (The entries of a Smith Normal Form are unique up to units.)

As the sequence 45.16 is exact, we have

$$\begin{aligned}\mathrm{coker}(t1_{F[t]^n} - T) &= F[t]^n / \mathrm{im}(t1_{F[t]^n} - T)\\ &= F[t]^n) / \ker(t1_{F[t]^n} - T) \cong V.\end{aligned}$$

By the Fundamental Theorem, we can decompose the finitely generated torsion $F[t]$-module V into cyclic $F[t]$-submodules, say

$$V = V_1 \coprod \cdots \coprod V_r$$

with $V_i = F[t]v_i$ some v_i for each i and satisfying

$$V_i \cong F[t]/\mathrm{ann}_{F[t]}(v_i) = F[t]/\mathrm{ann}_{F[t]}(V_i) = F[t]/(q_i),\ i = 1, \dots, r$$

with the descending chain of ideals

$$(q_1) \supset \cdots \supset (q_r)$$

unique. Moreover,

$$q_i = q_{T|_{V_i}},\ \ q_r = q_T,\ \ f_T = q_1 \dots q_r,$$

and

$$RCF(T) := RCF([T]_{\mathcal{S}_n} = \begin{pmatrix} C_{q_1} & & \\ & \ddots & \\ & & C_{q_r} \end{pmatrix}.$$

In particular, to find the invariant factors of T one computes the Smith Normal Form of the characteristic matrix using the algorithm given in the Appendix D.

We turn to the second form of the Fundamental Theorem. It is immediate that we have the following:

Theorem 45.17. *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then there exists a direct sum decomposition of $F[t]$-modules*

$$V = V_{11} \oplus \cdots V_{1r_1} \oplus \cdots \oplus V_{r1} \oplus \cdots \oplus V_{rn_r}$$

for some r and r_i with each V_{ij} a cyclic $F[t]$-module, $1 \le i \le r$ and $1 \le j \le i_r$, satisfying:

(1) $V_{ij} \cong F[t]/(q_i^{e_{ij}})$ with $q_{ij} \in F[t]$ monic.
(2) $1 \le e_{i1} \le e_{i2} \le \cdots \le e_{in_i}$.
(3) $(q_i^{e_{ij}}) = \mathrm{ann}_{F[t]} V_{ij}$ for $1 \le i \le r$ and $1 \le e_{i1} \le e_{i2} \le \cdots \le e_{in_i}$, are unique relative to the ordering in (2), and are the elementary divisors of V as an $F[t]$-module.
(4) The polynomial $q_i^{e_{ij}}$ satisfy $q_i^{e_{ij}} = q_T|_{V_{ij}} = f_T|_{V_{ij}}$ $1 \le i \le r$, $1 \le e_{i1} \le e_{i2} \le \cdots \le e_{in_i}$.
(5) $f_T = \prod_{i,j} q_i^{e_{ij}}$.
(6) There exist ordered bases $\mathcal{B}_{ij}$ for V_{ij}, $1 \le i \le r$ and $1 \le e_{i1} \le e_{i2} \le \cdots \le e_{in_i}$ satisfying:

(a) $T|_{V_{ij}}$ *lies in* $\operatorname{End}_F(V_{ij})$ *for* $i = 1, \ldots, r$ *and* $1 \leq e_{i1} \leq e_{i2} \leq \cdots \leq e_{in_i}$.
(b) $[T|_{V_{ij}}]_{\mathcal{B}_{ij}} = C_{q_{ij}}$, *for* $i = 1, \ldots, r$ *and* $1 \leq e_{i1} \leq e_{i2} \leq \cdots \leq e_{in_i}$.
(c) Let $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$. *Then* $\mathcal{B}$ *is an ordered basis for* V, *and*

$$[T]_{\mathcal{B}} = \begin{pmatrix} C_{q_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{q_{rn_r}} \end{pmatrix} \text{ (in block form.)}$$

This matrix is unique relative to the monic polynomials in (1) as ordered in (2) up to a permutation of the blocks determined by $\prod_{j=1}^{n_i} q_{ij}$.

We now shall apply the second form for matrices in the more interesting case when the field is an algebraically closed field, i.e., when non-constant polynomials always factor into a product of linear polynomials, as it is this case that the elementary divisors arising from this form are useful. Instead of the form above, we shall get a more useful form. We first establish the cyclic case when the minimal polynomial is a power of a linear polynomial.

Lemma 45.18. *Suppose* V *is a finite dimensional vector space over* F *and* $T : V \to V$ *a linear operator. Let* V *be a cyclic* $F[t]$*-module by evaluation at* T, *i.e., there exists a vector* v *in* V *satisfying* $V = F[t]v$. *Suppose that the minimal polynomial* q_T *of* T *is* $(t-a)^d$ *in* $F[t]$. *Then* $\mathcal{B} = \{v, (T-a)(v), \ldots, (T-a)^{d-1}(v)\}$ *is an ordered basis for* V *and*

$$[T]_{\mathcal{B}} = \begin{pmatrix} a & 0 & 0 & \cdots & 0 \\ 1 & a & 0 & \cdots & \\ 0 & 1 & a & & \\ \vdots & & \ddots & \ddots & \\ 0 & & \cdots & 1 & a \end{pmatrix} \text{ in } \mathbb{M}_d(F).$$

Proof. We know that $\mathcal{C} = \{v, T(v), \ldots, T^{d-1}(v)\}$ is an F-basis for V by our previous work. If $\mathcal{B}$ is linearly dependent, then there exists an equation

$$(*) \qquad 0 = \sum_{i=0}^{d-1} a_i (T-a)^i v \text{ for some } a_i \text{ in } F, \text{ not all zero.}$$

Choose N maximal such that $a_N \neq 0$ and expand (*) to get

$$0 = a_N T^N v + \sum_{i=0}^{N-1} b_i T^i v \text{ for some } b_i \text{ in } F.$$

As $\mathcal{C}$ is linearly independent, $a_N = 0$, a contradiction. Therefore, $\mathcal{B}$ is linearly independent hence a basis as $|\mathcal{B}| = |\mathcal{C}|$. We have

$$0 = q_r(T) = (T - a1_V)^r \text{ and}$$
$$T(T-a1_V)^i = \big((T-a1_V)+a\big)(T-a1_V)^i = (T-a1_V)^{i+1} + a(T-a1_V)^i$$

for $i = 0, \ldots r-1$. So

$$T(T-a)^i(v) = (T-a)^{i+1}(v) + a(T-a)^i(v) \text{ for } i = 0, \ldots, r-1$$

and the last statement follows. □

When the minimal polynomial of a linear operator on a finite dimensional vector space factors into a product of linear terms, the Fundamental Theorem 44.22 and the last lemma now immediately implies the following:

Theorem 45.19. (Jordan Canonical Form) *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Suppose that q_T factors into linear terms in $F[t]$. Then there exists an ordered basis $\mathcal{B}$ for V satisfying*

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_r \end{pmatrix} \quad \text{with} \quad A_i = \begin{pmatrix} a_i & 0 & & \cdots & & 0 \\ 1 & a_i & 0 & & & 0 \\ 0 & 1 & a_i & 0 & & \vdots \\ \vdots & & & \ddots & & \\ & & & 1 & a_i & 0 \\ 0 & & \cdots & & 1 & a_i \end{pmatrix}$$

for $i = 1, \dots, r$. The a_i, $i = 1, \dots, r$ are the eigenvalues of T (with $a_i = a_j$, $i \neq j$ possible). This matrix of these blocks is unique up to the order of the blocks. The monic polynomials f_{A_i}, $1 \leq i \leq r$, are the elementary divisors of V with the above $F[t]$-module structure and the characteristic polynomial of T satisfies $f_t = \prod_{i=1}^{r} f_{A_i}$

In the theorem, the blocks A_i are called *Jordan blocks* and, as the monic elementary divisors f_{A_i} are unique, they are called the *elementary divisors* of T. The matrix representing T in 45.19 is called the *Jordan canonical form* of T.

Corollary 45.20. *Let A and B lie in $\mathbb{M}_n(F)$. If each of the characteristic polynomials f_A and f_B factor into a product of linear terms in $F[t]$, then A and B are similar in $\mathbb{M}_n(F)$ if and only if A and B have the same Jordan canonical form (up to block order) if and only if they have the same elementary divisors (counted with multiplicity). In particular, if F is an algebraically closed field, then a matrix in $\mathbb{M}_n(F)$, $n \in \mathbb{Z}^+$, is determined up to similarity by its Jordan Canonical Form.*

Remark 45.21. The rational canonical form of a matrix is quite computable, as the division algorithm is computable, but the Jordan canonical form (when it exists) maybe harder to compute as it depends on factoring the invariant factors of the monic polynomials arising in the rational canonical form.

If A in $\mathbb{M}_n(F)$ is an upper triangular matrix, then the characteristic polynomial f_A of A is a product of linear polynomials (why?), hence the minimal polynomial q_A also is a product of linear polynomials. The next two standard results in linear algebra results now follow easily.

Corollary 45.22. *Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Then A is similar to a diagonal matrix if an only if the minimal polynomial q_A factors as a product of monic linear factors without multiple roots, i.e., $q_A = (t - \lambda_1) \cdots (t - \lambda_r)$ for some distinct $\lambda_1, \dots, \lambda_r$ in F.*

We say that a non-constant polynomial in $F[t]$ *splits* over F if it is a product of linear polynomials in $F[t]$. The corollary says that the matrix is *diagonalizable* if and only if q_A splits over F without any *multiple roots*.

Corollary 45.23. *Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Then A is similar to a triangular matrix if an only if the minimal polynomial q_A splits over F.*

We say that A is *triangularizable* if q_A splits over F.

Remark 45.24. Let V be a finite dimensional vector space over a field F and $T \in \operatorname{End}_F(V)$. We say that the linear operator T is *diagonalizable* if there exists a basis for V consisting of eigenvectors of T. The linear operator T is called *triangularizable* if there exists an ordered basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V such that $[T]_{\mathcal{B}}$ is upper triangular. (Equivalently, if $\mathcal{B}' = \{v_n, \dots, v_1\}$, then $[T]_{\mathcal{B}'}$ is lower triangular.) So in this terminology Corollaries 45.22 and 45.23 say:

(i) T is diagonalizable if and only if q_T splits over $F[t]$ and has no multiple roots.
(ii) T is triangularizable if and only if q_T splits over $F[t]$'

Examples 45.25. Let F be a field

1. Similarity classes of matrices in $\mathbb{M}_3(F)$ are completely determined by the minimal and characteristic polynomials of a matrix:
   Let A lie in $\mathbb{M}_3(F)$ and $q_1 \mid \cdots \mid q_r$ be the invariant factors of A. So $f_A = q_1 \cdots q_r$ with $q_r = q_A$. As $A \in \mathbb{M}_3(F)$, we know that $r \leq 3$.
   $r = 1$: We must have $q_1 = q_A = f_A$ determines A up to similarity, so A is similar to $C_{q_1} = C_{q_A}$. We can say more if q_A is reducible. First suppose that $q_A = (t - \lambda)h$ in $F[t]$ with h irreducible. By Exercise 45.26(2) below, A is similar to the matrix
$$\begin{pmatrix} \lambda & 0 \\ 0 & C_h \end{pmatrix},$$
   which is unique up to block order (as that $C_{t-\lambda} = (\lambda)$). Suppose instead that q_A splits over F, say $q_A = (t - \lambda_1)(t - \lambda_2)(t - \lambda_3)$ and A is triangularizable. As $r = 1$, we must have A is similar to
$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \text{if } \lambda_1, \lambda_2, \lambda_3 \text{ are distinct,}$$
$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 1 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2 \neq \lambda_3,$$
$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 1 & \lambda_1 & 0 \\ 0 & 1 & \lambda_1 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2 = \lambda_3,$$
   matrices in Jordan Canonical Form, unique up to order of the blocks.
   $r = 2$: We have $q_1 \mid q_2 = q_A$ and $f_A = q_1 q_2$, so $q_1 = f_A/q_A$ in $F[t]$. Therefore, q_A and f_A determine A up to similarity, and A is similar to
$$\begin{pmatrix} C_{q_1} & 0 \\ 0 & C_{q_2} \end{pmatrix}.$$

Note that if this is the case, $\deg q_1 = 1$ and $\deg q_2 = 2$, so $q_1 \mid q_2$ implies that q_2 splits so A is triangularizable, i.e., a Jordan canonical form for A exists. If $q_1 = t - \lambda_1$ and $q_2 = (t - \lambda_1)(t - \lambda_2)$, then A is similar to

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \qquad \text{if } \lambda_1 \neq \lambda_2,$$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 1 & \lambda_1 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2,$$

unique up to the order of the blocks.

$r = 3$: We have $q_1 \mid q_2 \mid q_3 = q_A$, so $q_1 = q_2 = q_3 = q_A$. So in this case q_A and f_A determine A up to similarity, and A is diagonalizable with a single eigenvalue, i.e., A is similar to the matrix λI with λ an (the) eigenvalue of A. In particular, A is similar to a matrix in Jordan canonical form.

2. Similarity classes of matrices in $\mathbb{M}_4(F)$ are not determined by minimal and characteristic polynomials. For example the matrices

$$\begin{array}{ccc} A & & B \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} & \text{and} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

have

$$\begin{array}{ccc} q_A = (t-1)^2 & & q_B = (t-1)^2 \\ f_A = (t-1)^4 & & f_B = (t-1)^4 \\ \text{elementary divisors} & & \text{elementary divisors} \\ t-1, t-1, (t-1)^2 & & (t-1)^2, (t-1)^2 \\ \text{invariant factors} & & \text{invariant factors} \\ t-1 \mid t-1 \mid (t-1)^2 & & (t-1)^2 \mid (t-1)^2 \end{array},$$

so they are not similar.

3. A system of representatives for the similarity classes of matrices in $\mathbb{M}_4(F)$ having characteristic polynomial $t(t-1)^3$ are:

$\begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&1&1 \end{pmatrix}$	$\begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&1&1&0\\ 0&0&1&1 \end{pmatrix}$
minimal polynomial	minimal polynomial	minimal polynomial
$t(t-1)$	$t(t-1)^2$	$t(t-1)^3$
elementary divisors	elementary divisors	elementary divisors
$t, t-1, t-1, t-1$	$t, (t-1), (t-1)^2$	$t, (t-1)^3$
invariant factors	invariant factors	invariant factors
$t-1 \mid t-1 \mid t(t-1)^2$	$(t-1) \mid t(t-1)^2$	$t(t-1)^3$

4. A system of representatives for the similarity classes of matrices in $\mathbb{M}_4(F)$ with minimal polynomial $t(t-1)$ are:
$$\begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}$$
as these matrices must be diagonalizable with eigenvalues 0 and 1.
5. Let A be a 3×3 matrix in $\mathbb{M}_3(\mathbb{Q})$ satisfying $A^3 = I$. Then $q_A \mid t^3-1 = (t-1)(t^2+t+1)$ in $\mathbb{Q}[t]$, so the only possibilities are $t-1$, t^2+t+1, or t^3-1. If $q_A = t^2+t+1$, then $q_1 \mid q_2 = q_A$, but q_A is irreducible, so no linear polynomial divides it, so this case cannot occur. If $q_A = t-1$, A is diagonalizable and $t-1 \mid t-1 \mid t-1$ are the invariants. So the possible rational canonical forms for A are:
$$\begin{pmatrix} 1&0&0\\ 0&1&0\\ 0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&0&1\\ 1&0&0\\ 0&1&0 \end{pmatrix}$$
[Note: Using Exercise 45.26(2) below, an alternate matrix similar to
$$\begin{pmatrix} 0&0&1\\ 1&0&0\\ 0&1&0 \end{pmatrix} \quad \text{is} \quad \begin{pmatrix} 1&0&0\\ 0&0&-1\\ 0&1&-1 \end{pmatrix}]$$

Exercises 45.26.

1. Prove Observation 45.10.
2. Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Let $f_1, \dots, f_r$ be the monic elementary divisors of A, i.e., those of $F^{n\times 1}$ as a $F[t]$ module via $tv = Av$ for all v in $F^{n\times 1}$. Show that A is similar to the matrix
$$\begin{pmatrix} C_{f_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & C_{f_r} \end{pmatrix}$$
and this matrix is unique up to the order of the blocks.

3. Let V be a finite dimensional vector space over F and $T \in \operatorname{End}_F(V)$
4. Let A be a matrix in $\mathbb{M}_5(\mathbb{Q})$ satisfying $A^3 = I$. Determine all possible rational canonical forms of A and justify your answer.
5. Determine the Rational or Jordan canonical form of all 3×3 matrices A over a field F satisfying $A^4 = I$, over each prime field and justify your answer.
6. Determine all 4×4 matrices A over a field F satisfying $A^5 = I$ in the following three cases and justify your answer.
   (i) F is the field of rational numbers.
   (ii) F is $\mathbb{Z}/2\mathbb{Z}$.
   (iii) F is $\mathbb{Z}/5\mathbb{Z}$.
7. Compute the Jordan and Rational Canonical Forms of the matrix
$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$
8. Prove Corollaries 45.22 and 45.23.
9. Let V be a real vector space, T an $\mathbb{R}$-endomorphism. Suppose that the minimal polynomial of T factors into linear terms over $\mathbb{R}$ with no repeated and no negative roots. Show that there exists an $\mathbb{R}$-endomorphism S of V such that $S^2 = T$.
10. Recall that a matrix $C \in \mathbb{M}_n\mathbb{C}$ is called *diagonalizable* if there exists a matrix $Q \in \mathrm{GL}_n(\mathbb{C})$ such that QAQ^{-1} is a diagonal matrix. Show if $A, B \in \mathbb{M}_n\mathbb{C}$ are both diagonalizable, then there is a matrix $P \in \mathrm{GL}_n(\mathbb{C})$ satisfying both PAP^{-1} and PBP^{-1} are diagonal matrices if and only if $AB = BA$.

46. Addendum: Jordan Decomposition

We prove a weaker form of Jordan canonical form that only needs the Primary Decomposition Theorem, not the full Fundamental Theorem 44.6. [The Cayley-Hamilton Theorem can be proven without the full version of the Fundamental Theorem.]

Definition 46.1. Suppose that F is a field, V a vector space over F, and T is a linear operator on V. Let
$$E_T(\lambda) := \{v \in V \mid T(v) = \lambda v\},$$
a subspace of V. If λ is an eigenvalue of T, then $E_T(\lambda)$ is called the *eigenspace* of λ relative to T, i.e., $E_T(\lambda)$ is an eigenspace if and only if it is not the zero subspace. We say that T is *semisimple* if V has a basis consisting of eigenvectors of T, equivalently, $V = \bigoplus_F E_T(\lambda)$. Of course, $T|_{E_T(\lambda)} = \lambda 1_{E_T(\lambda)}$ for all $\lambda \in F$.

Observation 46.2. Let V be a vector space over F and S, T two commuting linear operators on V, i.e., $T \circ S = S \circ T$. If λ is an eigenvalue of T, then $E_T(\lambda)$ is S-invariant, i.e., $S|_{E_T(\lambda)} : E_T(\lambda) \to E_T(\lambda)$:
If $v \in E_T(\lambda)$, then $T\big(S(v)\big) = S\big(T(v)\big) = S(\lambda v) = \lambda S(v)$.

The observation has the following consequence:

Theorem 46.3. *Let V be a finite dimensional vector space over the field F and $\mathcal{S}$ a set of commuting semisimple linear operators in $\operatorname{End}_F(V)$, i.e., if T and S lie in $\mathcal{S}$, they are semisimple and $ST = TS$. Then there exists a basis $\mathcal{B}$ for V consisting of eigenvectors for every T in $\mathcal{S}$ simultaneously. In particular, $T + S$ is semisimple for all $T, S \in \mathcal{S}$.*

PROOF. If $T = \rho 1_V$ for some $\rho \in F$ for every $T \in \mathcal{S}$, the result is trivial. So we may assume that there exists a T in $\mathcal{S}$ so that this is not true. As T can have only finitely many eigenvalues, say $\lambda_1, \dots, \lambda_r$, we have $V = E_T(\lambda_1) \oplus \cdots \oplus E_T(\lambda_r)$ with each $E_T(\lambda_j) < V$. By the observation, $S|_{E_T(\lambda_j)} \in \operatorname{End}_F(E_T(\lambda_j))$ for all $S \in \mathcal{S}$. As $q_{S|_{E_T(\lambda_j)}} \mid q_S$ (why?), we know that $q_{S|_{E_T(\lambda_j)}}$ splits over F, so $S|_{E_T(\lambda_j)}$ is semisimple. By induction, the result holds on each $E_T(\lambda_j)$, hence on V. $\square$

Definition 46.4. Let F be a field, V a nontrivial vector space over F, and T a linear operator on V. We say that T is *nilpotent* if there exists a positive integer N such that $T^N = 0$ and *unipotent* if $T - 1_V$ is nilpotent.

Note that a unipotent linear operator is always an isomorphism (why?) and an operator that is both nilpotent and semisimple must be the zero linear operator.

Theorem 46.5. *Let V a finite dimensional vector space over F, and T a linear operator on V. Suppose that f_T splits over F. Then there exist unique linear operators T_s and T_n on V semisimple and nilpotent respectively and if T is a linear automorphism, there exists a unique unipotent operator T_u satisfying the following:*

(1) (Additive Jordan Decomposition) $T = T_s + T_n$ *and* $T_sT_n = T_nT_s$.
(2) (Multiplicative Jordan Decomposition) *If T is a linear automorphism, then $T = T_sT_u$ and $T_sT_u = T_uT_s$.*
(3) *There exist polynomials g and h in $F[t]$ with zero constant term, i.e., $g(0) = 0 = h(0) = 0$ satisfying $T_s = g(T)$ and $T_n = h(T)$. In particular, if a linear operator S on V commutes with T, then it commutes with both T_s and T_n.*
(4) *If T is a linear automorphism and T commutes with a linear operator S on V, then it commutes with T_s and T_u.*

PROOF. Let V be an $F[t]$-module by evaluation at T and suppose that the characteristic polynomial of T is $f_T = \prod_{i=1}^r (t - \lambda_i)^{e_i}$ in $F[t]$ with $\lambda_i \neq \lambda_j$ if $i \neq j$, and $e_i > 0$ for all i. Then by the Cayley-Hamilton Theorem, $f_T v = 0$ for all v in V. Let

$$V_i = \ker(T - \lambda_i 1_V)^{e_i} \cong F[t]^{e_i}/\big((t - \lambda_i)^{e_i}\big)$$

for $i = 1, \dots, r$, so $V = V_1 \oplus \cdots \oplus V_r$ as $F[t]$-modules by the Primary Decomposition Theorem 44.20 (Why?). The ideals $(t - \lambda_i)^{e_i}$ and $\sum_{j \neq i} (t - \lambda_j)^{e_j}$ in the euclidean domain $F[t]$ are comaximal for all i. Therefore, there exists a polynomial g in $F[t]$ satisfying $g \equiv \lambda_i \mod (t - \lambda_i)^{e_i}$, $1 \leq i \leq r$ and $g \equiv 0 \mod t$ by the Chinese Remainder Theorem 27.19. (Of course, if some $\lambda_i = 0$, the last congruence is redundant.) In particular, $g(0) = 0$. Let $T_s = g(T)$. Each V_i is T-invariant, as it is an $F[t]$-module by evaluation at T, so each V_i is also T_s-invariant. We also have $T_s|_{V_i} = \lambda_i 1_{V_i}$ for $i = 1, \dots, r$, as the λ_i's are eigenvalues of T. Therefore, T_s is semisimple.

Let $h = t - g$ in $F[t]$, so $h(0) = 0$. Set $T_n = h(T)$. We have

$$T = g(T) + h(T) = T_s + T_n \text{ and } T_sT_n = T_nT_s.$$

In addition,

$$0 = (T|_{V_i} - \lambda_i 1_{V_i})_i^e = (T|_{V_i} - T_s|_{V_i})^{e_i} = T_n^{e_i}|_{V_i}$$

for each i, so T_n is nilpotent. This proves (3) and establishes the existence in (1).

If T is a linear automorphism, then 0 is not an eigenvalue of T, so T_s is invertible. Let $T_u = 1_V + T_s^{-1}T_n$. Then T_u is unipotent as $T_s^{-1}T_n = T_nT_s^{-1}$ is nilpotent. Statement (4) and the existence in (2) now follow.

We next show the uniqueness in (1). Suppose that we have a semisimple operator T_s' and a nilpotent operator T_n' satisfying

$$T_s + T_n = T_s' + T_n' \text{ and } T_s'T_n' = T_n'T_s'$$

Then $T_s - T_s' = T_n' - T_n$. By (3), T_s' and T_n' commute with both T_s and T_n, as they commute with $T = T_s' + T_n'$. Therefore, $T_n' - T_n$ is nilpotent and by Theorem 46.3 we have $T_s - T_s'$ is semisimple. Therefore, $T_s - T_s' = T_n' - T_n = 0$, i.e., $T_s = T_s'$ and $T_n = T_n'$.

Finally, we show the uniqueness in (3). Suppose that

$$T_sT_u = T_s'T_u' \text{ and } T_s'T_u' = T_n'T_u'$$

with T_u unipotent. Then $S = T_u' - 1_V$ is nilpotent and commutes with T_s', hence $T_s'S$ is nilpotent and $T = T_s' + T_s'S$ is a additive Jordan decomposition, so $T_s' = T_s$ (and $T_n = T_s'S$). Hence $T_u = T_s^{-1}T = T_s'^{-1}T = T_u'$, and we are done. $\square$

Our definition of a semisimple operator T on V is equivalent to the minimal polynomial q_T splitting without any repeated factors over F. The theorem generalizes if we can find an alternate condition of a linear operator on V when q_T does not split over F that will agree with our definition of a semisimple operator over a field containing F in which q_T splits. Then we can mimic the proof on a factorization of q_T, if it is not a product of distinct linear terms. To do this, one can redefine an operator T to be a *semisimple* operator on a vector space V over F if it satisfies the following: Let V be an $F[t]$-module by evaluation at T. If W is an $F[t]$-submodule of V then $V = W \oplus W'$ for some $F[t]$-module W', i.e., and exact sequence

$$0 \to W \to V \to W' \to 0$$

of $F[t]$-modules is split exact (cf. Exercise 39.12(12)). With this definition, the theorem still holds if the characteristic of F is zero. [If the characteristic of F is positive, irreducible polynomials can have multiple roots over some field containing F.]

Exercises 46.6.

1. Show a unipotent operator on a vector space is an isomorphism and an operator that is both nilpotent and semisimple on a vector space is the trivial operator.
2. In the proof of Theorem 46.5, show that $V_i = \ker(T - \lambda_i 1_V)^{e_i} \cong F[t]/\big((t-\lambda_i)^{e_i}\big)$ is the $(t - \lambda_i)$-primary submodule of V as an $F[t]$-module.
3. In Theorem 46.5, show if W is an $F[t]$-submodule of V, i.e., is T-invariant, then the Jordan decomposition(s) of T induce those on $T|_W$ and $T_{V/W}$ (where $T_{V/W}(x + W) = T(x)$).

47. Addendum: Cayley-Hamilton Theorem

In this addendum, we generalize the Cayley-Hamilton Theorem. Let R be a commutative ring and M a finitely generated R-module. Let

$$M[t] := \{\sum t^i x_i \mid x_i \in M \text{ almost all } x_i = 0\} = \coprod_{i=0}^{\infty} t^i M.$$

This is clearly an R-module and becomes a finitely generated $R[t]$-module in the obvious way.

For example, if V is a finite dimensional vector space over F on basis $\{v_i\}$, then $V[t]$ is the F-vector space on basis $\{t^i v_j \mid i \geq 0, j\}$ and a free $F[t]$-module on basis $\{v_i\}$. So if $\dim_F V = n$, $V[t] \cong F[t]^n$.

If $f : M \to N$ is an R-homomorphism, define $f[t] : M[t] \to N[t]$ by $\sum t^i x_i \mapsto \sum t^i f(x_i)$, an $R[t]$-homomorphism.

Suppose that N is an $R[t]$-module. Then it is also an R-module and N is completely determined by

(*i*) the R-module structure.
(*ii*) The R-endomorphism $f : N \to N$ given by $f(x) = tx$.

In particular, an R-module becomes an $R[t]$-module via (i) and (ii). We denote this $R[t]$-module N_f. We then have an $R[t]$-epimorphism $\varphi_f : N[t] \to N_f$ given by $\sum t^i x_i \mapsto \sum f^i(x_i)$.

Theorem 47.1. *Let R be a commutative ring and M a finitely generated R-module. Suppose that f is an R-endomorphism of M. Then, in the notation above, we have an exact sequence*

$$0 \to M[t] \xrightarrow{t1_{M[t]} - f[t]} M[t] \xrightarrow{\varphi_f} M_f \to 0$$

of $R[t]$-modules.

The exact sequence in the theorem is called the *characteristic sequence* for f.

Proof. **Exactness at the middle** $M[t]$: As

$$\begin{aligned}\big(\varphi_f \circ (t1_{M[t]} - f[t])\big)(\sum t^i x_i) &= \varphi_f\big(\sum t^{i+1} x_i - t^i f(x_i)\big)\\ &= \sum f^{i+1}(x_i) - f^{i+1}(x_i) = 0,\end{aligned}$$

we have $\operatorname{im}(t1_{M[t]} - f[t]) \subset \ker \varphi_f$. If $x = \sum t^i x_i \in \ker f$, then $\sum f^i(x_i) = 0$, hence

$$x = x - \sum f^i(x_i) = \sum_{i \geq 0} \big(t^i x_i - f^i(x_i)\big) = \sum_{i>0} (t^i 1_{M[t]} - f^i)(x_i).$$

Check if $h_i = \sum_{j=0}^{i} t^j 1_{M[t]} f^{i-j}[t]$, then

$$x = \sum_{i>0} (t1_{M[t]} - f[t])\big(\sum_{i>0} h_i(x_i)\big) \text{ lies in } \operatorname{im}(f(t1_{M[t]} - f[t]).$$

φ_f **is surjective**: This follows as $\varphi_f(x) = x$ for all $x \in M$.

$t1_{M[t]} - f[t]$ is **monic**: We leave this part of the proof as an exercise noting that $t1_{M[t]} - f[t]$ raises the "degree" by one and preserves the "leading coefficients", so induction on degree works. $\square$

Cramer's Rule can be stated in the following way: Let R be a commutative ring, f an R-endomorphism of the free R-module R^n. Set $M = \operatorname{coker} f = R^n / \operatorname{im} f$. Then there exists an R-endomorphism g of R^n such that $fg = gf = (\det f)1_{R^n}$. In particular, $R^n \det f \subset \operatorname{im} f$. If M is a free R-module of rank n on basis $\mathcal{B}$ and $f \in \operatorname{End}_R(M)$, then $M[t]$ is $R[t]$-free on $\mathcal{B}$ and $[f]_{\mathcal{B}} = [f[t]]_{\mathcal{B}}$. If follows that $P(t) := \det(t1_{M[t]} - f[t])$ is that characteristic polynomial of f. Applying the characteristic sequence of f in the theorem implies that $P(f)M_f = 0$. As the R-endomorphism of M_f defined by $P_f(t)$ is just $P_f(f)$, we have $P_f(f) = 0$, i.e., the Cayley-Hamilton Theorem.

Exercises 47.2.

1. Prove that $t1_{M[t]} - f[t]$ in the characteristic sequence is monic.
2. Let R be a commutative ring and $\mathfrak{A}$ an ideal, and M an R-module generated by n elements. Suppose that φ is an R-endomorphism satisfying $\varphi(M) \subset \mathfrak{A}M$. Show that there exists a a monic polynomial $f = t^n + a_1 t^{n-1} + \cdots + a_n$ in $F[t]$ satisfying $a_i \in \mathfrak{A}^i$ for $i = 1, \dots, n$, and $f(\varphi) = 0$ as an endomorphism of M.
3. Let R be a commutative ring and M a finitely generated R-module. Show that any R-epimorphism of M is an R-automorphism.
4. Let R be a commutative ring and M a free R-module isomorphic to R^n. Prove that any generating set $\mathcal{B}$ of M consisting of n elements is a basis. In particular, show that this gives another proof that the rank of M is well-defined.

Part 5

Field Theory

CHAPTER XI

Field Extensions

In this chapter, we begin our study of fields. Given a non-constant polynomial, it is natural to ask what are its roots? To make this more precise, if R is a ring and f is a non-constant polynomial in a single variable over R, one asks does f have any roots in R. If so, how many, and if not, does there exist a ring S with R a subring such that f has a root in S? We have seen that if R is a domain, then the number of distinct roots f has in R is bounded by the degree of f. The first domain one encounters is the integers. The arithmetic problem, i.e., for f in $\mathbb{Z}[t]$ and monic leads to the study of algebraic number theory that we shall investigate in Chapter XV and is harder than that over fields. If $R = F$ is a field, we already know how to obtain a field K containing F such that f has a root in K. By induction, it is easy to construct a field E containing F for which the polynomial f splits. We shall show that the smallest such E is essentially unique. Generalizing this construction, we indicate, using Zorn's Lemma, how to do this for sets of polynomials in $F[t]$. In particular, we construct a field over which every polynomial in $F[t]$ splits (and is the smallest possible). This field turns out to be an algebraically closed field, i.e., one in which every non-constant polynomial over it splits. We also use the theory developed to answer the classical euclidean construction problems using only straight-edge and compass: the trisection of an angle, doubling of a cube, squaring of the circle, and construction of regular n-gons, showing how to turn a geometric problem into an algebraic one. The solution of the first two will be done in this chapter as well as the last (assuming a result proven in the next chapter) and the third will be done in §73. Finally, we investigate the difficulty arising from the case of positive characteristic of a field, where irreducible polynomials can have multiple roots over some bigger field.

48. Algebraic Elements

In this section, given a field F, we are mostly interested in roots of polynomials with coefficients in F. We have seen before that if f is a non-constant polynomial in $F[t]$, then there exists a field K containing F such that f has a root in K. In fact, we explicitly constructed such a field. We shall investigate this situation in depth.

Let K be a field containing F as a subfield, so $F \subset K$. We shall write this as K/F. We call F the *base field* of K/F and K the *extension field* of K/F. If E is a field with $F \subset E \subset K$ (as subfields), written $K/E/F$, we call E an *intermediate field* of K/F. We allow E to be K or F. A sequence of fields $F = F_1 \subset \cdots \subset F_n \subset \cdots$ is called a *tower of fields*. It may be infinite.

Definition 48.1. Let K/F be an extension of fields. Then K is a vector space over F by restriction of scalars. We denote the dimension, $\dim_F K$, of K as an F-vector space by

$[K : F]$ and call it the *degree* of K/F. We call K/F a *finite extension* if $[K : F]$ is finite, and an *infinite extension* otherwise.

Examples 48.2. 1. If F is any field, thenF/F is a finite extension of degree one.

2. $\mathbb{C}/\mathbb{R}$ is a finite extension of degree two.

3. $\mathbb{R}/\mathbb{Q}$ is an infinite extension.

Proposition 48.3. *Let $L/K/F$ be an extension of fields. Then L/F is a finite extension if and only if both L/K and K/F are finite field extensions. Moreover, if L/F is a finite extension, then*

$$[L : F] = [L : K][K : F].$$

PROOF. We prove a stronger statement. By Proposition 28.6, we know that all vector spaces have bases. Let

$$\begin{aligned}
\mathcal{B} &= \{x_i\}_I \text{ be an } F\text{-basis for } K,\\
\mathcal{C} &= \{y_j\}_J \text{ be a } K\text{-basis for } L, \text{ and}\\
\mathcal{D} &= \{x_i y_j\}_{I\times J}.
\end{aligned}$$

Claim. $\mathcal{D}$ is an F-basis for L and $|\mathcal{D}| = |I \times J|$.

Note $x_i y_j = x_{i'} y_{j'}$ if and only if $x_i = x_{i'}$ and $y_j = y_{j'}$ as $\mathcal{C}$ is a K-basis of L and $x_i, x_{i'}$ lie in K, hence $|\mathcal{D}| = |\mathcal{B}||\mathcal{C}| = |I \times J|$.

If we show this, then we are done.

$\mathcal{D}$ **F-spans** L: Let $z \in L$. Then $z = \sum_J \alpha_j y_j$ for some $\alpha_j \in K$ almost all zero, as $\mathcal{C}$ spans L as a vector space over K. As $\mathcal{B}$ spans K as a vector space over F, each $\alpha_j = \sum_I c_{ij} x_i$ for some $c_{ij} \in F$ almost all zero. Hence we have $z = \sum_I \sum_J c_{ij} x_i y_j$ lies in the F-span of $\mathcal{D}$.

$\mathcal{D}$ **is F-linearly independent**: Suppose that $\sum_{I\times J} c_{ij} x_i y_j = 0$ for some c_{ij} in F almost all zero. Then $\sum_J (\sum_I c_{ij} x_i) y_j = 0$. As $\mathcal{C}$ is K-linearly independent, $\sum_I c_{ij} x_i = 0$ for all j. As $\mathcal{B}$ is F-linearly independent, $c_{ij} = 0$ for all i and j. $\square$

This field theoretic analogue of Lagrange's Theorem has two analogous consequences.

Corollary 48.4. *Let $L/K/F$ be an extension of fields with L/F finite. Then $[K : F] \mid [L : F]$ and $[L : K] \mid [L : F]$. In particular, if $[L : F]$ is prime, then $K = L$ or $K = F$.*

If $[L : F]$ is prime in the above, we say that there exist no *nontrivial intermediate fields* in L/F.

Corollary 48.5. *If $F_1 \subset \cdots \subset F_n$ is a tower of fields with F_n/F_1 finite, then*

$$[F_n : F_1] = [F_n : F_{n-1}] \cdots [F_2 : F_1].$$

Given a field extension, we want to look at intermediate fields. We do this just as we do to get subgroups of groups, submodules of modules, etc.

Lemma 48.6. *Let K be a field and S a nonempty subset of K. Then there exists a unique minimal subfield F_0 of K containing S, i.e., if $S \subset F \subset K$ with F a field, then $F_0 \subset F$.*

PROOF. Let

$$\mathcal{F} = \{F \mid F \text{ a field satisfying } S \subset F \subset K\}.$$

As $K \in \mathcal{F}$, the set $\mathcal{F}$ is not empty. Set $F_0 = \bigcap_{\mathcal{F}} F$. Certainly, $S \subset F_0$. We show that F_0 works. We first prove that F_0 is a field. Indeed if $x, y \in F_0$ with $x \neq 0$, then x and y lie in any $F \in \mathcal{F}$, hence $x \pm y$, xy, x^{-1} also lie in any F in $\mathcal{F}$, establishing F_0 is a field. To finish, we show that F_0 is the unique minimal field containing S. But this is clear, for if $E \in \mathcal{F}$, then $F_0 \subset \bigcap_{\mathcal{F}} F \subset E$. $\square$

Notation 48.7. Let K/F be a field extension and X a nonempty subset of K. Let $S = F \cup X$. The field F_0 in Lemma 48.6 will be denoted by $F(X)$. If $X = \{\alpha_1, \ldots, \alpha_n\}$, we shall write $F(\alpha_1, \ldots, \alpha_n)$ for $F(X)$. In particular, if $\alpha \in K$, then

$$F(\alpha) = \text{the unique minimal intermediate field of } K/F \text{ containing } \alpha$$

and will be called a *simple* (or *primitive*) extension of F.

Remark 48.8. Let K/F be a field extension and X a nonempty subset of K. As usual $F[X]$ will denote the 'polynomials' in the $x \in X$ with coefficients in F, i.e., if we let t_x, with $x \in X$, be independent variables and $T = \{t_x \mid x \in X\}$, then $F[X]$ is the homomorphic image of the evaluation map $e : F[T] \to K$, a ring homomorphism as F is commutative, given by $t_x \mapsto x$ for all x in X (where $F[T] = F[t_x]_X$). As $F[X] \subset K$, the ring $F[X]$ is a domain. (So $\ker e$ is a prime ideal in $F[T]$.) By the universal property of quotient fields (Theorem 27.14), we must have $F \subset F[X] \subset qf(F[X]) \subset K$ and $qf(F[X]) \subset F(X)$. It follows by Lemma 48.6 that $qf(F[X]) = F(X)$. In particular, if $x = \{\alpha\}$, then $F \subset F[\alpha] \subset F(\alpha) = qf(F[\alpha]) \subset K$, with all of these domains and, in fact, except for possibly $F[\alpha]$, all are fields. Our first problem is to determine when $F[\alpha]$ is a field.

Remark 48.9.

Let $F \subset \mathbb{C}$ be a subfield and $d \in F$. Note that we must have $\mathbb{Q} \subset F$ is a subfield. If $\sqrt{d}$ lies in F, then $F = F[\sqrt{d}]$. So suppose not, i.e., suppose that d is not a square in F. Let $\alpha = a + b\sqrt{d}$ and $0 \neq \beta = c + e\sqrt{d}$ in $\mathbb{C}$ with $a, b, c, e \in F$. We must have $c^2 - e^2 d$ is not zero, since d is not a square in F. Therefore,

$$\frac{\alpha}{\beta} = \frac{a + b\sqrt{d}}{c + e\sqrt{d}} = \frac{a + b\sqrt{d}}{c + e\sqrt{d}} \cdot \frac{c - e\sqrt{d}}{c - e\sqrt{d}} = \frac{(a + b\sqrt{d})(c + e\sqrt{d})}{c^2 - e^2 d}$$

makes sense and lies in $F[\sqrt{d}]$. In particular, $F[\sqrt{d}] = F(\sqrt{d})$, hence a field. In this case, we also have $\dim_F F[\sqrt{d}] = 2$ with $\{1, \sqrt{d}\}$ an F-basis for $F(\sqrt{d})$. We now compare this to complex conjugation of $\mathbb{C}$ over $\mathbb{R}$. Consider the map

$$\sigma : F(\sqrt{d}) \to F(\sqrt{d}) \text{ given by } a + b\sqrt{d} \mapsto a - b\sqrt{d} \text{ for } a, b \in F.$$

It is a field automorphism satisfying $\sigma(a) = a$ for all $a \in F$. We call σ an F-*automorphism*. In fact, $\sigma(\alpha) = \alpha$ for $\alpha \in F(\sqrt{d})$ if and only if α lies in F, so the set $\{\alpha \in F(\sqrt{d}) \mid \sigma(\alpha) = \alpha\}$, called the *fixed field* of σ, is precisely F. This means that σ *moves* every element in $F(\alpha) \setminus F$. It is also noteworthy to observe that $\{1, \sigma\}$ (with 1 the identity automorphism $1_{F(\sqrt{d})}$) is a group of order two which is precisely the F-dimension of $F(\sqrt{d})$.

Examples 48.10. 1. $\mathbb{R}[\sqrt{-1}] = \mathbb{R}(\sqrt{-1}) = \mathbb{C}$, as all are 2-dimensional real vector spaces.

2. Let $d \in \mathbb{Z}$. Then
$$\mathbb{Q}[\sqrt{d}] = \mathbb{Q}(\sqrt{d}) = qf(\mathbb{Z}[\sqrt{d})])$$
is a vector space over $\mathbb{Q}$ of dimension at most two and exactly two if $d \notin \{x^2 \mid x \in \mathbb{Z}\}$, in which case $\{1, \sqrt{d}\}$ is a $\mathbb{Q}$-basis.

3. If F is a field, then we have
$$F < F[t] < F(t) := qf(F[t]),\ F[t]^\times = F^\times,\ \text{and}\ F(t)^\times = F(t) \setminus \{0\}.$$
The last follows as $F(t) = \{f/g \mid f, g \in F[t] \text{ with } g \neq 0\}$ is a field.

Definition 48.11. Let K/F be a field extension and α an element of K. We say

(1) α is *algebraic over* F if there exists a nonzero polynomial f in $F[t]$ satisfying $f(\alpha) = 0$, i.e., α is a root of a nonzero polynomial with coefficients in F (hence of a monic polynomial in $F[t]$).

(2) α is a *transcendental element over* F if it is not algebraic over F.

Our main interest in field theory is the study of algebraic elements over a fixed field. A good analog is to consider linear operators on a finite dimensional vector space. Historically the study of algebraic elements was the study of roots of polynomials in $\mathbb{Q}[t]$.

Examples 48.12. Let F be a field.

1. Every element α in F is algebraic over F as α is a root of the polynomial $t - \alpha$ in $F[t]$.
2. Let a, b be elements in $\mathbb{Q}$ and $\alpha = a + b\sqrt{-1}$ in $\mathbb{C}$. Then α is a root of $t^2 - 2at + (a^2 + b^2)$ in $\mathbb{Q}[t]$, so α is algebraic over $\mathbb{Q}$.
3. Suppose $F \subset \mathbb{C}$ and $d \in F$. Then $F[\sqrt{d}] = F(\sqrt{d})$. Let α be an element in $F[\sqrt{d}]$. We can write $\alpha = a + b\sqrt{d}$ for some $a, b \in F$. Then α is a root of the polynomial $t^2 - 2at + (a^2 - b^2 d)$ in $F[t]$, hence is algebraic over F.
4. Lindemann's Theorem says that the element π is transcendental over $\mathbb{Q}$. This is not easy [cf. Section 73]. It is easier to show that the real number e is transcendental over $\mathbb{Q}$ [cf. Section 71], as is the Liouville number $\sum 1/10^{n!}$ [cf. Section 70]. We have seen that $\mathbb{R}$ is uncountable. The set of roots of all nonzero polynomials over $\mathbb{Q}$ is countable, since there are countably many polynomials over $\mathbb{Q}$ each having finitely many roots. Therefore, "most" elements in $\mathbb{R}$ are transcendental. However, it is usually very difficult to show that a specific real number is transcendental over $\mathbb{Q}$. It is still unknown if π^e is transcendental over $\mathbb{Q}$.
5. The real numbers π and e are algebraic over the field $\mathbb{R}$. The notion of being transcendental is a relative notion, it depends on the field F unlike the indeterminant t which by definition is transcendental over any field F. Of course, if an element is algebraic over a field, it is algebraic over any extension field.
6. The complex number $e^{\sqrt{-1}\pi}$ is algebraic over $\mathbb{Q}$ as it is equal to -1, but it can be shown that the number e^π is transcendental over $\mathbb{Q}$ (cf. Theorem 1.4).
7. Let $L/K/F$ be field extensions. If α is an element in L algebraic over F then it is algebraic over K, since $F[t] \subset K[t]$. However, in general, if α is algebraic over K it need not be algebraic over F.

8. Let K/F be a field extension and u an element in K that is transcendental over F. We know that the evaluation map $e_u : F[t] \to F[u]$ given by $f \mapsto f(u)$ is a ring epimorphism and an F-linear transformation. Suppose that f lies in $\ker e_u$. As u is transcendental over F, we must have $f = 0$, hence e_u is a ring monomorphism. As it is also surjective, $e_u : F[t] \xrightarrow{\sim} F[u]$ is an isomorphism. The variable t is not a unit in $F[t]$, since $F[t]^\times = F^\times$. Therefore, we also have $F[t] < F(t)$, so $F[u] < qf(F[u]) = F(u)$. In addition, as $\{t^i \mid i \geq 0\}$ is an F-basis for $F[t]$, we have $\{u^i \mid i \geq 0\}$ is an F-basis for $F[u]$. In particular, $F[u]$ and $F(u)$ are both infinite dimensional vector spaces over F.
   Note: By (4), this means that $\mathbb{Q}[\pi] \cong \mathbb{Q}[t]$. But $\mathbb{R}[\pi] = \mathbb{R} < \mathbb{R}[t]$, so being transcendental is a relative notion depending on the base field F.

The main theorem about an algebraic element over a field is given by the following foundational result, much of which we have implicitly seen before.

Theorem 48.13. *Let K/F be an extension of fields and α an element in K that is algebraic over F. Then there exists a unique monic irreducible polynomial $m_F(\alpha)$ in $F[t]$ satisfying all of the following:*

(1) α is a root of $m_F(\alpha)$.
(2) If α is a root of a nonzero polynomial g in $F[t]$, then $m_F(\alpha) \mid g$ in $F[t]$ and $\deg g \geq \deg m_F(\alpha)$ with equality if and only if $g = am_F(\alpha)$ for some nonzero element $a \in F$.
(3) Let $n = \deg m_F(\alpha)$. Then $F(\alpha) = F[\alpha]$ and $\{1, \alpha, \dots, \alpha^{n-1}\}$ is an F-basis for $F[\alpha] = F(\alpha)$. In particular,

$$[F(\alpha) : F] = \dim_F F[\alpha] = \deg m_F(\alpha).$$

The polynomial $m_F(\alpha)$ in the theorem is called the *minimal* or *irreducible polynomial of α over F*. The integer $\deg m_F(\alpha) = [F(\alpha) : F]$ is called the *degree* of *α over F*.

Proof. We begin the proof by modifying the argument in Construction 45.1. Let $e_\alpha : F[t] \to F[\alpha]$ be given by $f \mapsto f(\alpha)$, evaluation at α, a ring epimorphism and F-linear transformation of vector spaces over F. As α is algebraic over F, α is the root of a nonzero polynomial in $F[t]$ which may be assumed to be monic. Therefore, there exists a positive integer N such that α^N lies in the F-vector space $\sum_{i=0}^{N-1} F\alpha^i$. It follows by induction that α^{N+j} lies in $\sum_{i=0}^{N-1} F\alpha^i$ for all $j \geq 0$, so $F[\alpha] = \sum_{i=0}^{N-1} F\alpha^i$ is a finite dimensional vector space over F. Since $F[t]$ is an infinite dimensional vector space over F, it follows that e_α cannot be a monomorphism, so $0 < \ker e_\alpha < F[t]$ ($e_\alpha(1) = 1$, so e_α is not the trivial map).

Since $F[t]$ is euclidean, it is a PID, hence there exists a unique non-constant monic polynomial f in $F[t]$ such that $\ker e_\alpha = (f)$. [It is unique, as generators of $\ker e_\alpha$ only differ by a nonzero element in F.] By the First Isomorphism Theorem, $F[\alpha] = \operatorname{im} e_\alpha \cong F[t]/(f)$. Let $n = \deg f$.

Claim. $f = m_F(\alpha)$ works.

As $F[t]/(f) \cong F[\alpha] \subset K$, it is a domain, so $0 < (f) < F[t]$ is a prime ideal. In particular, f must be an irreducible polynomial. Since $F[t]$ is a PID, it must be a maximal ideal, hence $F[t]/(f) \cong F[\alpha]$ is a field, so $F[\alpha] = F(\alpha)$.

We next show that $\mathcal{B} = \{1, \alpha, \dots, \alpha^{n-1}\}$ is an F-basis for $F[\alpha]$. Let $g \in F[t]$. As $F[t]$ is euclidean (under the degree map), $g = fq + r$ for some $q, r \in F[t]$ with $r = 0$ or

$\deg r < \deg f$. Since e_α is a ring homomorphism, $g(\alpha) = f(\alpha)q(\alpha) + r(\alpha) = r(\alpha)$, which lies in $\sum_{i=0}^{n-1} F\alpha^i$, the span of $\mathcal{B}$. Moreover, $g(\alpha) = 0$ if and only if $r(\alpha) = 0$. As $r = 0$ or $\deg r < \deg f$, we have $r(\alpha) = 0$ if and only if $r \in \ker e_\alpha$ if and only if $f \mid r$ in $F[t]$ if and only if $r = 0$. We conclude that if $g(\alpha) = \sum_{i=0}^{n-1} a_i\alpha^i = 0$ with $a_i \in F$, then each $a_i = 0$. Consequently, $\mathcal{B}$ is also F-linearly independent, hence an F-basis for $F[\alpha]$. The other needed properties of f now follow easily. □

Remarks 48.14. Let $K/E/F$ be field extensions and α in K an algebraic element over F.

1. $m_E(\alpha) \mid m_F(\alpha)$ in $E[t]$ and $\deg m_E(\alpha) \le \deg m_F(\alpha)$. [Of course, they can be different.]
2. If $f \in F[t]$ is monic and irreducible with $f(\alpha) = 0$, then $f = m_F(\alpha)$.
3. If $g \in F[t]$ satisfies $\deg g = [F(\alpha) : F]$ and $g(\alpha) = 0$, then g is irreducible.

Proposition 48.15. (Characterization of algebraic elements) *Let K/F be a field extension and $\alpha \in K$. Then α is algebraic over F if and only if $[F(\alpha) : F]$ is finite.*

Proof. ($\Rightarrow$): This follows from Theorem 48.13.

($\Leftarrow$): Suppose that $n = [F(\alpha) : F]$. Then $\{1, \alpha, \dots, \alpha^n\}$ is F-linearly dependent. Therefore, there exist $a_i \in F$, $1 \le i \le n$, not all zero such that $\sum_{i=0}^{n} a_i\alpha^i = 0$, so α is a root of the nonzero polynomial $\sum_{i=0}^{n} a_i t^i$ in $F[t]$. □

Corollary 48.16. *Let K/F be a finite field extension. Then every element in K is algebraic over F.*

Proposition 48.17. (Characterization of transcendental elements) *Let K/F be a field extension and $u \in K$. Then the following are equivalent:*

(1) u is transcendental over F.
(2) $F[u] \cong F[t]$.
(3) $F[u] < F(u)$.
(4) $[F(u) : F]$ is infinite.
(5) $\dim_F F(u)$ is infinite.

Proof. By Example 48.12(8), we know that (1) implies (2), (3), (4), and (5), and if (2), (3), (4), or (5) holds, then u cannot be algebraic and (1) must also hold. □

Corollary 48.18. *Let K/F be a field extension. If $a_1, \dots, a_r$ are elements of K algebraic over F, then $F[a_1, \dots, a_r] = F(a_1, \dots, a_r)$ and $[F(a_1, \dots, a_r) : F] \le \prod_{i=1}^{r}[F(a_i) : F]$ (with strict inequality possible).*

Proof. As a_r is algebraic over F, it is algebraic over $F(a_1, \dots, a_{r-1})$. By induction and the $r = 1$ case (that we have done), $F[a_1, \dots, a_r] = F[a_1, \dots, a_{r-1}][a_r] =$

$F(a_1, \ldots, a_{r-1})[a_r] = F(a_1, \ldots, a_r)$ and

$$\begin{aligned}
[F(a_1, \ldots, a_r) : F] &\leq [F(a_1, \ldots, a_r) : F(a_1, \ldots, a_{r-1})] \prod_{i=1}^{r-1} [F(a_i) : F] \\
&= \deg m_{F(a_1, \ldots, a_{r-1})}(a_r) \prod_{i=1}^{r-1} [F(a_i) : F] \\
&\leq \prod_{i=1}^{r} [F(a_i) : F].
\end{aligned}$$

□

Corollary 48.19. *Let A be a domain containing a field F. If A is a finite dimensional vector space over F, then A is a field.*

Proof. If A has an F-basis $\mathcal{B}$, then $A = F[\mathcal{B}]$ and its quotient field $K = F(\mathcal{B})$ is a finite dimensional vector space over F on basis $\mathcal{B}$, hence each element of $\mathcal{B}$ is algebraic over F. It follows that $A = K$ by Corollary 48.18. □

If K/F is a field extension, we arrive at the important conclusion that the set of elements in K algebraic over F form an intermediate field.

Theorem 48.20. *Let K/F be a field extension with a, b elements in K algebraic over F. Then $a \pm b$, ab, and b^{-1}, if $b \neq 0$, are algebraic over F. In particular, the set of elements in K algebraic over F forms an intermediate field of K/F.*

Proof. Let $\gamma = a \pm b$, ab, or b^{-1}, if $b \neq 0$. Then $F \subset F(\gamma) \subset F(a, b)$, so

$$[F(\gamma) : F] \leq [F(a, b) : F] \leq [F(a) : F][F(b) : F] < \infty.$$

The result follows. □

Definition 48.21. Let K/F be a field extension. We say that it is an *algebraic extension* (or simply *algebraic*) if every element of K is algebraic over F. If K/F is not algebraic, we say that it is a *transcendental extension*.

We have shown that every finite field extension is algebraic. The converse is false, as we shall see below. First, however, we show that being an algebraic extension is transitive. In fact, we have

Theorem 48.22. *Let $L/K/F$ be field extensions. Then L/F is algebraic if and only if both L/K and K/F are algebraic.*

Proof. ($\Rightarrow$): Any element in L algebraic over F is clearly algebraic over K and any element in K lies in L so is algebraic over F.

($\Leftarrow$): Let α be an element of L. We must show that α is algebraic over F. As it is algebraic over K, the minimal polynomial $m_K(\alpha) = \sum_{i=0}^{n} c_i t^i$ in $K[t]$ exists. Let $E = F(c_0, \ldots, c_{n-1}) \subset K$. Since the elements $c_0, \ldots, c_{n-1}$ are algebraic over F, we have $[E : F] < \infty$. As $m_K(\alpha) \in E[t]$, we also have $[E(\alpha) : E] < \infty$, so

$$[F(\alpha) : F] \leq [E(\alpha) : F] = [E(\alpha) : E][E : F] < \infty$$

and α is algebraic over F. $\square$

The above trick is very useful. If you want to prove a result and can reduce to a finite amount of data, you can often reduce to checking in a nicer set, e.g., in the above we reduced from an algebraic extension to a finite extension because we needed to attach only finitely many elements algebraic over the base field. In commutative ring theory, a finite amount of data over a commutative ring often reduces a proof to a Noetherian ring containing it by the Hilbert Basis Theorem.

Examples 48.23. 1. $\mathbb{C}/\mathbb{R}$ and $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$, with d an integer, are algebraic extensions.

2. $\mathbb{R}/\mathbb{Q}$ is transcendental as $\mathbb{R}$ is uncountable (by Cantor's Theorem) and $\mathbb{Q}$ is countable.
3. Let
$$\Omega := \{\alpha \in \mathbb{C} \mid \alpha \text{ algebraic over } \mathbb{Q}\}.$$
Ω is a countable field as it is the set of roots of countably many nonzero polynomials, each having finitely many roots.

   Claim. $\Omega/\mathbb{Q}$ is not finite:

   Let p be a (positive) prime in $\mathbb{Z}$ and n a positive integer. By Eisenstein's Criterion (35.11(1)), $t^n - p$ in $\mathbb{Q}[t]$ is irreducible and has a root α in Ω. In particular, $[\Omega : \mathbb{Q}] \geq [\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg m_{\mathbb{Q}}(\alpha) = \deg(t^n - p) = n$, establishing the claim.
4. Ω in (3) is *algebraically closed*, i.e., every non-constant polynomial in $\Omega[t]$ has a root in Ω. To see this, we use the Fundamental Theorem of Algebra (to be proved in 57.12 below) which says that $\mathbb{C}$ is algebraically closed. If f is a non-constant polynomial in $\Omega[t]$, it has a root α in $\mathbb{C}$, so $\Omega(\alpha)/\Omega$ is algebraic. As $\Omega/\mathbb{Q}$ is algebraic, it follows that so is $\Omega(\alpha)/\mathbb{Q}$, hence α is algebraic over $\mathbb{Q}$ hence is in Ω. As $\mathbb{C}$ is uncountable, we must have $\Omega < \mathbb{C}$.
5. The same argument as in (4) shows if K/F with K algebraically closed and Ω the set of elements of K that are algebraic over F, then Ω is an algebraically closed field. Moreover, by definition, if $F < E < \Omega$ is a field, then E is not algebraically closed. We call Ω an *algebraic closure* of F. We shall see in §51 that all such algebraic closures of a field F are unique up to a field isomorphism fixing F.

Calculations 48.24. 1. Let $p \in \mathbb{Z}^+$ be a prime. Then the polynomial $f = t^{p-1} + t^{p-2} + \cdots + t + 1$ is irreducible in $\mathbb{Q}[t]$ by the application of Eisenstein's Criterion in 35.11(3). Let ζ_p be a root of f in $\mathbb{C}$. Then $f = m_{\mathbb{Q}}(\zeta_p)$, so $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = \deg f = p - 1 = \varphi(p)$ (where φ is the Euler ϕ-function) and $\{1, \zeta_p, \ldots, \zeta_p^{p-2}\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\zeta_p)$. The element ζ is called a *primitive pth root of unity*.

   [Note that $t^p - 1 = (t-1)(t^{p-1} + t^{p-2} + \cdots t + 1)$ is a factorization into irreducibles in $\mathbb{Q}[t]$. There exists an irreducible polynomial of degree n dividing $t^n - 1$ in $\mathbb{Q}[t]$ when n is not a prime, but it is not so easy.]
2. Let α in $\mathbb{C}$ be a root of the irreducible polynomial $f = t^3 - 2t + 2$ in $\mathbb{Q}[t]$ (irreducible by Eisenstein's Criterion). So $[\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg f = \deg m_{\mathbb{Q}}(\alpha) = 3$ and $\mathcal{B} = \{1, \alpha, \alpha^2\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\alpha)$. As $\alpha^3 = 2\alpha - 2$ in $\mathbb{Q}(\alpha)$, we have $-\alpha^{-1} = (\alpha^2/2) - 1$ in $\mathbb{Q}(\alpha)$. Let $\beta \in \mathbb{Q}(\alpha)$, so $\mathbb{Q}(\beta) \subset \mathbb{Q}(\alpha)$, hence $[\mathbb{Q}(\beta) : \mathbb{Q}] \mid [\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$. Therefore, either $\beta \in \mathbb{Q}$ or $\mathbb{Q}(\beta) = \mathbb{Q}(\alpha)$. For example, suppose that $\beta = \alpha^2 - \alpha$. As $\mathcal{B}$ is $\mathbb{Q}$-linearly independent,

$\beta \notin \mathbb{Q}$, so $\mathbb{Q}(\beta) = \mathbb{Q}(\alpha)$ and $\mathcal{C} = \{1, \beta, \beta^2\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\beta)$ and $\{1, \beta, \beta^2, \beta^3\}$ is $\mathbb{Q}$-linearly dependent. Computation shows that

$$1 = 1, \quad \beta = \alpha^2 - \alpha, \quad \beta^2 = 3\alpha^2 - 6\alpha + 4$$

and

$$(*) \qquad \beta^3 = 16\alpha^2 - 28\alpha + 18.$$

It follows that the change of basis matrix for the basis $\mathcal{C}$ to the basis $\mathcal{B}$ is

$$[Id]_{\mathcal{C},\mathcal{B}} = \begin{pmatrix} 1 & 0 & 4 \\ 0 & -1 & -6 \\ 0 & 1 & 3 \end{pmatrix}.$$

Inverting this matrix to $[Id]_{\mathcal{B},\mathcal{C}}$ then shows that

$$1 = 1, \quad \alpha = -\frac{\beta^2}{3} + \beta + \frac{4}{3}, \quad \alpha^2 = -\frac{\beta^2}{3} + 2\beta + 4 + \frac{4}{3},$$

and substituting this into (*) yields $\beta^3 - 4\beta^2 - 4\beta - 2 = 0$, i.e., β is a root of monic $g = t^3 - 4t^2 - 4t - 2$ in $\mathbb{Q}[t]$. This must be the irreducible polynomial $m_{\mathbb{Q}}(\beta)$ as g has the same degree and satisfies $g(\beta) = 0$. [Of course, we also know that it is irreducible by Eisenstein's Criterion.]

3. It is easy to check that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$ with $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ (by showing that $\sqrt{3}$ cannot lie in $\mathbb{Q}(\sqrt{2})$). [Cf. Exercise 50.18(12) for the generalization.] Using this we determine $[\mathbb{Q}(\sqrt{2} + \sqrt{3}) : \mathbb{Q}]$ and $m_{\mathbb{Q}}(\sqrt{2} + \sqrt{3})$. Let $K = \mathbb{Q}(\sqrt{2} + \sqrt{3}) \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$. So $[\mathbb{Q}(\sqrt{2} + \sqrt{3}) : \mathbb{Q}] = 1, 2$ or 4. Let $\alpha = \sqrt{2} + \sqrt{3}$. Then
$$1 = 1, \quad \alpha = \sqrt{2} + \sqrt{3}, \quad \alpha^2 = 5 + 2\sqrt{6}.$$
These must be $\mathbb{Q}$-linearly independent (why?). It follows that the set $\{1, \alpha, \alpha^2, \alpha^3\}$ must be a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. In particular, $\mathbb{Q}(\alpha) = K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Now $\alpha^2 = 5 + 2\sqrt{6}$, so the polynomial $g = t^2 - (5 + 2\sqrt{6})$ lying in $\mathbb{Q}(\sqrt{6})[t]$ has a root α in K with $\alpha \notin \mathbb{Q}(\sqrt{6})$. Consequently, $[K : \mathbb{Q}(\sqrt{6})] = 2$ and $g = m_{\mathbb{Q}(\sqrt{6})}(\alpha)$. Computation shows that
$$h = (t^2 - \big(5 + 2\sqrt{6})\big)\big(t^2 - (5 - 2\sqrt{6})\big) = t^4 - 10t^2 + 1,$$
so it lies in $\mathbb{Q}[t]$ and has α as a root in $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ with $\deg h = \deg m_{\mathbb{Q}}(\alpha) = 4 = [K : \mathbb{Q}]$. Thus $h = m_{\mathbb{Q}}(\alpha)$.

We shall show later that if F is a field of characteristic zero and K/F a finite extension, that there always exists a θ in K such that $K = F(\theta)$. [Cf. Theorem 57.9.] In general, such a θ is hard to find, although many different ones work. This result is also not true for arbitrary fields of positive characteristic $p > 0$.

Exercises 48.25.

1. Let D is a division ring containing a field F and $\alpha \in D$ a root of a nonzero polynomial in $F[t]$. Show that $F[\alpha]$ is a field. [Cf. A finite division ring is a field.]
2. (a) Find $u \in \mathbb{R}$ such that $\mathbb{Q}(u) = \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$.

   (b) Describe how you how you would find all $w \in \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$ such that $\mathbb{Q}(w) = \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$.
3. Let $a, b \in K$ be algebraic over F of degrees m, n respectively with m, n relatively prime. Show that $[F(a, b) : F] = mn$.
4. Show that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$ with $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.
5. If $|F| = q < \infty$ show:
   (a) There exists a prime p such that char $F = p$.
   (b) $q = p^n$ some n.
   (c) $a^q = a$ for all $a \in F$.
   (d) If $b \in K$ is algebraic over F then $b^{q^m} = b$ for some $m > 0$.
6. Let u be a root of $f = t^3 - t^2 + t + 2$ in $\mathbb{Q}[t]$ and $K = \mathbb{Q}(u)$.
   (a) Show that $f = m_{\mathbb{Q}}(u)$.
   (b) Express $(u^2 + u + 1)(u^2 - u)$, and $(u - 1)^{-1}$ in the form $au^2 + bu + c$, for some $a, b, c \in \mathbb{Q}$.
7. Let $\zeta = \cos \frac{\pi}{6} + \sqrt{-1} \sin \frac{\pi}{6}$ in $\mathbb{C}$. Show that $\zeta^{12} = 1$ but $\zeta^r \neq 1$ for $1 \leq r < 12$. Show also that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = 4$ and find $m_{\mathbb{Q}}(\zeta)$.
8. Let $K = F(u)$, u algebraic over F and degree of u odd. Show that $K = F(u^2)$.
9. Let u be transcendental over F and $F < E \subset F(u)$. Show that u is algebraic over E.
10. Let $f, g \in F[t]$ be relatively prime and suppose that $u = f/g$ lies in $F(t) \setminus F$. Show that $F(t)/F(u)$ is finite of degree $d = \max\{\deg(f), \deg(g)\}$.
11. If $f = t^n - a \in F[t]$ is irreducible, $u \in K$ is a root of f, and $n/m \in \mathbb{Z}$, show that $[F(u^m) : F] = \dfrac{n}{m}$. What is $m_F(u^m)$?
12. If a^n is algebraic over a field F for some $n > 0$, show that a is algebraic over F.

49. Addendum: Transcendental Extensions

What can we say about a field extension of a field that is not algebraic, i.e., is transcendental? In this short section, we answer this question, leaving most of the details to the reader.

Definition 49.1. Let K/F be an extension of fields and suppose that $\mathcal{B} = \{x_i \mid i \in I\}$ is a subset of K. We say that the set $\mathcal{B}$ is *algebraically independent* over F if $x_i \neq x_j$ for $i \neq j$ and if $f \in F[\mathcal{B}]$ satisfies $f(x_i)_I = 0$, then $f = 0$. This is equivalent to

$$\{x_{i_1}^{n_{i_1}} \cdots x_{i_r}^{n_{i_r}} \mid n_{i_j} \geq 0 \text{ for all } i_j \text{ in } I, \text{ some } r \geq 0\}$$

is an F-linearly independent set for the F-vector space $F[\mathcal{B}]$.

For example, a set of indeterminants $T = \{t_i \mid i \in I\}$ is an algebraically independent set in the quotient field $F(T)$ of $F[T]$. A maximal algebraically independent subset $\mathcal{B}$ in an extension field K of F is called a *transcendence basis* for K over F.

Remarks 49.2. Let K/F be an extension of fields.

1. There exists a transcendence basis for K over F. If K/F is algebraic, it is the empty set. A proof of this is analogous to the proof of the existence of bases for vector spaces.

2. If $\mathcal{B}$ is a transcendence basis for K over F then $K/F(\mathcal{B})$ is algebraic.
3. The set $\{t_1, \ldots, t_n\}$ is a transcendence basis for $F(t_1, \ldots, t_n)$ over F.
4. The set $\mathcal{B} = \{x_i\}_I \subset K$ is a transcendence basis for K over F if and only if $\mathcal{B}$ is a maximal subset of K satisfying $F(\mathcal{B}) \cong F(\{t_i\}_I)$.
5. The extension K/F is called *finitely generated* if there exists a finite set $S \subset K$ such that $K = F(S)$. If K/F is finitely generated, then K has a finite transcendence basis over F.

The analogue of dimension of vector spaces holds for the cardinality of transcendence bases. The proof for the case in which K/F finitely generated is analogous to the so-called Replacement Theorem used to prove that dimension is well-defined for finite dimensional vector spaces. We shall only prove this case in some detail.

Proposition 49.3. *Let K/F be an extension of fields. Then any two transcendence bases of K over F have the same cardinality.*

PROOF. Although this follows from the fact that the cardinality of any basis for a fixed vector space is well-defined, we present a proof in the case that K/F is a field extension and K has a finite transcendence basis $\{x_1, \ldots, x_n\}$ over F. Suppose that $\{y_1, \ldots, y_m\}$ is a subset of K that is algebraically independent over F. We must show $m \leq n$. As $\{x_1, \ldots, x_n\}$ is a maximal algebraically independent set over F, the set $\{y_1, x_1, \ldots, x_n\}$ must be *algebraically dependent*, i.e., is not algebraically independent. Therefore, there exists a nonzero polynomial $f \in F[t_0, t_1, \ldots, t_n]$ satisfying $f(y_1, x_1, \ldots, x_n) = 0$. By assumption, y_1 occurs nontrivially in f as well as some x_i, say x_1. Then x_1 is algebraic over $F(y_1, x_2, \ldots, x_n)$. Hence both $F(x_1, \ldots, x_n)$ and $F(y_1, x_1, x_2, \ldots, x_n)$ are algebraic over $F(y_1, x_2, \ldots, x_n)$. It follows that $K/F(y_1, x_2, \ldots x_n)$ is algebraic. A subset $\mathcal{B}$ of $\{y_1, x_2, \ldots, x_n\}$ must be a transcendence basis of K over F and it must include y_1, for if $\mathcal{B} < \{y_1, x_2, \ldots, x_n\}$, reversing the argument would show that a proper subset of $\{x_1, \ldots, x_n\}$ would be a transcendence basis for K over F. It follows that $\{y_1, x_2, \ldots, x_n\}$ is a transcendence basis for K over F. The argument continues, as in the proof of the Replacement Theorem for vector spaces, by throwing in y_2 and seeing that we have to throw away some x_i in $\{x_2, \ldots, x_n\}$, etc. It follows that $m \leq n$. □

Let K/F be a field extension. Then the cardinality of a transcendence basis $\mathcal{B}$ for K over F is called the *transcendence degree* of K/F and is denoted by $\operatorname{tr}\deg_F K$. It follows that K/F is finitely generated if and only if $\operatorname{tr}\deg_F K$ is finite and $K/F(\mathcal{B})$ is a finite extension.

Proposition 49.4. *Let $L/K/F$ be field extensions. Then* $\operatorname{tr}\deg_F L$ *is finite if and only if both* $\operatorname{tr}\deg_F K$ *and* $\operatorname{tr}\deg_K L$ *are finite. Moreover, if* $\operatorname{tr}\deg_F L$ *is finite, then*

$$\operatorname{tr}\deg_F L = \operatorname{tr}\deg_K L + \operatorname{tr}\deg_F K.$$

Exercises 49.5.

1. Prove that if K/F is an extension of fields, then there exists a transcendence basis for K over F.

2. If K/F is an extension of fields, prove that $\mathcal{B} = \{x_i\}_I \subset K$ is a transcendence basis for K/F if and only if $\mathcal{B}$ is a maximal subset of K satisfying $F(\mathcal{B}) \cong F(\{t_i\}_I)$ and $F[\mathcal{B}] \cong F[\{t_i\}_I]$.
3. Prove Proposition 49.4.

50. Splitting Fields

We have shown in Kronecker's Theorem 34.12 that an irreducible polynomial over a field F has a root in an extension field. Indeed we construct such an extension as follows: Let F be a field and f an irreducible polynomial in $F[t]$. As $F[t]$ is a PID, the nontrivial prime ideal (f) in $F[t]$ is maximal, so $K = F[t]/(f)$ is a field. Let $^{-} : F[t] \to K$ be the canonical epimorphism given by $g \mapsto g + (f)$. The composition of maps $F \subset F[t] \to K$ is monic, since F is simple. Consequently, we can, and do, view this composition as an inclusion of F into K, i.e., we identify a in F and $\overline{a}$ in K. Therefore, we view K/F is a field extension. Under this identification,

$$g := \sum_i a_i t^i \mapsto \overline{g} = \sum \overline{a_i}\overline{t}^i = \sum a_i \overline{t}^i = g(\overline{t})$$

is evaluation at $\overline{t}$. In particular, $\overline{0} = \overline{f} = f(\overline{t})$ in K, so $f \in F[t] \subset K[t]$ has a root in K. If h lies in $F[t]$ then, as $F[t]$ is euclidean, $h = fq + r$ in $F[t]$ for some $q, r \in F[t]$ with $r = 0$ or $\deg r < \deg f$. Therefore, $\overline{h} = \overline{f}\overline{q} + \overline{r} = \overline{r}$. It follows that if $n = \deg f$, then $\mathcal{B} = \{1, \overline{t}, \dots, \overline{t}^{n-1}\}$ spans the vector space K over F. As $\overline{h} = 0$ in K if and only $f \mid h$ in $K[t]$, we have $h = 0$ or $\deg f \leq \deg h$. It follows that $\mathcal{B}$ is an F-basis for K (why?), $f = (\text{lead } f) m_F(\overline{t})$, and $n = \deg f = \deg m_F(\overline{t}) = [K : F]$.

So we have proved

Proposition 50.1. *Let F be a field and f an irreducible polynomial in $F[t]$. Then there exists an extension K/F of degree* $\deg f$ *such that f has a root in K.*

Generalizing (as we did before), we have:

Theorem 50.2. (Kronecker's Theorem) *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists an algebraic extension K/F such that f has a root in K and $[K : F] \leq \deg f$.*

Proof. Let $f_1 \mid f$ in the UFD $F[t]$ with f_1 irreducible. Then f has a root in the field $K = F[t]/(f_1)$, as f_1 does, with $[K : F] = \deg f_1 \leq \deg f$. □

Definition 50.3. Let K/F be an extension of fields and f a non-constant polynomial in $F[t]$. We say that f *splits over* K if f factors into a product of linear polynomials in $K[t]$ and that K is a *splitting field* of f *over* F or K/F is a *splitting field* of f if K is a minimal field extension of F such that f splits over K, i.e.,

(1) f splits over K.
(2) Either $K = F$ or whenever $K/E/F$ are field extensions with $E < K$, the polynomial f does not split over E.

The existence of splitting fields of a non-constant polynomial now will follow by an easy induction using Kronecker's Theorem.

Theorem 50.4. *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists a splitting field K of f over F satisfying* $[K : F] \leq (\deg f)!$.

Proof. Let $n = \deg f$. By Kronecker's Theorem, there exists a field extension K_1/F such that $f = (t - \alpha_1)f_1$ in $K_1[t]$ with $f_1 \in K_1[t]$ and $[K_1 : F] \leq n$. By induction, there exists a splitting field K_2 of f_1 over K_1 satisfying $[K_2 : K_1] \leq (\deg f_1)! = (n-1)!$. Thus $[K_2 : F] = [K_2 : K_1][K_1 : F] \leq n!$ and f splits over K_2. By well-ordering, there exists an intermediate field $K_2/K/F$ with K a splitting field of f over F and $[K : F] \leq n!$. [Is $K = K_2$?] □

Remarks 50.5. Let F be a field and f a non-constant polynomial in $F[t]$ of degree n, K a splitting field of f over F, and $\alpha_1, \dots, \alpha_n$ the roots of f in K. [They may not be distinct.]

1. We have $K = F(\alpha_1, \dots, \alpha_n)$ and the proof of the theorem shows that $[K : F] \leq (\deg f)! = n! = |S_n|$. One often calls the splitting field K of f over F the *root field* of *f over F* for this reason.
2. Let L/F be a field extension with both K and L lying in some larger field. Then $L(\alpha_1, \dots, \alpha_n)$ is a splitting field of f over L. Of course, it is possible that $L = L(\alpha_1, \dots, \alpha_n)$.
3. If $K/L/F$ are field extensions, then K is a splitting field of f over L.
4. If $g \mid f$ in $F[t]$ with g a non-constant polynomial, then g splits over K, hence there exists an intermediate field $K/E/F$ such that E is a splitting field of g over F.

We turn to uniqueness statements for splitting fields. We shall do this in greater generality, so we need some notation.

Definition 50.6. Let $R_1 \subset S_1$ and $R_2 \subset S_2$ be commutative rings and $\varphi : R_1 \to R_2$ and $\tau : S_1 \to S_2$ be ring homomorphisms. We say that τ *lifts* or *extends* φ if $\tau|_{R_1} = \varphi$, i.e., we have a commutative diagram:

$$\begin{array}{ccc} S_1 & \xrightarrow{\tau} & S_2 \\ {\scriptstyle inc}\uparrow & & \uparrow{\scriptstyle inc} \\ R_1 & \xrightarrow{\varphi} & R_2 \end{array}$$

where *inc* is the inclusion map. If $R = R_1 = R_2$, $\varphi = 1_R$, and τ lifts φ, we call τ an *R-algebra homomorphism*. If, in addition, τ is an isomorphism (respectively, monomorphism, epimorphism), then we call τ an *R-algebra isomorphism* (respectively, *R-algebra monomorphism, R-algebra epimorphism*). (Cf. Remark 26.11.) If $S = S_1 = S_2$ and τ is an automorphism, we call it an *R-algebra automorphism*. As is standard, if $R = F$ is a field, we call such an F-algebra homomorphism (respectively, F-algebra monomorphism, F-algebra epimorphism, F-algebra isomorphism, F-algebra automorphism (when appropriate) just an *F-homomorphism* (respectively, *F-monomorphism, F-epimorphism, F-isomorphism, F-automorphism*). Of course, an F-algebra homomomorphism is also a F-vector space linear transformation, but the converse is not true in general. Since these standard notations conflict with our usage of a module homomorphism if F is a field, a F-vector space homomorphism will be called an *F-linear homomorphism* or an *F-linear transformation* from now on.

Examples 50.7. 1. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. Define $\tilde{\varphi} : R[t] \to S[t]$ by $\sum a_i t^i \to \sum \varphi(a_i)t^i$, a ring homomorphism extending φ. *Given such a φ, we shall always let $\tilde{\varphi}$ denote this map.* If φ is a monomorphism, respectively an epimorphism, isomorphism, then $\tilde{\varphi}$ is a monomorphism, respectively an epimorphism, isomorphism.

2. Complex conjugation $^- : \mathbb{C} \to \mathbb{C}$ given by $a + b\sqrt{-1} \mapsto a - b\sqrt{-1}$, for $a, b \in \mathbb{R}$ is an $\mathbb{R}$-automorphism. In fact, it is an $\mathbb{R}$-*involution*, i.e., an automorphism of order two.
3. If F is a field, as we always view $F \subset F[t]/(f)$ for any non-constant polynomial f in $F[t]$, $F[t]/(f)$ is an F-algebra. [By the definition of R-algebra A given in Remark 26.11, we only need a ring homomorphism of a commutative ring R into the center of a ring A.]

To prove a uniqueness statement for splitting fields, we use an approach that we shall often use, viz., first prove a result for irreducible polynomials and then proceed to arbitrary ones. We begin with an analysis of irreducible polynomials.

Lemma 50.8. *Let $\varphi : F \to F'$ be a field isomorphism, f an irreducible polynomial in $F[t]$ and $f' = \tilde{\varphi}(f)$. Then*

(1) f' is an irreducible polynomial in $F'[t]$.

(2) There exists a field isomorphism $\tau : F[t]/(f) \to F'[t]/(f')$ extending φ and taking the root $\overline{t} = t + (f)$ in $F[t]/(f)$ to the root $\overline{t} = t + (f')$ in $F'[t]/(f')$.

Proof. (1) is clear, so both $F[t]/(f)$ and $F'[t]/(f')$ are fields. Let $^- : F'[t] \to F'[t]/(f')$ be the canonical epimorphism and μ the composition

$$F[t] \xrightarrow{\tilde{\varphi}} F'[t] \xrightarrow{-} F'[t]/(f').$$

As φ is an isomorphism, so is $\tilde{\varphi}$. In particular, μ is epic. By definition, $\ker{}^- = (f')$, so $\tilde{\varphi}$ an isomorphism implies that

$$\begin{aligned}\ker \mu &= \{h \in F[t] \mid \mu(h) = 0\} = \{h \in F[t] \mid \tilde{\varphi}(h) \in (f')\}\\ &= \{h \in F[t] \mid h \in (f)\} = (f).\end{aligned}$$

Therefore, μ induces an isomorphism $\overline{\mu} : F[t]/(f) \to F'[t]/(f')$ given by $g + (f) \mapsto \tilde{\varphi}(g)+(f')$ by the First Isomorphism Theorem. As $\overline{\mu}|_F = \mu|_F = \varphi$ and $\overline{\mu}\big(t+(f)\big) = t+(f')$, the map $\tau = \overline{\mu}$ works. $\square$

Lemma 50.9. *Let K/F be an extension of fields and f an irreducible polynomial in $F[t]$. If α in K is a root of f, then there exists an F-isomorphism*

$$\theta : F[t]/(f) \to F(\alpha) \ \textit{ given by } \ t + (f) \mapsto \alpha.$$

Proof. Let $e_\alpha : F[t] \to F[\alpha]$ by $g \mapsto g(\alpha)$ be the evaluation map at α, a ring epimorphism, and, in fact, an F-epimorphism. As $f(\alpha) = 0$, the element α is algebraic over F, so, as before, e_α induces an F-isomorphism $\theta : F[t]/\big(m_F(\alpha)\big) \to F[\alpha] = F(\alpha)$. As f is irreducible with root α, we have $(f) = \big(m_F(\alpha)\big)$ (by the proof of Theorem 48.13), so θ works. $\square$

By combining these two lemmas, we shall easily deduce the following:

Proposition 50.10. *Let $\varphi : F \to F'$ be a field isomorphism, f an irreducible polynomial in $F[t]$ and $f' = \tilde{\varphi}(f) \in F'[t]$. Suppose that K/F is a field extension with α a root of f in K and K'/F' a field extension with α' a root of f' in K'. Then there exists an isomorphism $\sigma : F(\alpha) \to F'(\alpha')$ extending φ and satisfying $\sigma(\alpha) = \alpha'$. Moreover, the multiplicity of the root α of f in $F(\alpha)$ is the same as the multiplicity of the root β of f in $F(\alpha')$.*

PROOF. The two lemmas determine isomorphisms

$$F(\alpha) \xleftarrow[\text{lifts } 1_F]{\theta} F[t]/(f) \xrightarrow[\text{lifts } \varphi]{\tau} F'[t]/(f') \xrightarrow[\text{lifts } 1_{F'}]{\theta'} F'(\alpha')$$

with $\theta\big(t + (f)\big) = \alpha$, $\tau\big(t + (f)\big) = t + (f')$, and $\theta'\big(t + (f')\big) = \alpha'$. The map $\theta' \circ \tau \circ \theta^{-1}$ works. Moreover, if $f = (t - \alpha)^r h$ in $F(\alpha)[t]$ with $f'(\alpha) \neq 0$, then the isomorphism $\tilde{\sigma}$ takes f to $f' = (t - \alpha')^r h'$ in $F(\alpha)[t]$ with $f(\alpha') \neq 0$ where $h' = \tilde{\sigma}(h)$. □

The following important corollary follows immediately:

Corollary 50.11. *Let K/F be an extension of fields, f an irreducible polynomial in $F[t]$ and α, β two roots of f in K. Then the map*

$$\sigma : F(\alpha) \to F(\beta) \ \text{ given by } \sum c_i \alpha^i \mapsto \sum c_i \beta^i$$

is an F-isomorphism. Moreover, the multiplicity of the root α of f in $F(\alpha)$ is the same as the multiplicity of the root β of f in $F(\beta)$.

The previous proposition is the first step in the induction proof of the result to which we have been striving.

Theorem 50.12. *Let $\varphi : F \to F'$ be an isomorphism of fields, f a non-constant polynomial in $F[t]$ and $f' = \tilde{\varphi}(f) \in F'[t]$. Suppose that E is a splitting field of f over F and E' is a splitting field of f' over F'. Then there exists an isomorphism $\tau : E \to E'$ extending φ and taking the set of roots of f in E bijectively onto the set of roots of f' in E' (and preserving multiplicities).*

An immediate consequence of this theorem is the result that we had wished to prove.

Corollary 50.13. (Uniqueness of Splitting Fields) *Let F be a field and f a non-constant polynomial in $F[t]$. If E and E' are splitting fields of f over F, then there exists an F-isomorphism $E \to E'$ taking the set of roots of f in E bijectively onto the set of roots of f in E' (and preserving multiplicities). In particular, a splitting field of f over F is unique up to an F-isomorphism.*

Of course, one would like to say given a polynomial there is a unique splitting field. We could do this if we assume that all the fields in which we work lie in some larger field Ω. Then one could say that a field extension E/F is *the* splitting field of a non-constant polynomial f in $F[t]$, as the splitting field of f is uniquely determined by the roots of f in Ω. For example, we view all splitting fields of rational polynomials to lie in $\mathbb{C}$.

PROOF. (of the theorem.) We induct on $n = [E : F]$.

$n = 1$: We have $E = F$, so f splits over F, hence f' splits over F'. In particular, if $f = a \prod(t - \alpha_i)$ in $F[t]$, then $f' = \varphi(a) \prod \big(t - \varphi(\alpha_i)\big)$ in $F'[t]$, and the result follows.

$n > 1$: If f does not split over F, then $f = gh$, for some polynomials g and h in $F[t]$, with g an irreducible polynomial satisfying $\deg g > 1$. As $\tilde{\varphi} : F[t] \to F'[t]$ is an isomorphism, $f' = g'h'$ with $g' = \tilde{\varphi}(g)$ irreducible of degree greater than one, and $h' = \tilde{\varphi}(h)$. Let $\alpha \in E$ be a root of g in E and $\alpha' \in E'$ be a root of g' in E' (which exist as f, f' split over E, E', respectively). By Proposition 50.10, there exists an isomorphism $\tau_1 : F(\alpha) \to F'(\alpha')$ extending φ and mapping α to α'. Let $f = (t-\alpha)f_1$ in $F(\alpha)[t]$. Then $\tilde{\varphi}(f) = (t-\alpha')f_1'$ in $F'(\alpha')[t]$ with $f_1' = \tilde{\tau}_1(f_1)$ and $\tilde{\varphi}(f) = \tilde{\tau}_1(f)$.

We know that E is a splitting field of f_1 over $F(\alpha)$ and E' is a splitting field of f_1' over $F(\alpha')$. Since $[E : F(\alpha)] < [E : F]$, induction provides an isomorphism $\tau : E \to E'$ lifting τ_1, hence φ, and taking the set of roots of f_1 bijectively onto the set of roots of f_1' (and preserving multiplicity of roots). The result now follows. $\square$

The proof above filled in the maps in the following picture:

$$\begin{tikzcd}
E \arrow[r, dashed, "\tau"] \arrow[d, dash] & E' \arrow[d, dash] \\
F(\alpha) \arrow[r, dashed, "\tau_1"] \arrow[d, dash] & F'(\alpha') \arrow[d, dash] \\
F \arrow[r, "\varphi"] & F'
\end{tikzcd}$$

where we shall always use the vertical lines in such pictures as the appropriate set inclusion.

We now look at what we have accomplished, in the following crucial discussion.

Remark 50.14. Let F be a field, f a non-constant polynomial in $F[t]$ of degree n, and L a splitting field of f over F. Let $L/K/F$. Define

$$\mathrm{Aut}_F(K) := \{\tau : K \to K \mid \tau \text{ is an } F\text{-automorphism}\},$$

a group called the *Galois group* of K *over* F. The Galois group $\mathrm{Aut}_F(L)$ is also called the *Galois group* of f, although if we omit having a fixed splitting field, then it is only unique up to isomorphism. Now let g be a non-constant polynomial in $F[t]$ and set $S = \{x \in K \mid g(x) = 0\}$, i.e., the set of roots of g in K. This is of interest when S is not empty. Let $\tau \in \mathrm{Aut}_F(K)$, then $\tilde{\tau}$ fixes the coefficients of any g in $F[t]$, so $\tilde{\tau}(g) = g$. Therefore, if $x \in K$, we have

$$\tau\big(g(x)\big) = \big(\tilde{\tau}(g)\big)\big(\tau(x)\big) = g\big(\tau(x)\big).$$

Thus, if S is non-empty and $x \in S$, we have

$$0 = \tau\big(g(x)\big) = g\big(\tau(x)\big),\ i.e., \tau(x) \in S.$$

So we have a map,

$$\mathrm{Aut}_F(K) \to \Sigma(S) \text{ given by the restriction } \tau \mapsto \tau|_S.$$

Check that this is, in fact, a group homomorphism.

Now let $L = K$ and $f = g$ and let $\alpha_1, \dots, \alpha_n$ be the roots of f in K, not necessarily distinct. We know that $K = F(\alpha_1, \dots, \alpha_n) = F[\alpha_1, \dots, \alpha_n]$. Let τ and ψ lie in $\mathrm{Aut}_F(K)$, and suppose that $\tau|_S = \psi|_S$. Then $\tau^{-1} \circ \psi$ fixes F and S, so $\tau^{-1} \circ \psi = 1_K$, as $K = F(S) = F[S]$. We conclude that $\tau|_S = \psi|_S$ if and only if $\tau = \psi$. Therefore, the map

$\mathrm{Aut}_F(K) \to \Sigma(S)$ given by $\tau \mapsto \tau|_S$ is injective, so a group monomorphism, and we can view $\mathrm{Aut}_F(K)$ as a subgroup of $\Sigma(S)$ and $\Sigma(S)$ a subgroup of S_n with $\Sigma(S) = S_n$ if and only if $\alpha_1, \dots, \alpha_n$ are distinct, i.e., f has no multiple roots. Suppose now, in addition, that f is irreducible in $F[t]$. Fix $i, j,\ 1 \le i, j \le n$, then K is a splitting field of f over $F(\alpha_i)$ and K is a splitting field of f over $F(\alpha_j)$. We have shown that there exists an F-isomorphism $\sigma : F(\alpha_i) \to F(\alpha_j)$ sending α_i to α_j. The theorem now says that there exists τ in $\mathrm{Aut}_F(K)$ extending σ and taking S onto S bijectively. This means that

$$\mathrm{Aut}_F(K) \subset \Sigma(S) \text{ is a transitive subgroup,}$$

i.e., for every pair of roots α_i and α_j in S, there exists a $\tau \in \mathrm{Aut}_F(K)$ satisfying $\tau(\alpha_i) = \alpha_j$. We shall see in the sequel, that this will essentially reduce the theory of algebraic extensions to group theory. This is what Galois discovered. In fact, it is important to note that the proof above shows more. If $h \in F[t]$ is irreducible and has roots β and β' in K, the splitting field of f over F, then there exists a automorphism $\tau \in \mathrm{Aut}_R(K)$ such that $\tau(\beta) = \beta'$. Therefore, $\mathrm{Aut}_F(K)$ also acts transitively on all the roots (if any) in K of an irreducible polynomial in $F[t]$. In fact, we shall prove in Proposition 56.2 below that any irreducible polynomial in $F[t]$ which has a root in K splits in K when K/F is a splitting field of some polynomial over F. You have the tools to prove this now. Can you?

We now give a number of examples to illustrate what we have done.

Examples 50.15. Let F be a field.

1. Let $f = t^2 + bt + c$ be a polynomial in $F[t]$ and K/F an extension field with α in K a root of f. If α' is another root of f or $\alpha = \alpha'$ is a *multiple root* of f, then $f = (t-\alpha)(t-\alpha') = t^2 - (\alpha+\alpha')t + \alpha\alpha'$ in $K[t]$. So we have $b = -\alpha - \alpha'$ and $c = \alpha\alpha'$. In particular, $\alpha' = -\alpha - b$ lies in K. If $\alpha = \alpha'$ is a multiple root, then $f = (t-\alpha)^2$ in $K[t]$ and $2\alpha = -b$. If, in addition, $\mathrm{char}\, F \neq 2$, then 2 is a unit and we have $\alpha = -b/2$ and $f = (t-\alpha)^2$ in $F[t]$ splits over F. In the general case, $f = (t-\alpha)(t-\alpha')$ splits over K and $F(\alpha)$ is a splitting field of f over F, so either $F = F(\alpha)$ and f splits over F or $[F(\alpha) : F] = 2$ depending on whether f is irreducible or not. If $\mathrm{char}\, F \neq 2$, this depends on whether $b^2 - 4c$ is a square in F or not, as the quadratic formula can be derived over F if 2 is a unit, just as over the real numbers.
2. The field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $(t^2-2)(t^2-3)$ in $\mathbb{C}$. We already know that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$ and $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a $\mathbb{Q}$-basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.
3. Let $f = t^3 - 2$ in $\mathbb{Q}[t]$. It is irreducible by Eisenstein's Criterion. Let $\omega = (-1 + \sqrt{-3})/2 = \cos(\frac{2}{3}\pi) + \sqrt{-1}\sin(\frac{2}{3}\pi)$ We have the complex conjugate of ω satisfies $\overline{\omega} = \omega^2$ and $\omega^3 = 1$. The roots of f in $\mathbb{C}$ are: $\sqrt[3]{2}$, $\sqrt[3]{2}\,\omega$, $\sqrt[3]{2}\,\omega^2$, all distinct. Then $K = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}\,\omega, \sqrt[3]{2}\,\omega^2) = \mathbb{Q}(\sqrt[3]{2}, \omega)$ is a splitting field of f over $\mathbb{Q}$ with $[K : \mathbb{Q}] \le 3! = 6$. As f is monic irreducible, we have $f = m_{\mathbb{Q}}(\sqrt[3]{2})$ and, therefore, $3 = \deg f = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] \mid [K : \mathbb{Q}]$. As $\omega = \sqrt[3]{2}\,\omega/\sqrt[3]{2}$ is not real, it does not lie in $\mathbb{Q}(\sqrt[3]{2})$. Thus $[K : \mathbb{Q}] > 3$, hence $[K : \mathbb{Q}] = 6$. [Alternatively, $m_{\mathbb{Q}}(\omega) = t^2 + t + 1$ splits in K, so $\mathbb{Q}(\omega) \subset K$ and

$2 = \deg m_{\mathbb{Q}}(\omega) \mid [K : \mathbb{Q}]$, so $[K : \mathbb{Q}] = 6$.] We also have

$$\begin{tikzcd}
& K = \mathbb{Q}(\sqrt[3]{2}, \omega) \arrow[dl, dash, "2"'] \arrow[dd, dash, "6"] \arrow[dr, dash, "3"] & \\
\mathbb{Q}(\sqrt[3]{2}) \arrow[dr, dash, "3"'] & & \mathbb{Q}(\omega) \arrow[dl, dash, "2"] \\
& \mathbb{Q} &
\end{tikzcd}$$

(where the integers in the diagram are the field extension degrees.)

4. Let $p > 1$ be a prime and $f = t^{p-1} + t^{p-2} + \cdots + t + 1$ in $\mathbb{Q}[t]$. We know that f is irreducible by Remark 35.11(3). Let $\zeta = e^{2\pi\sqrt{-1}/p} = \cos(2\pi/p) + \sqrt{-1}\sin(2\pi/p)$. Then $t^p - 1 = (t-1)f$ is a factorization into irreducibles in $\mathbb{Q}[t]$. So $\zeta^p = 1$ with $\zeta \neq 1$. As $f(\zeta) = 0$, we have that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p - 1 = \varphi(p)$. Similarly, $\zeta^2, \dots, \zeta^{p-1}$ are roots of f, all distinct and different than ζ. It follows that $\mathbb{Q}(\zeta)$ is a splitting field of f (and of $t^p - 1$) over $\mathbb{Q}$. The polynomial f is usually denoted by Φ_p and called the *pth cyclotomic polynomial* in $\mathbb{Q}[t]$ and satisfies $\deg \Phi_p = \varphi(p)$.

5. The polynomial $f = t^4 + t^2 + 1 = (t^2 + t + 1)(t^2 - t + 1)$ in $\mathbb{Q}[t]$ has roots $\zeta = (-1 + \sqrt{-3})/2$, $\overline{\zeta} = \zeta^2$, $\eta = (1 + \sqrt{-3})/2$, and $\overline{\eta}$, where $^-$ is complex conjugation. We have $K = \mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta, \eta)$ is a splitting field of f over $\mathbb{Q}$ and $[K : \mathbb{Q}] = 2$. Note that if x is a root of f, then $x^6 = 1$. The polynomial $t^2 - t + 1 = m_{\mathbb{Q}}(\eta)$ is the *6th cyclotomic polynomial* over $\mathbb{Q}$. It is denoted by Φ_6 and satisfies $\deg \Phi_6 = \varphi(6)$.

6. Let p be an odd prime, r a positive integer, and $\zeta = e^{2\pi\sqrt{-1}/p^r}$.

   Claim: $\mathbb{Q}(\zeta)$ is the splitting field of $t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$ and $t^{p^r} - 1 \in \mathbb{Q}[t]$ over $\mathbb{Q}$ and $[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(p^r) = p^{r-1}(p-1)$ (The same is true for $\zeta = e^{2\pi n\sqrt{-1}/p^r}$ for any n not divisible by p satisfying $1 \leq n < p^r$). In particular, $\Phi_{p^r} = m_{\mathbb{Q}}(\zeta)$, the p^rth *cyclotomic polynomial* over $\mathbb{Q}$ satisfies $\deg \Phi_{p^r} = \varphi(p^r)$:

   By Exercise 6.14(7), we know that $\varphi(p^r) = p^{r-1}(p-1)$, so it suffices to show $m_{\mathbb{Q}}(\zeta) = t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$, as ζ^i, $1 \leq i \leq p^r$, are p^r distinct roots of $t^{p^r} - 1$. We need the following identity of binomial coefficients that is easily proven by induction:

$$\sum_{i=j}^{n} \binom{i}{i-j} = \binom{n+1}{n-j}. \tag{50.16}$$

   Let $f = t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1 \in \mathbb{Z}[t]$ and $t = y + 1$. If we can show that $g(y+1) = f(t)$ satisfies Eisenstein's Criterion for p, we are done by Remark 35.11(2). By the Children's Binomial Theorem (Exercise 27.20(10)),

$$(y+1)^{p^{r-1}i} \equiv (y^{p^{r-1}} + 1)^i = \sum_{j=0}^{i} \binom{i}{j} y^{p^{r-1}j} \mod p\mathbb{Z}[t].$$

We have

$$g(y+1) = \sum_{i=0}^{p-1}\sum_{j=0}^{i}\binom{i}{j}y^{p^{r-1}j} \mod p\mathbb{Z}[t].$$

The coefficient of $y^{p^{r-1}j}$ modulo $p\mathbb{Z}[t]$ for $j \neq 0$ is

$$\sum_{i=j}^{p-1}\binom{i}{i-j} = \binom{p}{p-j} \equiv 0 \mod p.$$

The constant term is precisely p, so the result follows.

7. Let $f = t^3 + t + 1$ in $(\mathbb{Z}/2\mathbb{Z})[t]$. Check that f has no roots in $(\mathbb{Z}/2\mathbb{Z})$.

Observation 50.17. If F is a field and a non-constant polynomial g in $F[t]$ has no roots in F and is of degree at most three, then g is irreducible in $F[t]$.

Therefore, the polynomial f is irreducible in $(\mathbb{Z}/2\mathbb{Z})[t]$. Suppose that $L/(\mathbb{Z}/2\mathbb{Z})$ is an extension field such that f has a root α in L and let $K = (\mathbb{Z}/2\mathbb{Z})(\alpha)$. We have $\alpha^3 = -\alpha - 1 = \alpha + 1$ in K, so $\alpha^{-1} = \alpha^2 + 1$. Let $f = (t-\alpha)(t^2 + at + b)$ in $K[t]$. Multiplying out, yields $\alpha = a$ and $b = 1 + \alpha^2 = \alpha^{-1}$. It follows that $t^2 + at + b = t^2 + \alpha t + (1 + \alpha^2)$ in $K[t]$. If $e_{\alpha^2} : K[t] \to K$ is the evaluation map at α^2, then

$$\begin{aligned}
e_{\alpha^2}(t^2 + at + b) &= e_{\alpha^2}\big(t^2 + \alpha t + (1+\alpha^2)\big) = \alpha^4 + \alpha^3 + (1+\alpha^2) \\
&= (\alpha^2 + \alpha) + \alpha^3 + (1+\alpha^2) \\
&= (\alpha^2 + \alpha) + (\alpha + 1) + (1 + \alpha^2) = 0.
\end{aligned}$$

Consequently, α^2 is also a root of f in K with $\alpha \neq \alpha^2$. So we have $f = (t-\alpha)(t-\alpha^2)g$ in $K[t]$ with $\deg g = 1$. Therefore, f splits over $K = (\mathbb{Z}/2\mathbb{Z})(\alpha)$, hence K is a splitting field of f over $(\mathbb{Z}/2\mathbb{Z})$ with $[K : (\mathbb{Z}/2\mathbb{Z})] = \deg f = 3$. Note that K is a field having $2^3 = 8$ elements.

8. Let $p > 0$ be a prime, F a field of characteristic p, and α an element in F. Then $t^p - \alpha^p = (t-\alpha)^p$ in $F[t]$ by the Children's Binomial Theorem (Exercise 27.20(10)), has F as a splitting field of f and f has only one distinct root of multiplicity p. Now let L/F be an extension field, α an element in L satisfying $a = \alpha^p$ lies in F. Then $K = F(\alpha)$ is a splitting field of $t^p - a$ in $F[t]$ over F.

Claim. If α is not an element in F, then $[K : F] = p$:

Suppose that $t^p - \alpha = fg$ in $F[t]$ with f and g lying in $F[t]$. We may assume that both f and g are monic and f not a unit. As $fg = t^p - a = (t-\alpha)^p$ in the UFD $K[t]$, we must have $f = (t-\alpha)^r$ in $K[t]$ for some r satisfying $1 \leq r \leq p$. As f lies in $F[t]$, its constant term $\pm\alpha^r$ must lie in F. If $r < p$, then r and p are relatively prime, so α^r, α^p lying in F implies that α lies in F, a contradiction. Therefore, $g = 1$ and $r = p$, so $t^p - a$ is irreducible in $F[t]$.

Exercises 50.18.

1. If $f \in \mathbb{Q}[t]$ and K is a splitting field of f over $\mathbb{Q}$, determine $[K : \mathbb{Q}]$ if f is:
   (a) $t^4 + 1$.
   (b) $t^6 + 1$.

(c) $t^4 - 2$.
(d) $t^6 - 2$.
(e) $t^6 + t^3 + 1$.

2. Find the splitting fields K for $f \in \mathbb{Q}[t]$ and $[K : \mathbb{Q}]$ if f is:
   (a) $t^4 - 5t^2 + 6$.
   (b) $t^6 - 1$.
   (c) $t^6 - 8$.
3. Show both of the following:
   (a) If $K/\mathbb{Q}$ and $\sigma \in \operatorname{Aut} K$ then σ fixes $\mathbb{Q}$.
   (b) The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ are not isomorphic.
4. Prove Identity 50.16.
5. A *primitive nth root of unity* is an element $z \in \mathbb{C}$ such that $z^n = 1$ and $z^r \neq 1$ for $1 \leq r < n$. Show the following:
   (a) There exist $\varphi(n) := |\{d \mid 0 \leq d \leq n, \quad (d, n) = 1\}|$ primitive nth roots of unity.
   (b) If ω is a primitive nth root of unity, then $\mathbb{Q}(\omega)$ is a splitting field of $t^n - 1 \in \mathbb{Q}[t]$.
   (c) Let $\omega_1, ..., \omega_{\varphi(n)}$ be the $\varphi(n)$ primitive nth roots of unity of $t^n - 1 \in \mathbb{Q}[t]$ and $\sigma \in \operatorname{Aut} \mathbb{Q}(\omega_1)$ then $\sigma(\omega_1) = \omega_i$ for some $i,\ 1 \leq i \leq \varphi(n)$.
6. Continued from Exercise 5. Show:
   (a) Let $\Phi_n(t) = (t - \omega_1) \cdots (t - \omega_{\varphi(n)})$. Then show $\Phi_n(t) \in \mathbb{Q}[t]$. $\Phi_n(t)$ is called the nth *cyclotomic polynomial*.
   (b) $\Phi_n(t) \in \mathbb{Z}[t]$.
7. Continued from Exercise 6. Show:
   (a) $\Phi_n(t) \in \mathbf{Z}[t]$ is irreducible.
   (b) Calculate $\Phi_n(t)$ for $n = 3, 4, 6, 8$ explicitly and show directly that $\Phi_n(t) \in \mathbb{Z}[t]$ is irreducible.
8. Let $F = \mathbb{Z}/p\mathbb{Z}$. Show all of the following:
   (a) There exists a polynomial $f \in F[t]$ satisfying $\deg f = 2$ and f irreducible.
   (b) Use the f in (a) to construct a field with p^2 elements.
   (c) If $f_1, f_2 \in F[t]$ have $\deg f_i = 2$ and f_i irreducible for $i = 1, 2$, show that their splitting fields are isomorphic.
9. Let F be a field of characteristic different from two and $f = at^2 + bt + c$ in $F[t]$. State and prove the quadratic formula.
10. Let K/F be an extension of fields and $f \in F[t]$.
    (a) If $\varphi : K \to K$ is an F-automorphism, then φ takes roots of f in K to roots of f in K.
    (b) If $F \subset \mathbb{R}$ and $\alpha = a + b\sqrt{-1}$ is a root of f with $a, b \in \mathbb{R}$ then $\overline{\alpha} = a - b\sqrt{-1}$ is also a root of f.
    (c) Let $F = \mathbb{Q}$. If $m \in \mathbb{Z}$ is not a square and $a + b\sqrt{m}$ in $\mathbb{C}$ is a root of f with a, b in $\mathbb{Q}$ then $a - b\sqrt{m}$ is also a root of f in $\mathbb{C}$.
11. Any (field) automorphism $\varphi : \mathbb{R} \to \mathbb{R}$ is the identity automorphism.
12. Let $f = (t^2 - p_1) \cdots (t^2 - p_n)$ in $\mathbb{Q}[t]$ with $p_1, \ldots, p_n$ be n distinct prime numbers. Show that $K = \mathbb{Q}[\sqrt{p_1}, \ldots, \sqrt{p_n}]$ is a splitting field of f over $\mathbb{Q}$ and $[K : \mathbb{Q}] = 2^n$. Formulate a generalization of the statement for which your proof still works.

13. Find a splitting field of $f \in F[t]$ if $F = \mathbb{Z}/p\mathbb{Z}$ and $f = t^{p^e} - t$, $e > 0$.
14. Let K be a splitting field over F of f in $F[t] \setminus F$ and g an irreducible polynomial in $F[t]$. Suppose that g has a root in K. Show that g splits over K.
15. Suppose F is a field and K is a splitting field of $f \in F[t]$ over F. If $K/E/F$ is an intermediate field, $\sigma : E \to K$ be a monomorphism fixing F, show that there exists an F-automorphism $\widehat{\sigma} : K \to K$ lifting σ, i.e., such that $\widehat{\sigma}|_E = \sigma$.
16. Let F be a field of characteristic $p > 0$. Show that $f = t^4 + 1 \in F[t]$ is not irreducible. Let K be a splitting field of f over F. Determine which finite fields F satisfies $K = F$.
17. Let $f = t^6 - 3 \in F[t]$. Construct a splitting field K of f over F and determine $[K : F]$ for each of the cases: $F = \mathbb{Q}, \mathbf{Z}/5\mathbb{Z}$, or $\mathbb{Z}/7\mathbb{Z}$. Do the same thing if f is replaced by $g = t^6 + 3 \in F[t]$ for the same fields F.

51. Algebraically Closed Fields

Recall that a field K is called *algebraically closed* if every non-constant polynomial in $K[t]$ has a root in K, equivalently that every non-constant polynomial in $K[t]$ splits over F. Clearly, this is also equivalent to the condition that if L/K is an algebraic extension of fields, then $L = K$. We call an extension field K of F an *algebraic closure* of F if K/F is algebraic and K is algebraically closed. We shall later prove the Fundamental Theorem of Algebra that $\mathbb{C}$ is an algebraically closed field. We know that $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$. In fact, assuming the Fundamental Theorem of Algebra, we showed in Example 48.23(4) that $\Omega := \{x \in \mathbb{C} \mid x \text{ algebraic over } \mathbb{Q}\}$ is an algebraic closure of $\mathbb{Q}$ and is countable. In this section, we shall show that given any field F, there exists an algebraic closure of F, and it is unique up to an F-homomorphism. By the argument in Example 48.23(5), to establish the existence of an algebraic closure of F, it suffices to show that F is a subfield of some algebraically closed field.

We know that if we have a finite set of polynomials, $\{f_1, \ldots, f_n\}$, then we can construct a field K such that each f_i splits in K and K is the smallest such field by letting K be the splitting field of $\prod f_i$ over F. Another way of constructing this K would be to find a splitting field K_1 of f_1 over F, then K_2 a splitting field of f_2 over K_1, and continue in this way. We use both for the case of an infinite set of polynomials.

To use these two ideas, we first set up some notation. Let I be an indexing set and $S = \{t_i \mid i \in I\}$ be a set of distinct indeterminants. Then $F[S]$ is the set of polynomials, coefficients in F in the t_i's, $i \in I$. If $J \subset I$, let $S_J = \{t_j \mid j \in J\}$. As each such polynomial needs only finitely many indeterminants to determine it, i.e., $f \in F[S_J]$, for some finite subset $J \subset I$, we have

$$F[S] = \bigcup_{\substack{J \subset I \\ J \text{ finite}}} F[S_J].$$

Note that $F[S]$ is a domain, and in fact, a UFD. Using this notation, we give Artin's proof of the following theorem.

Theorem 51.1. *Let F be a field. Then there exists an algebraically closed field K containing F.*

PROOF. Let $I = F[t] \setminus F$ be our indexing set and $S = \{t_f \mid f \in I\}$. So if $f \in F[t]$, we have $f(t_f)$ lies in $F[t_f] \subset F[S]$. Set $T = \{f(t_f) \mid f \in F[t] \setminus F\}$ and $\mathfrak{A}$ the ideal in $F[S]$ generated by the set T. So

$$\mathfrak{A} = \{\sum_{finite} g_i f_i(t_{f_i}) \mid g_i \in F[S]\}.$$

Claim. $\mathfrak{A} < F[S]$:
Suppose that $\mathfrak{A} = F[S]$. Then there exist $f_i(t_{f_i}) \in T$ and $g_i \in F[S]$, $i = 1, \dots, n$, some n, satisfying

$$1 = g_1 f_1(t_{f_1}) + \cdots + g_n f_n(t_{f_n})$$

and a finite subset $J \subset I$ with $g_1, \dots, g_n$ lying in $F[S_J]$. Relabel each t_{f_i} as t_i and the t_j's in J such that $J = \{t_1, \dots, t_N\}$ some N, so $f_i(t_i), g_i \in F[t_1, \dots, t_N]$. We know that there exists a field extension K/F such that each $f_i \in F[t]$ has a root α_i in K for $1 \le i \le n$ by the preliminary remarks. Set $\alpha_i = 0$ for $n < i \le N$. Applying the evaluation map $e_{\alpha_1,\dots,\alpha_n}$ yields

$$0 = \sum g_i(\alpha_1, \dots, \alpha_n) f_i(\alpha_i) = 1$$

in K, a contradiction. This establishes the claim.
By the claim, there exists a maximal ideal $\mathfrak{m}$ satisfying $\mathfrak{A} \subset \mathfrak{m} < F[S]$ (which we know exists by Zorn's Lemma). Define $L_1 := F[S]/\mathfrak{m}$, viewed as a field extension of F in the usual way (i.e., identifying F and $\varphi(F)$ where $\varphi : F[S] \to L$ is given by $g \mapsto g + \mathfrak{m}$). By construction, if f is a non-constant polynomial in $F[t]$, then f has a root $t_f + \mathfrak{m}$ in L_1. Inductively, define L_i such that L_i/L_{i-1} and every non-constant polynomial $g \in L_{i-1}[t]$ has a root in L_i by repeating the same construction. Set $L = \cup L_i$. If α, β lie in L with β nonzero, then $\alpha \pm \beta, \alpha\beta, \beta^{-1}$ lie in some subfield L_i of L, hence L is a field. Finally, let f be a non-constant polynomial in $L[t]$. As f has finitely many nonzero coefficients, there exists an i such that f lies in $L_i[t]$ hence has root in $L_{i+1} \subset L$. Therefore, L is algebraically closed. □

We showed in the argument in Example 48.23(5) that if F is a field and K/F a field extension with K algebraically closed, then the elements in K algebraic over F form an algebraic closure of F. Therefore, we have established the following result.

Corollary 51.2. *Let F be a field. Then an algebraic closure of F exists.*

Algebraic closures are "super" splitting fields, so we expect a uniqueness statement. We prove this.

Theorem 51.3. *Let K/F be an algebraic extension of fields and L an algebraically closed field. Suppose that there exists a homomorphism $\sigma : F \to L$. Then there exists a homomorphism $\tau : K \to L$ that lifts σ. In particular, if K is also algebraically closed and $L/\sigma(F)$ is algebraic, then τ is an isomorphism.*

PROOF. We use a Zorn Lemma argument that is often repeated in algebra to define maps. Let

$$\begin{aligned} S := \{(E, \eta) \mid & K/E/F \text{ an intermediate field} \\ & \eta : E \to L \text{ a homomorphism lifting } \sigma\}. \end{aligned}$$

Partially order the set S by $\leq$ defined by

$$(E,\eta) \leq (E',\eta') \text{ if } E \subset E' \text{ and } \eta'|_E = \eta.$$

As (F,σ) lies in S, the set S is nonempty. Let $\mathcal{C} = \{(E_i,\eta_i) \mid i \in I\}$ be a chain in S. We know that $E = \cup_I E_i$ is a field. Define

$$\eta : E \to L$$

as follows: If α lies in E, choose $i \in I$ such that $\alpha \in E_i$ and set $\eta(\alpha) := \eta_i(\alpha)$. By the definition of our partial ordering, $\eta(\alpha)$ is independent of i, i.e., η is well-defined. It follows that (E,η) is an upper bound for $\mathcal{C}$. By Zorn's Lemma, there exists a maximal element (E,η) in S, so $K/E/F$ and $\eta : E \to L$ extends σ.

Claim. $K = E$:

Suppose this is false and $\alpha \in K \setminus E$. As α is algebraic over F, it is algebraic over E. Let β be a root of $\widetilde{\eta}\big(m_E(\alpha)\big)$ in $\eta(E)[t]$ in the algebraically closed field L. Then we know by Proposition 50.10 that there exists an isomorphism

$$\eta' : E(\alpha) \to \eta(E)(\beta) \text{ with } \alpha \mapsto \beta \text{ lifting } \eta.$$

It follows that $(E(\alpha),\eta')$ lies in S with $(E,\eta) < (E(\alpha),\eta')$, contradicting the maximality of (E,η). This proves the claim.

If K is algebraically closed so is the subfield $\tau(K)$ in L. If $L/\sigma(F)$ is algebraic, then $L/\tau(K)$ is algebraic, hence $L = \tau(K)$ and the theorem if proven. □

We immediately see that this implies the desired result.

Corollary 51.4. *Let F be a field. Then any algebraic closure of F is unique up to an F-isomorphism.*

Because of the corollary, one usually fixes an algebraic closure K of a field F and views all roots of all polynomials in $F[t]$ as lying in K. More generally, one fixes some E/F with E algebraic closed and only works with fields K with $E/K/F$. Then E contains a unique algebraic closure of F, hence for each polynomial in $F[t]$ a unique splitting field in E. One then views every algebraic extension of F to be in E. For example, one can let $E = \mathbb{C}$ when $F = \mathbb{Q}$. More generally, if F is a field of characteristic zero and $\widetilde{F}$ an algebraically closed field, we can view $F/\mathbb{Q}$ and the algebraic closure of $\mathbb{Q}$ in $\widetilde{F}$ in $\mathbb{C}$ by replacing it with an isomorphic copy. We shall always do so implicitly.

Exercises 51.5.

1. Let F be a finite field and K an algebraic closure of F. Show that K is countable and infinite. (In all other cases, an algebraic closure of a field has the same cardinality of the given field).
2. Let S be a collection of polynomials over a field F. Define what it means for an extension field K/F to be a *splitting field* of the collection S and prove that one exists and is unique up to an F-isomorphism. Also prove that an algebraic closure of F is a splitting field for the collection of irreducible polynomials in $F[t]$.

52. Constructible Numbers

The ancient Greeks were very interested in constructing geometric objects under certain rules. The most famous ones that they could not solve were the following:

Euclidean Construction Problems. Using a straight-edge and compass, can you:

I. Trisect a given angle?
II. (Delian Problem) Double a cube, i.e., construct the edge of a cube whose volume is twice the volume of a given cube?
III. Square a circle, i.e., construct a square whose area equals the area of a given circle?
IV. Construct a regular n-gon, $n \geq 3$?

Given a specific set of initial points, the following constructions were allowed:

C1. You can draw a line through two initial points.
C2. You can draw a circle with center at an initial point with radius equal to the length between two initial points.

From these constructions on your given set of initial points, you were allowed to enlarge this original set by adding the points of intersection arising from the constructions C1 and C2 and add further points by iterating this process.

We shall solve the Euclidean Problems, by converting them into problems in field theory. In the case of squaring the circle, some analysis will also be needed.

Algebraic Reformulation (Preliminaries). Let $S = \{P_1, \ldots, P_n\}$ be a finite nonempty set of points in the euclidean plane. Recursively define S_r as follows:

$$S_0 = S \text{ and having defined } S_r \text{ define } S_{r+1} := S_r \cup T_r$$

where T_r is the union of the following sets:

1. The set of points of intersections of lines through two distinct points in S_r (that we call S_r-*lines*).
2. The set of points of the intersection of circles having center in S_r and radii equal in length to line segments connecting two distinct points in S_r (that we call S_r-*circles*).
3. The set of intersections of S_r-circles and S_r-lines.

Now set

$$C(P_1, \ldots, P_n) := \bigcup_{r=0}^{\infty} S_r, \tag{52.1}$$

the *set of points constructible* from $P_1, \ldots, P_n$. We say a point P in the euclidean plane is *constructible from* $P_1, \ldots, P_n$, if $P \in C(P_1, \ldots, P_n)$ and *not constructible* otherwise. If $C = C(P_1, \ldots, P_n)$, call a line segment a C-*line* if an S_r-line for some r and a C-*circle* if an S_r-circle for some r.

Remarks 52.2. Let $P_1, \ldots, P_n$ be our initial given points in the euclidean plane and $C = C(P_1, \ldots, P_n)$ Using elementary euclidean geometry, we deduce the following:

1. $C(P_1) = \{P_1\}$. As this is not interesting, we shall always assume that we begin with at least two given points, i.e., $n \geq 2$.

2. We can begin to define a coordinate system of our euclidean plane by setting $P_1 = (0, 0)$, the origin, and $P_2 = (1, 0)$.

3. Given a line segment of length d between two points in C, we can construct a line segment of length md for all integers m, can bisect the given line segment between two points in C, and can construct a perpendicular at each endpoint of the line segment between two points in C. In particular, we can construct axes.

4. We can drop a perpendicular from a point in C to a given C-line as well as construct a parallel through a point in C to a C-line. In particular, we see that any point in $\mathbb{Z}^2$ is constructible from $P_1, \ldots, P_n$.

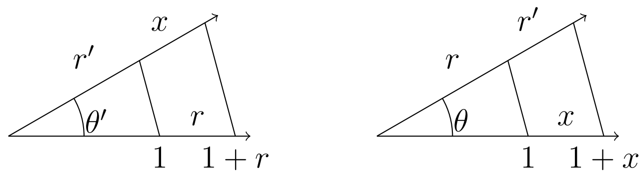
5. Given two C-lines of angles θ and φ with the X-axis respectively, we can construct rays (C-lines through the origin) of angles $\theta \pm \varphi,\ -\theta,\ \theta/2$. In fact, this can be done with the X-axis replaced by any allowable C-line.

It is now convenient to complexify everything, i.e., to replace the euclidean plane with the complex plane. Let $P_1, \ldots, P_n,\ n \geq 2$, be points in the euclidean plane. As above, we let $P_1 = (0, 0)$ and $P_2 = (1, 0)$. To complexify, if $P_i = (x_i, y_i)$ lie in $\mathbb{R}^2$, let $z_i = x_i + y_i\sqrt{-1}$, e.g., we have $z_1 = 0$ and $z_2 = 1$, and let $C(z_1, \ldots, z_n)$ denote the complex numbers constructible by straight-edge and compass from $z_1, \ldots, z_n$. We call $C(z_1, \ldots, z_n)$ the *set of complex constructible numbers from* $z_1, \ldots, z_n$. We have $(x, y) \in C(P_1, \ldots, P_n)$ if and only if $x + y\sqrt{-1} \in C(z_1, \ldots, z_n)$. Note that we already know that $C(z_1, \ldots, z_n)$ contains the Gaussian integers. But we have much more.

Theorem 52.3. *Let $n \geq 2$ and $z_1\,(= 0), z_2\,(= 1), \ldots, z_n$ be complex numbers. Then $C(z_1, \ldots, z_n)$ is a field satisfying $\mathbb{Q} \subset C(z_1, \ldots, z_n) \subset \mathbb{C}$.*

Proof. As $C(z_1, \ldots, z_n) \subset \mathbb{C}$ with $\operatorname{char} \mathbb{C} = 0$, it suffices to show that $C(z_1, \ldots, z_n)$ is a field. Let z, z' lie in $C(z_1, \ldots, z_n)$, with z nonzero. As we can draw parallel lines, we can add given vectors (line segments starting at the origin), so we see that $z \pm z'$ lies in $C(z_1, \ldots, z_n)$. We need to show that zz' and z^{-1} lie in $C(z_1, \ldots, z_n)$. Write z and z' in polar form, say $z = re^{\sqrt{-1}\theta}$ and $z' = r'e^{\sqrt{-1}\theta'}$ with r, r', θ, θ' real numbers, r, r' nonnegative. Then $zz' = rr'e^{\sqrt{-1}(\theta+\theta')}$ and $z^{-1} = r^{-1}e^{-\sqrt{-1}\theta}$. We know that we can construct rays of angle $\theta \pm \theta'$ and $-\theta$ and line segments of length $r = |z|$ and $r' = |z'|$. On the X-axis in the complex plane mark off 1 and $1 + r$ and the ray of angle θ' through z', which we may assume to be nonzero. Then the line through $1 + r$ parallel to the line through z' and 1 intersects the ray of an angle θ' with the X-axis at $(r + x)e^{\sqrt{-1}\theta'}$, and x is seen to be rr' by similar triangles. Reversing, the roles of x and r' constructs r/r'. (Cf. Figure 52.4 below.) As associativity, commutativity, etc. hold in $\mathbb{C}$, we see that $C(z_1, \ldots, z_n)$ is a field. $\square$

(52.4)



The field $C(z_1, \dots, z_n)$ is easily characterized using computation of the intersection points of lines and circles.

Theorem 52.5. *Let $z_1 \,(=0), z_2 \,(=1), \dots, z_n$ be points in $\mathbb{C}$ with $n \geq 2$. Then the field $C(z_1, \dots, z_n)$ has the following properties:*

(1) The point z_i lies in $C(z_1, \dots, z_n)$ for $i = 1, \dots, n$.
(2) If z lies in $C(z_1, \dots, z_n)$, so does its complex conjugate $\overline{z}$.
(3) if z^2 lies in $C(z_1, \dots, z_n)$, so does z, i.e., if z lies in $C(z_1, \dots, z_n)$, so does $\sqrt{z}$.

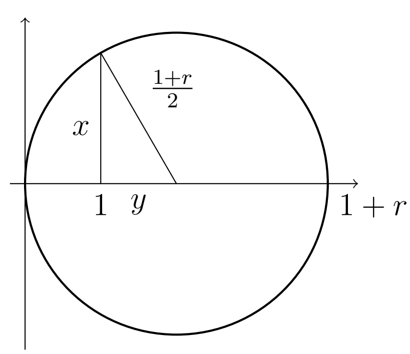
Moreover, $C(z_1, \dots, z_n)$ is the unique smallest subfield of $\mathbb{C}$ satisfying these properties, i.e., if a subfield C' of $\mathbb{C}$ satisfies (1), (2), and (3), then $C(z_1, \dots, z_n) \subset C'$.

PROOF. (1) is clear.

(2): If $z = re^{\sqrt{-1}\theta}$, then $\overline{z} = re^{-\sqrt{-1}\theta}$, and (2) follows.

(3): Since we can construct rays of angle $\theta/2$ from rays of angle θ, it suffices to show that we can construct $\sqrt{r}$ given a segment of length r. Draw a circle centered at the point $(1+r)/2$ on the X-axis of radius $(1+r)/2$. Erect a perpendicular at the point 1 on the X-axis parallel to the Y-axis. This line segment has length $\sqrt{r}$ by the Pythagorean Theorem. (Cf. Figure 52.6.)

(52.6)



This shows that $C(z_1, \dots, z_n)$ satisfies the desired properties. Suppose that C' also does. As $z_1, \dots, z_n$ all lie in C', by the definition (equation (52.1)) of $C(z_1, \dots, z_n)$, it suffices to show that intersection points of two C'-lines, two C'-circles, and a C'-line and a C'-circle, lie in C'. To do this note that $\sqrt{-1}$ lies in C' by (3) and if $z = a + b\sqrt{-1}$ with a, b in $\mathbb{R}$, then z lies in C' if and only if a and b lie in the field $C' \cap \mathbb{R}$, e.g., if z lies in C', then by (2), so does $b = -\sqrt{-1}(z - \overline{z})/2$ as C' is a field. This means that the equations defining C'-lines and C'-circles have coefficients in $C' \cap \mathbb{R}$, so solving these equations leads to solutions in C' (using (3) for intersections with circles). $\square$

We wish to further characterize the field theoretic properties of fields of constructible numbers. Property (3) in the theorem gives the key. Such fields are closed under taking square roots of elements in the field.

Definition 52.7. Let K/F be an extension of fields. We call K/F a *square root tower over* F if there exist elements $u_1, \dots, u_n$ in K for some positive integer n, satisfying:

1. $K = F(u_1, \dots, u_n)$.
2. u_1^2 lies in F.
3. u_i^2 lies in $F(u_1, \dots, u_{i-1})$ for each $i > 1$.

Remarks 52.8. Let K/F be a square root tower defined by $K = F(u_1, \dots, u_n)$ satisfying the above properties and E/F an extension field. Assume that both K and E lie in some larger field.

1. We have
$$[F(u_1, \dots, u_i) : F(u_1, \dots, u_{i-1})] = \begin{cases} 1 & \text{if } u_i \in F(u_1, \dots, u_{i-1}) \\ 2 & \text{otherwise.} \end{cases}$$
In particular, $[F(u_1, \dots, u_i) : F] = 2^e$ for some $e \leq n$.
2. The extension $E(K)/E$ is a square root tower. (Of course, $E(K) = E(u_1, \dots, u_n)$.)
3. If E/F is also a square root tower, then so is $E(K)/F$.
4. If E/K is a square root tower, then so is E/F
5. Suppose that $K \subset \mathbb{C}$ and $\overline{F} := \{\overline{z} \mid z \in F\}$, where $\overline{z}$ is the complex conjugate of z. Then $\overline{F}$ is a field and $\overline{K}/\overline{F}$ is a square root tower, with $\overline{K} = \overline{F}(\overline{z_1}, \dots, \overline{z_n})$.
6. Suppose that $\operatorname{char} F \neq 2$ and E/F satisfies $[E : F] \leq 2$. Then E/F is a square root tower:
We may assume that $[E : F] = 2$, so there exists an $\alpha \in E \setminus F$ satisfying $E = F(\alpha)$. Let $m_F(\alpha) = t^2 + bt + a$ in $F[t]$. Then $m_F(\alpha)$ splits over E and has roots α and $\alpha_1 = -b - \alpha$. . Let $\beta = 2\alpha + b$. As $\alpha^2 + b\alpha = -a$ and β does not lie in F, we see that $E = F(\beta)$ and $\beta^2 = -4a + b^2$ lies in F.

In general, it is not true if a field extension E/F has degree a power of two, then it is a square root tower. [Can you find an example?] However, we shall prove the following theorem later (cf. Theorem 57.11):

Theorem 52.9. (Square Root Tower Theorem) *Let F be a field of characteristic zero and K/F an extension of fields that is a splitting field of some non-constant polynomial in $F[t]$. Then K/F is a square root tower if and only if K/F is a field extension of degree a power of two.*

In fact, we shall prove a stronger result, that will be applicable to fields of arbitrary characteristic. We shall use this result when needed later in this section.

The notion of square root tower allows us to give a sufficient condition for a complex number to be constructible from a set of complex numbers.

Theorem 52.10. (Constructibility Criterion) *Let $z_1 \,(= 0), z_2 \,(= 1), \dots, z_n$ be complex numbers with $n \geq 2$. Set $F = \mathbb{Q}(z_1, z_2, \dots, z_n, \overline{z_1}, \dots, \overline{z_n})$. If z is a complex number, then*

z is constructible from $z_1, z_2, \ldots, z_n$, i.e., z lies in $C(z_1, z_2, \ldots, z_n)$ if and only if there exists a square root tower K/F (with $K \subset \mathbb{C}$) satisfying $z \in K$.

PROOF. Let

$$C' = \{z \in \mathbb{C} \mid \text{There exists a square root tower } K/F \text{ with } z \in K\}.$$

By definition, $F \subset C'$. We must show that $C' \subset C(z_1, \ldots, z_n)$. We know that $C(z_1, \ldots, z_n)$ is a field containing $z_1, \ldots, z_n$ closed under complex conjugation and extracting square roots. In particular, F is a subfield of $C(z_1, \ldots, z_n)$ as is any square root tower K/F. In particular, $C' \subset C(z_1, \ldots, z_n)$. To show that $C(z_1, \ldots, z_n) \subset C'$, it suffices to show by Theorem 52.5 that C' is a field closed under complex conjugation and extracting square roots. Let z and z' be elements in C' satisfying $z \in K$ and $z' \in K'$ with K/F and K'/F square root towers. Then $K(K')/F$ is a square root tower by Remark (3) above containing zz', $z \pm z'$, and z^{-1} (if $z \neq 0$). As $K(K') \subset C'$, we have C' is a field. Moreover, $K(\sqrt{z})/K$ is a square root tower, hence $K(\sqrt{z})/F$ is also a square root tower as K/F is. Since $F = \overline{F}$, we also have $\overline{K}/F$ is a square root tower. Therefore, $\sqrt{z}$ and $\overline{z}$ lie in C' and $C(z_1, \ldots, z_n) \subset C'$ as needed. □

Corollary 52.11. *Let $z_1\,(=0), z_2\,(=1), \ldots, z_n$ be complex numbers with $n \geq 2$ and z a complex number in $C(z_1, \ldots, z_n)$. Then z is algebraic over $F = \mathbb{Q}(z_1, z_2, \ldots, z_n, \overline{z_1}, \ldots, \overline{z_n})$ and $[F(z) : F]$ is a power of two.*

PROOF. As z lies in $C(z_1, \ldots, z_n)$, there exists a square root tower K/F with $z \in K$, hence $[F(z) : F] \mid [K : F]$, a power of two. □

In many euclidean constructions, one starts with two points, P_1 and P_2, which we may assume correspond to complex numbers $z_1 = 0$ and $z_2 = 1$. In this case, $\mathbb{Q} = \mathbb{Q}(z_1, z_2, \overline{z_1}, \overline{z_2})$ and $C(z_1, z_2)$ is called the field of (*complex euclidean*) *constructible numbers* and any $z \in C(z_1, z_2)$ is called *constructible*. So if z is a constructible number, we have $[\mathbb{Q}(z) : \mathbb{Q}]$ is a power of two.

We now can solve the (Greek) Euclidean Construction Problems. Recall that a cubic (or quadratic) polynomial in $F[t]$, F a field, is irreducible if and only if it does not have a root in F.

Trisection of an angle θ: To solve this construction problem, we need the following lemma.

Lemma 52.12. *Let α be a ray with the X-axis as one side. Then it is constructible form $z_1\,(=0), z_2\,(=1), \ldots, z_n$ if and only if $e^{\sqrt{-1}\alpha}$ is constructible from $z_1\,(=0), z_2\,(=1), \ldots, z_n$ if and only if $\cos\alpha$ is constructible from $z_1\,(=0), z_2\,(=1), \ldots, z_n$.*

PROOF. As $z = e^{\sqrt{-1}\alpha} = \cos\alpha + \sqrt{-1}\sin\alpha$ and we can erect and drop perpendiculars, this is immediate. □

The trisection of an angle θ is, therefore, equivalent to:

Problem. Given $z_1 = 0, z_2 = 1, z_3 = \cos\theta$, can we construct $\cos(\theta/3)$ from $z_1 = 0, z_2 = 1, z_3$?

Let $F = \mathbb{Q}(\cos\theta, \overline{\cos\theta}) = \mathbb{Q}(\cos\theta)$. For any angle α, we have the trigonometric identity

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha.$$

So $\beta = \cos(\theta/3)$ is a root in $\mathbb{C}$ of $f = 4t^3 - 3t - \cos\theta$ in $F[t]$. The polynomial is *reducible*, i.e., not irreducible in $F[t]$ if and only if f has a root in F if and only if for any root γ of f in $\mathbb{C}$, we have $[F(\gamma) : F] = \deg m_F(\gamma) < 3$ if and only if $F(\beta)/F$ is a square tower if and only if β is constructible from z_1, z_2, z_3. Hence $\beta \notin C(z_1, z_2, z_3)$ if and only if f has no root in F. We, therefore, have proven the following:

Theorem 52.13. *An angle θ can be trisected with straight-edge and compass if and only if the polynomial $f = 4t^3 - 3t - \cos\theta$ in $\mathbb{Q}(\cos\theta)$ is reducible if and only if f has a root in $\mathbb{Q}(\cos\theta)$*

Examples 52.14. 1. Let $\theta = 60^o$, so $\cos\theta = 1/2$, $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. The only possible roots of $f = 4t^3 - 3t - 1/2$ in $\mathbb{Q}$ are ± 1, $\pm 1/2$, $\pm 1/4$, $\pm 1/8$, none of which are, so f is irreducible in $\mathbb{Q}[t]$, hence an angle of 60^0 cannot be trisected with straight-edge and compass.

2. Let $\theta = 90^o$, so $\cos\theta = 0$ and $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. As the polynomial $f = 4t^3 - 3t$ is reducible in $\mathbb{Q}[t]$, the angle 90^o can be trisected with straight-edge and compass. (Of course, we already knew this).
3. Let $\theta = 180^o$, so $\cos\theta = -1$ and $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. As the polynomial $f = 4t^3 - 3t + 1$ has -1 as a root, it is reducible in $\mathbb{Q}[t]$, so the angle 180^o can be trisected with straight-edge and compass. [Of course, this also implies that an angle of 90^o can be trisected.] Note that this means that an equilateral triangle can be constructed from two given points.
4. Let $\theta = 45^o$, so $\cos\theta = 1/\sqrt{2}$. The angle θ can be constructed from $0, 1, \sqrt{2}$ if and only if $f = 4t^3 - 3t - 1/\sqrt{2} \in \mathbb{Q}(\cos 45^0)$ is not irreducible. But $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is a square root tower and an angle of 90^o can be trisected, hence so can an angle of 45^o. Therefore, f has a root in $\mathbb{Q}(\sqrt{2})$. Of course, you can use the cubic formula (if you know it) to find the root.

Doubling the cube: We are given $z_1 = 0, z_2 = 1$ and we wish to construct a line segment of length a such that $a^3 = 2$ from $z_1 = 0, z_2 = 1$. As the polynomial $m_{\mathbb{Q}}(\sqrt[3]{2}) = t^3 - 2$ in $\mathbb{Q}[t]$ is irreducible (by Eisenstein's Criterion), $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, and hence this construction cannot be done with straight-edge and compass from the given two points.

Squaring the circle: Given a unit circle, $z_1 = 0$ and $z_2 = 1$, can we construct a segment of length $\sqrt{\pi}$ from these two points? We can construct such a segment of this length from $z_1 = 0$ and $z_2 = 1$ if and only if we can construct a segment of length π from these two points and if this can be done, then $[\mathbb{Q}(\pi) : \mathbb{Q}]$ is a power of two. By Lindemann's Theorem 73.1 below, π is transcendental, so this cannot occur and it is impossible to square a circle with straight-edge and compass.

Construction of regular n-gons: We must determine when it is possible to construct a ray of angle $2\pi/n$ radians from $z_1 = 0$, $z_2 = 1$, equivalently, when it is possible to construct $z = \cos\frac{2\pi}{n} + \sqrt{-1}\sin\frac{2\pi}{n}$ from z_1, z_2. To solve this construction problem, we must use the Square Root Tower Theorem (to be proven in 57.11 below). We begin with two lemmas.

Lemma 52.15. *Any 2^n-gon, $n > 1$, is constructible from two points.*

Proof. As we can erect perpendiculars, i.e., angles $2\pi/4$ radians, we can construct a square. As we can bisect any angle, we can construct a ray of $\frac{1}{2^m}(\pi/2)$ for any m by induction. □

Lemma 52.16. *Let m and n be relatively prime integers each at least three. Then*

(1) A regular (mn)-gon is constructible from two points if and only if both a regular m-gon and a regular n-gon can be constructible from two points.

(2) a regular n-gon can be constructed from two points if and only if a regular $2n$-gon can be constructed from two points.

Proof. (2): Angles can be bisected or doubled using straight-edge and compass.
(1): Suppose that a regular (mn)-gon can be constructed, i.e., a ray of angle $\theta = 2\pi/mn$ radians. Then rays of angle $m\theta$ and $n\theta$ can be constructed. Conversely, suppose that $\theta = 2\pi/m$ radians and $\varphi = 2\pi/n$ radians are constructible. As m and n are relatively prime, we have an equation $1 = am + bn$ for some integers a and b. So

$$\frac{1}{mn} = \frac{a}{n} + \frac{b}{m} \quad \text{hence} \quad \frac{2\pi}{mn} = a\frac{2\pi}{n} + b\frac{2\pi}{m}$$

is constructible. □

We call a number $f_n = 2^{2^n} + 1$ a *Fermat number*. If it is a prime, we call it a *Fermat prime*.

Remarks 52.17. 1. $f_0 = 3$, $f_1 = 5$, $f_2 = 17$, $f_3 = 257$, and $f_4 = 65537$ are Fermat primes.

2. (Euler) $f_5 = 4294967297 = 641 \cdot 6700417$.
3. $f_6 = 274177 \cdot 67280421310722$ is composite.
4. f_n for $n = 5$—32, 36—39, 42—43, 48, 52, 55, 58, 61—66, 71—77, 81, 83, 88, 90—91, 93—94, 96, 99 are known to be composite.

At the time of this writing $f_{5523858}$ is the largest composite Fermat number known and f_4 is the largest Fermat prime known.

The natural question is whether there exist infinitely many Fermat primes, or even one f_n, with $n > 4$.

Fermat primes are applicable to solving the construction of regular n-gons, because the following theorem solves this Euclidean problem.

Theorem 52.18. *Let n be an integer at least three. Then a regular n-gon can be constructed from two points if and only if $n = 2^k p_1 \cdots p_r$ with k non-negative, and $p_1, \dots, p_r$ distinct Fermat primes for some r or $n = 2^k$ for some integer k at least two.*

Proof. By the lemmas, it suffices to determine when a regular p^r-gon can be constructed with p an odd prime, $r \geq 1$, i.e., to determine when $\zeta = \cos\frac{2\pi}{p^r} + \sqrt{-1}\sin\frac{2\pi}{p^r}$ is constructible. By Example 50.15(6), we know that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p^{r-1}(p-1)$. By the Square Root Tower Theorem, ζ is constructible if and only if $p^{r-1}(p-1) = 2^e$ for some $e \geq 0$. For this to be true, we must have $r = 1$ and $p = 2^e + 1$ a prime. If $e = ab$, for positive integers a and b with a odd, then

$$p = 2^{ab} + 1 = (2^b + 1)(2^{b(a-1)} - 2^{b(a-2)} + \cdots - 2^b + 1)$$

is a composite number unless $a = 1$, i.e., $p = 2^{2^s} + 1$ is a Fermat prime. This proves the theorem. $\square$

Examples 52.19. 1. A regular 9-gon cannot be constructed either by the n-gon theorem or by our results on trisection of an angle. Indeed, if the 9-gon could be constructed, a ray of 40^o could be constructed, hence an angle of 120^o could be trisected, so an angle of 60^o could be trisected which we know is false.

2. We show that a regular pentagon can be constructed, by producing the necessary square root tower. We must construct $z = e^{2\pi\sqrt{-1}/5}$, a root of $f = t^4 + t^3 + t^2 + t + 1$ in $\mathbb{Q}[t]$. We have $z^{-1} = z^4 = e^{8\pi\sqrt{-1}/5} = e^{-2\pi\sqrt{-1}/5} = \overline{z}$ and $z^4 + z^3 + z^2 + z + 1 = 0$. So z, $\overline{z}$ z^2, $\overline{z}^2$ are the roots of f in $\mathbb{C}$. Let

$$x_1 = z + \overline{z} \text{ and } x_2 = z^2 + \overline{z}^2.$$

[Note that $\dfrac{x_1}{2} = \dfrac{z+\overline{z}}{2} = \cos(2\pi/5)$ and $\dfrac{z-\overline{z}}{2} = \sin(2\pi/5)$.]
Then

$$\begin{aligned} x_1 + x_2 &= -1 \\ x_1x_2 &= (z+\overline{z})(z^2+\overline{z}^2) = -1, \end{aligned}$$

so $(t - x_1)(t - x_2) = t^2 + t - 1$. This is not quite what we want, so we complete the square. Let $y_1 = 2x_1 + 1$ and $y_2 = 2x_2 + 1$, then

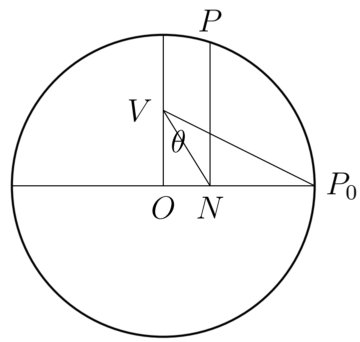
$$y_1 + y_2 = 0 \text{ and } y_1y_2 = 5.$$

Hence $t^2 - 5 = (t - y_1)(t - y_2)$ with $t^2 - 5$ irreducible in $\mathbb{Q}[t]$. We may assume that $y_1 = \sqrt{5}$. We have $x_1, x_2 \in \mathbb{Q}(y_1, y_2) = \mathbb{Q}(y_1) = \mathbb{Q}(\sqrt{5})$ and

$$\begin{aligned} x_1^2 &= (z+\overline{z})^2 = z^2 + 2z\overline{z} + \overline{z}^2 = x_2 + 2 \\ (z-\overline{z})^2 &= z^2 - 2z\overline{z} + \overline{z}^2 = x_2 - 2. \end{aligned}$$

Let $K = \mathbb{Q}(y_1, z - \overline{z}) = \mathbb{Q}(y_1, z)$. As $(z - \overline{z})^2 \in \mathbb{Q}(y_1) = \mathbb{Q}(\sqrt{5})$, the field extension $K/\mathbb{Q}$ is a square root tower. As $z = (z - \overline{z} + x_1)$ lies in K, a regular pentagon can be constructed. [Note that by the quadratic formula, either x_1 or x_2 is equal to $(\sqrt{5}-1)/2$ say x_1. Then we already know that K/F is a square root tower and $x_1/2 = \cos(2\pi/5)$.] To actually do the construction in the complex plane, draw a circle at the origin of radius two. Let O be the origin, $V = \sqrt{-1}$, and $P_0 = 2$ in the plane. Let N be the intersection point of the X-axis and the bisector of the angle $\theta = OVP_0$ and P be the intersection point of the circle (in upper half plane) and the vertical constructed at N. Then the point P is $2z = 2\cos(2\pi/5) + 2\sqrt{-1}\sin(2\pi/5)$. (Cf. Figure 52.20.)

(52.20)



In the figure $\tan\theta = 1/2$, so the length of the line segment $\overline{VP_0}$ is $\sqrt{5}$.

Gauss proved that one could construct the regular 17-gon in 1796 at the age of 19. This was the first important euclidean construction since ancient times. He showed

$$16\cos\frac{2\pi}{17} = -1 + \sqrt{17} + \sqrt{34 - 2\sqrt{17}} + 2\sqrt{17 + 3\sqrt{17} - \sqrt{34 - 2\sqrt{17}} - 2\sqrt{34 + 2\sqrt{17}}}.$$

An actual construction was shown a few years later by Johannes Erchinger. Gauss also proved that if a regular n-gon was constructible when n was of the form in the theorem in 1801. He also stated that this was a necessary condition, but his proof (if he had one) has never been found. Indeed it is doubtful that Gauss actually had a valid proof, since he made this statement prior to the birth of Galois and Galois theory is needed to prove necessity. The first published proof of necessity was given by Pierre Wantzel, who also proved that you could not trisect and angle or double a cube in 1837. Few have heard of Wantzel, so the truism that solving well-known problems will bring you lasting fame is not so well-founded. A regular 257-gon was constructed by Friedrich Richelot (and Schwendenwein) in 1832. The regular 65537-gon was constructed by Johann Hermes in 1895. Apparently Hermes' actual construction rests in a trunk in the attic at the Mathematics Institute at Göttingen (presumably studied by no one). The squaring of the circle was the last of the problems solved when Lindemann proved the transcendence of π in 1882.

Exercises 52.21.

1. Prove that the equations in the last paragraph in the proof of Theorem 52.5 have a solution in $C' \cap \mathbb{R}$.
2. Show the field of complex constructible numbers over $\mathbb{Q}$ is closed under quadratic extensions, but has extensions of every odd prime degree. Does it have extensions of any odd degree? of nonzero even degree? of any degree but two?
3. Show that $\mathbb{Z}/2\mathbb{Z}$ has no square root towers over if of degree > 1.
4. Show the construction of the regular 5-gon in Figure 52.20 works.
5. Show an angle of $2\pi/n$ radians, n an integer, can be trisected by straight-edge and compass if and only if n is not divisible by 3.

53. Separable Elements

A real polynomial that has multiple roots can cause problems as that means that such a root is also a root of its derivative, hence its graph is more complicated. If the polynomial is irreducible, this cannot happen. However, if the irreducible polynomial has coefficients in a field of positive characteristic, this can occur. In this section, we study this phenomena, formally defining some concepts that we have already used and leaving many of the details as exercises (although some will be proven later). We recall some definitions.

Definition 53.1. Let K/F be an extension of fields, f a non-constant polynomial in $F[t]$ with a root α in K. We say that α is a *multiple root* of g of *multiplicity* r in K if $f = (t-\alpha)^r g$ for some g in $K[t]$ with α not a root of g. If α is a root but not a multiple root, we call it a *simple* root of f. If $f = \sum_{i=0}^n a_i t^i$ in $F[t]$, the *derivative* of f is defined to be the polynomial $f' = \sum_{i=1}^n i a_i t^{i-1}$ in $F[t]$.

We have, that we (again) leave as an exercise:

Properties 53.2. Let f and g be polynomials in $F[t]$. Then

(1) $(f+g)' = f' + g'$.
(2) $(af)' = af'$ for all $a \in F$.
(3) $(fg)' = f'g + fg'$.

As mentioned above, multiple roots are also roots of the derivative. We formally write this as a lemma. This lemma and its proof also says that the notion of multiple root and its muliplicity is independent of the field extension of the base field.

Lemma 53.3. *Let K/F be an extension of fields, f a non-constant polynomial with a root α in K. Then α is a simple root of f if and only if α is not a root of f'.*

Proof. We can write $f = (t-\alpha)^r g$ for some g in $K[t]$ with $g(\alpha) \neq 0$ for some integer r. Therefore, $f' = r(t-\alpha)^{r-1} g + (t-\alpha)^r g'$ in $K[t]$. As $g(\alpha) \neq 0$ in the domain K, we have $t - \alpha \mid f'$ in $K[t]$ if and only if $t-\alpha \mid r(t-\alpha)^{r-1} g$, i.e., $r-1 \geq 1$ or $r = 0$ in K. If $r = 0$ in K, then $\operatorname{char} F \mid r$, so $r \geq \operatorname{char} F \geq 2$. The result follows. $\square$

Corollary 53.4. *Let f be a non-constant polynomial in $F[t]$. Then f has a multiple root in some extension field K of F if and only if the ideal (f, f') in $F[t]$ is not the unit ideal.*

Proof. ($\Rightarrow$): Let α be a multiple root of f in an extension field K of F. Then $m_F(\alpha) \mid f$ and $m_F(\alpha) \mid f'$ in $F[t]$, so $(f, f') \subset \big(m_F(\alpha)\big) < F[t]$.
($\Leftarrow$): $F[t]$ is a PID, so $(f, f') = (g) < F[t]$ for some non-constant polynomial g in $F[t]$. Let h be any non-constant polynomial in $F[t]$ satisfying $h \mid g$. Let K/F be a field extension such that h has a root α in K. Then $f(\alpha) = f'(\alpha) = h(\alpha) = 0$, so f has a multiple root in K. $\square$

The corolllary says that one can determine when a polynomial has a multiple root in some extension is intrinsic to the base field. Moreover, since $F[t]$ is a euclidean domain, this can be determined quite easily. It also answers the question of multiple roots for irreducible polynomials.

Corollary 53.5. *Let f be an irreducible polynomial in $F[t]$.*

(1) If $\operatorname{char} F = 0$, then f has no multiple roots in any field extension of F. In particular, f can have only simple roots, if any, in a field extension of F.

(2) If $\operatorname{char} F = p > 0$ and F has a multiple root in some field extension K of F, then there exists a polynomial g in $F[t]$ satisfying $f = g(t^p)$, i.e., $f = \sum a_i t^{pi}$ in $F[t]$.

Proof. We begin the proof, leaving the rest of the proof as exercises. Let K/F be a field extension such that f has a root α in K, so $f = am_F(\alpha)$ for some nonzero a in F. In particular, if α is a multiple root, we must have $f \mid f'$ in $F[t]$. So $f' = 0$ or $\deg f' \geq \deg f$. Consequently, $f' = 0$. The result now follows by Exercises 53.10(2a) and(2b). □

Definition 53.6. An irreducible polynomial f in $F[t]$ is called *separable over* F if it has no multiple roots in any field extension of F. A non-constant polynomial is called *separable over* F if all of its irreducible factors are separable. If f is not separable over F, it is called *inseparable*. Let K/F be a field extension and α an element in K. We say that α is *separable* over F if it is the root of a separable polynomial defined over F and the extension K/F is *separable* if every element of K is separable over F.

Remark 53.7. In reading other books, it is much more common to say that a non-constant polynomial f in $F[t]$ is separable if it has no multiple roots in any extension field of F. It is, however, convenient to use the above definition for a separable polynomial in the sequel.

Remarks 53.8. Let K/F be an extension of fields and f, g non-constant polynomials in $F[t]$.

1. Every element of F is separable over F.
2. An element α in K is separable over F if and only if α is algebraic over F and $m_F(\alpha)$ is separable over F. In particular, if K/F is separable, then it is algebraic.
3. If $\operatorname{char} F = 0$ and K/F is algebraic, then K/F is separable. A field F in which every element algebraic over F is separable is called a *perfect field*. Therefore, any field of characteristic zero is perfect. It turns out that every finite field is also perfect, but there exist fields that are not perfect.
4. The polynomial f is separable over F if and only if every divisor of f is separable over F.
5. If both f and g are separable over F then so is fg.
6. If $L/K/F$ are field extensions with L/F separable, then L/K and K/F are separable.

Theorem 53.9. *Let $K/E/F$ be extensions of field. Then K/F is separable if and only if K/E and E/F are both separable.*

We shall not prove this theorem here. We know that if K/F is separable, then K/E and E/F are separable. The converse is not so easy. To reduce the problem, we can try a trick that we did before. If α in K is separable over E, then $m_E(\alpha)$ in $E[t]$ is separable. If $m_E(\alpha) = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $E[t]$, then $m_E(\alpha)$ is separable over $E_0 = F(a_0, \ldots, a_{n-1})$, with a_i, $i = 0, \ldots, n-1$ separable over F. So we are reduced to showing if $\beta_1, \ldots, \beta_n$ are separable over F, then $F(\beta_1, \ldots, \beta_n)/F$ is separable and if, in addition, α is separable over $F(\beta_1, \ldots, \beta_n)$, then α is separable over F. This is still not easy. One way of doing this is given in the exercises. Of course, we may assume that $\operatorname{char} F$ is positive.

Exercises 53.10.

1. Suppose that f be a non-constant polynomial in $F[t]$, $L/K_i/F$ field extensions with $i = 1, 2$, and α an element of $K_1 \cap K_2$. Show that α is a root of f of multiplicity r in K_1 if and only if α is a root of f of multiplicity r in K_2.
2. Let F be a field and f a polynomial in $F[t]$. Show the following:
   (a) If $\operatorname{char} F = 0$, then $f' = 0$ if and only if f is a constant polynomial.
   (b) If $\operatorname{char} F = p$ is positive, then $f(t)^p = f^p(t^p)$ (The Children's Binomial Theorem) and if $f' = 0$, then there exists a polynomial g in $F[t]$ satisfying $f(t) = g(t^p)$.
3. If x is transcendental over a field F, then $t^p - x \in F(x)[t]$ is irreducible for any prime p.
4. Suppose that F is a field of positive characteristic p. Show the following:
   (a) The map $F \to F$ given by $x \mapsto x^p$ is a monomorphism. Denote its image by F^p.
   (b) If K/F is algebraic and $\alpha \in K$ is separable over $F(\alpha^p)$, then $\alpha \in F(\alpha^p)$.
   (c) Every finite field is *perfect*, i.e., every algebraic extension is separable.
5. Suppose that F is a field of positive characteristic p. Show all of the following:
   (a) If K/F is separable field extension, then $K = F(K^p)$, where $K^p := \{x^p \mid x \in K\}$.
   (b) Suppose that K/F is finite and $K = F(K^p)$. If $\{x_1, ..., x_n\} \subset K$ is linearly independent over F then so is $\{x_1^p, ..., x_n^p\}$.
   (c) If K/F is finite and $K = F(K^p)$, then K/F is separable.
6. Let K/F be an extension of fields. Show all of the following:
   (a) If $\alpha \in K$ is separable over F then $F(\alpha)/F$ is separable.
   (b) If $\alpha_1, ..., \alpha_n \in K$ are separable over F, then $F(\alpha_1, ..., \alpha_n)/F$ is separable.
   (c) Let $F_{\text{sep}} = \{\alpha \in K \mid \alpha \text{ separable over } F\}$. Then F_{sep} is a field.
7. Let $L/K/F$ be field extensions. Show that L/F is separable if and only if L/K and K/F are separable using the previous two exercises.
8. Show that any algebraic extension of a perfect field is perfect.
9. Let F_0 be a field of positive characteristic p. Let $F = F_0(t_1^p, t_2^p)$ and $L = F_0(t_1, t_2)$. Show all of the following:
   (a) If $\theta \in L \setminus F$ then $[F(\theta) : F] = p$.
   (b) There exist infinitely many fields K satisfying $F < K < L$.
10. Let F be a field of positive characteristic p. Show
    (a) If $F = F^p$, then F is perfect.
    (b) If F is not perfect, then $F \neq F^{p^r} := \{x^{p^r} \mid x \in F\}$ for any $r \geq 1$.
11. Let K/F be a field extension and $\alpha \in K$. We say that α is *purely inseparable* over F if $m_F(\alpha)$ has only one root in a splitting field of α and K/F is *purely inseparable* if every element of K is purely inseparable over F. Show that α in K is pure inseparable and separable over K if and only if α lies in F.
12. Let K/F be a field extension with $\operatorname{char}(F) = p > 0$. Show that if α in K is algebraic then there exists a nonnegative integer n satisfying α^{p^n} is separable over F.

13. Let K/F be an algebraic extension with $\text{char}(F) = p > 0$. Show the following are equivalent:
    (a) K/F is purely inseparable.
    (b) If α in K, then there exists a nonnegative integer n such that $m_F(\alpha) = t^{p^n} - a \in F[t]$.
    (c) If α lies in K, then there exists a nonnegative integer n such that α^{p^n} lies in F.
    (d) $K_{sep} = F$. (Cf. Exercise 6 above.)
    (e) K is generated by purely inseparable elements over F.
14. Let K/F be an algebraic field extension. Show that K/F_{sep} is purely inseparable and $K_{psep}; = \{\alpha \in K | \alpha \text{ is purely inseparable over } F\}$ is a field.

CHAPTER XII

Galois Theory

In the previous chapter, we studied field theory, i.e., fields and field extensions. In the chapter, we interrelate field theory with group theory. The goal will be to interwine the intermediate fields of finite field extension K/F when K is the splitting field of a separable polynomial in $F[t]$ over F and subgroups of the Galois group $\mathrm{Aut}_F(K)$. This will allow us to prove some fundamental results. For example, we shall give an algebraic proof of the Fundamental Theorem of Algebra using only the Intermediate Value Theorem from analysis, prove that there is no formula involving only addition, multiplication, and the extraction of nth roots for various n for the fifth degree polynomial over the rational numbers using the results that were established about solvable groups, and finish the proof for the existence of the regular n-gon for the allowable n. We shall show how the study of roots of unity leads to Quadratic Reciprocity and a special case of Dirichlet's Theorem of Primes in an Arithmetic Progression.

54. Characters

This section is the key to the interaction of field theory and group theory. We follow the approach of E. Artin, who proved the main results using systems of linear equations. The idea is to look at a set S of field homomorphisms from a field K and to determine the subfield of K on which each homomorphism in S acts in the same way. We begin with the following definition:

Definition 54.1. Let F be a field and G be a (multiplicative) group. Then a group homomorphism $\sigma : G \to F^\times$ is called an (*linear*) *character*. [Note that $F^\times = \mathrm{GL}_1(F)$.] We say that distinct characters $\sigma_1, \dots, \sigma_n : G \to F^\times$ are *dependent* (or the set $\{\sigma_1, \dots, \sigma_n\}$ is *dependent*) if there exist $a_1, \dots, a_n$ in F, not all zero, satisfying $\sum_i a_i\sigma_i(g) = 0$ for all g in G, i.e., the function $\sum a_i\sigma_i : G \to F$ is the zero map; and *independent* otherwise.

Remarks 54.2. 1. A field homomorphism $\sigma : E \to F$ induces an character by restriction, as a ring homomorphism takes units to units. As field homomorphisms are monic, such a character is always injective.

2. The set
$$\{\sigma \mid \sigma : G \to F^\times \text{ is a character.}\}$$
is not a group under $+$, but is a group under multiplication of functions, i.e., $\sigma\tau(g) := \sigma(g)\tau(g)$ for all g in G, with the identity the map $1 : G \to F^\times$ given by $g \mapsto 1_F$.

The key result is:

Lemma 54.3. (Dedekind's Lemma) *Let $\sigma_1, \dots, \sigma_n : G \to F^\times$ be distinct characters. Then $\sigma_1, \dots, \sigma_n : G \to F^\times$ are independent.*

PROOF. We prove this by induction on n.

$n = 1$: If $a\sigma_1(g) = 0$ for all g in G, then $a = 0$ as $\sigma_1(g) \neq 0$ for all g in G.

$n > 1$: Suppose that we have an equation

$$(*) \qquad \sum_{i=1}^{n} a_i\sigma_i(g) = 0 \text{ for all } g \text{ in } G \text{ with } a_i \text{ in } F \text{ not all zero.}$$

By induction any proper subset of $\{\sigma_1, \ldots, \sigma_n\}$ is independent, so a_i is nonzero for every i. Multiplying the equation in (*) by a_n^{-1}, we may assume that that $a_n = 1$. So (*) now reads

$$(**) \qquad \sum_{i=1}^{n-1} a_i\sigma_i(g) + \sigma_n(g) = 0 \text{ for all } g \text{ in } G.$$

By assumption, $\sigma_1 \neq \sigma_n$, so there exists an element x in G satisfying $\sigma_1(x) \neq \sigma_n(x)$. As G is a group, $xG = G$; and as the σ_i are group homomorphisms, we have

$$0 = \sum_{i=1}^{n-1} a_i\sigma_i(xg) + \sigma_n(xg) = \sum_{i=1}^{n-1} a_i\sigma_i(x)\sigma_i(g) + \sigma_n(x)\sigma_n(g)$$

for all g in G. Since $\sigma_n(x) \neq 0$, we see that

$$(\dagger) \qquad 0 = \sum_{i=1}^{n-1} \sigma_n(x)^{-1}a_i\sigma_i(x)\sigma_i(g) + \sigma_n(g) \text{ for all } g \text{ in } G.$$

Subtracting the equation in (†) from the equation in (**) yields

$$0 = \sum_{i=1}^{n-1} [a_i - \sigma_n(x)^{-1}a_i\sigma_i(x)]\sigma_i(g) \text{ for all } g \text{ in } G.$$

By induction, we conclude that

$$a_i = \sigma_n(x)^{-1}a_i\sigma_i(x) \text{ for all } i$$

in the domain F with $a_i \neq 0$, hence $\sigma_n(x) = \sigma_i(x)$ for all i, contradicting $\sigma_1(x) \neq \sigma_n(x)$. □

The proof above also shows, with the obvious definitions, that any set of characters $S = \{\sigma_i : G \to F^\times \mid \sigma_i \text{ a character for all } i \in I\}$ is independent.

Corollary 54.4. *Let $\sigma_1, \ldots, \sigma_n : F \to K$ be distinct field homomorphisms. Then $\sigma_1, \ldots, \sigma_n : F^\times \to K^\times$ are independent characters.*

Definition 54.5. Let S be a nonempty subset of the full set of all field homomorphisms $F \to K$, $\{\sigma : F \to K \mid \sigma \text{ a field homomorphism}\}$. We call an element a in F a *fixed point under S*, if $\sigma(a) = \tau(a)$ for all σ and τ in S.

Example 54.6. Let S be a nonempty set of field homomorphisms $F \to K$. Assume that both F and K contain the same prime subfield Δ (so $\Delta \cong \mathbb{Q}$ if the characteristic of F is zero and $\Delta \cong \mathbb{Z}/p\mathbb{Z}$, if the characteristic of F is the prime p.) As every homomorphism takes $1_F \to 1_K$ and $1_\Delta = 1_F = 1_K$, we have $\sigma(x) = x$ for all $x \in \Delta$, Therefore, the term 'fixed point' is actually appropriate for elements in Δ.

Note that if S is a set of field homomorphisms $\sigma : K \to K$ containing 1_K, then the fixes points of this set are all really fixed points, i.e., $\sigma(x) = x$ for all $\sigma \in S$.

As should be expected, the set of fixed points has additional structure. We immediately see that we have:

Lemma 54.7. *Suppose that S is a nonempty set of field homomorphisms $F \to K$. Let $E = \{x \in F \mid x \text{ is a fixed point of } S\}$. Then E is a subfield of F.*

If S is a nonempty set of field homomorphisms $F \to K$, we denote the field $\{x \in F \mid x \text{ is a fixed point of } S\}$ by F^S and call it the *fixed field* of S. We are interested in the fixed field of a finite set of field homomorphisms. Following Artin, we use linear algebra to obtain nontrivial solutions to an appropriate set of linear equations. The first main result about such is the following which gives a lower bound on a finite set of field homomorphism. This bound will be often used in the sequel.

Lemma 54.8. (Artin's Lemma) *Suppose that S is a finite nonempty set of field homomorphisms $F \to K$. Then $[F : F^S] \geq |S|$.*

Proof. Let $S = \{\sigma_1, \dots, \sigma_n\}$ with $n \geq 1$. Suppose that $r = [F : F^S] < |S| = n$ and $\{\omega_1, \dots, \omega_r\}$ is an F^S-basis for F. Consider the following system of equations over K:

$$(*) \qquad \begin{array}{ccccccc} \sigma_1(\omega_1)x_1 & + & \cdots & + & \sigma_n(\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_1(\omega_r)x_1 & + & \cdots & + & \sigma_n(\omega_r)x_n & = & 0. \end{array}$$

As $r < n$, there exists a nontrivial solution to the system of equations (*) over K, say

$$x_1, \dots, x_n \text{ with each } x_i \text{ in } K \text{ and not all zero}$$

is such a solution. Let a be an element of F. We can write $a = \sum a_i\omega_i$ with each $a_i \in F^S$. For each i, multiply the ith equation in (*) by $\sigma_i(a_i)$ to get a new equivalent system of equations

$$\begin{array}{ccccccc} \sigma_1(a_1)\sigma_1(\omega_1)x_1 & + & \cdots & + & \sigma_1(a_1)\sigma_n(\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_n(a_r)\sigma_1(\omega_r)x_1 & + & \cdots & + & \sigma_n(a_r)\sigma_n(\omega_r)x_n & = & 0. \end{array}$$

Since $\sigma_i(a_i) = \sigma_j(a_i)$ for all i and j and the σ_i are homomorphisms, we have

$$(\dagger) \qquad \begin{array}{ccccccc} \sigma_1(a_1\omega_1)x_1 & + & \cdots & + & \sigma_n(a_1\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_1(a_r\omega_r)x_1 & + & \cdots & + & \sigma_n(a_r\omega_r)x_n & = & 0. \end{array}$$

Adding all the equations in (†) yields

$$0 = \sum_{i=1}^{r}\sum_{j=1}^{n} \sigma_j(a_iw_i)x_j = \sum_{j=1}^{n}\sum_{i=1}^{r} \sigma_j(a_iw_i)x_j = \sum_{j=1}^{n} \sigma_j(a)x_j$$

for all a in F with x_j not all zero. This means that the characters $\sigma_1, \dots, \sigma_n$ are dependent, contradicting Dedekind's Lemma. □

A special case of Artin's Lemma, is the one in which we shall be interested.

Corollary 54.9. *Let S be a nonempty finite set of field automorphisms of a field K. Then $[K : K^S] \geq |S|$.*

Of course, we should be very interested in the case when we have equality in Artin's Lemma. In general this is not true. We can even have a finite set of field automorphisms S of a field K with $[K : K^S]$ infinite [can you give an example?]. What we need is to look at those S with additional structure. For example, let K/F be an extension of fields. Recall the *Galois group* of K/F is defined to be the set

$$\begin{aligned} G(K/F) &= \mathrm{Aut}_F(K) \\ &:= \{\sigma : K \to K \mid \text{an } F\text{-automorphism satisfying } \sigma|_F = 1_F\}. \end{aligned}$$

It is immediate that the following is true.

Lemma 54.10. *Let K/F be an extension of fields. Then $G(K/F)$ is a subgroup of the (field) automorphism group of K and F is a subfield of $K^{G(K/F)}$.*

As a consequence, we have:

Corollary 54.11. *Let K/F be a finite extension of fields. Then the Galois group $G(K/F)$ is a finite group and satisfies $[K : F] \geq |G(K/F)|$.*

PROOF. As F is a subfield of K^S for any nonempty subset S of $G(K/F)$, we have

$$[K : F] \geq [K : K^S] \geq |S| \text{ for any finite } S,$$

i.e., $[K : F]$ is a uniform upper bound for all such S. It follows that $G(K/F)$ is finite and satisfies $[K : F] \geq |G(K/F)|$. □

We arrive at the group theoretic condition in which we are interested.

Definition 54.12. Let K/F be a finite extension of fields. We say that the extension is *Galois* if F is the fixed field of $G(K/F)$, i.e., $F = K^{G(K/F)}$.

We have defined a field extension K/F to be *Galois* if K is a finite extension with F the fixed field of $G(K/F)$. It would be more appropriate to call this a *finite Galois* extension and call K/F a *Galois extension* if it is algebraic with F the fixed field of $G(K/F)$.

Remarks 54.13. Let K/F be a field extension.

1. If $F = K^{G(K/F)}$, then for any element $x \in K \setminus F$, there exists an automorphism in $G(K/F)$ that *moves* x, i.e., there exists a σ in $G(K/F)$ satisfying $\sigma(x) \neq x$.
2. If $K/E/F$ is an intermediate field, then $G(K/E) \subset G(K/F)$.
3. Let $E = K^{G(K/F)}$, then by definition of $K/E/F$, we have $G(K/E) \supset G(K/F)$, so by (2), we have
$$G(K/F) = G(K/K^{G(K/F)}).$$
4. If K/F is finite, then a field homomorphism $\sigma : K \to K$ fixing F necessarily is an element of $G(K/F)$ as σ is monic and F-linear.

We give examples of Galois groups.

Examples 54.14. 1. Let $K = F$ be a field. Then $G(K/F) = \{1\}$ and K/F is Galois with $[K : F] = 1 = |G(K/F)|$. Moreover, K is the splitting field of $t - a$ in $F[t]$ over F for any $a \in F$.

2. Let $F = \mathbb{R}$ and $K = \mathbb{C}$. Then $G(K/F) = \{1, \bar{\ }\}$ (with $1 = 1_K$ and $\bar{\ }$ complex conjugation) and K/F is Galois with $[K : F] = 2 = |G(K/F)|$. Moreover, K is the splitting field of $t^2 + 1$ over F.

3. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt{2})$. Let $\sigma : K \to K$ be defined by $a+b\sqrt{2} \mapsto a-b\sqrt{2}$ for all a and b in $\mathbb{Q}$. Then $G(K/F) = \{1, \sigma\}$ (with $1 = 1_K$) and K/F is Galois with $[K : F] = 2 = |G(K/F)|$. Moreover, K is the splitting field of $t^2 - 2$ over F.

4. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt[3]{2})$. Then $G(K/F) = \{1\}$, so K/F is not Galois. It satisfies $[K : F] = 3 > |G(K/F)|$. Moreover, K is not a splitting field of any polynomial over F:

   We indicate why these facts are true. Any field automorphism of K must take $\sqrt[3]{2}$ to another root of $t^3 - 2$, but $\mathbb{R}$ contains only one root of $t^3 - 2$. If K was the splitting field of some polynomial g in $F[t]$ over F, then it would be the splitting field of an irreducible polynomial of degree three. (Why?) Such a g would have three distinct roots, say $\alpha_1, \alpha_2, \alpha_3$ in K and we would have $K = \mathbb{Q}(\alpha_i)$, $i = 1, 2$, or 3. But then there must be an F-automorphism σ satisfying $\sigma(\alpha_1) = \alpha_2$.

5. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{R}$. Then $G(K/F) = \{1\}$ and $K^{G(K/F)} = K$ with K/F is infinite and not even algebraic. (Cf. Exercise 50.18(11).) However, (by Exercise 56.22(7)) the Galois group $G(\mathbb{C}/\mathbb{Q})$ is uncountable!

6. Suppose that the characteristic of a field F is $p > 0$ and $K = F(\alpha)$ for some α satisfying $\alpha^p = a$ with $a \in F$. If $\alpha \notin F$, then $G(K/F) = \{1\}$ and K/F is not Galois. It satisfies $[K : F] = p > |G(K/F)|$ and K is the splitting field of $t^p - a$ over F.

Artin's Lemma becomes much stronger, when the nonempty finite set of field automorphisms is a group. This results in the fundamental link between group theory and field theory on the group theory side. It is the following:

Theorem 54.15. (Artin's Theorem) *Let K be a field and G a finite subgroup of field automorphisms of K. Then*

(1) $[K : K^G] = |G|$.
(2) $G = G(K/K^G)$.
(3) K/K^G is Galois.

Proof. (1): Let $n = |G|$ and $G = \{\sigma_1, \ldots, \sigma_n\}$. By Artin's Lemma, we know that $[K : K^G] \geq |G| = n$. Suppose that $[K : K^G] > n$. Choose $a_1, \ldots, a_{n+1}$ in $K^\times$ that are K^G-linearly independent. The system of n linear equations in $(n+1)$-unknowns

$$(*)\qquad \begin{array}{ccccccc} x_1\sigma_1(a_1) & + & \cdots & + & x_{n+1}\sigma_1(a_{n+1}) & = & 0 \\ \vdots & & & & \vdots & & \\ x_1\sigma_n(a_1) & + & \cdots & + & x_{n+1}\sigma_n(a_{n+1}) & = & 0 \end{array}$$

has a nontrivial solution over K. Among all the nontrivial solutions $x_1, \ldots, x_{n+1}$ over K, choose one with the least number of x_i nonzero. Relabeling, we may assume this solution is

$$x_1, \dots, x_r, 0, \dots, 0 \text{ with } x_i \in K^\times,\ 1 \le i \le r$$

with r minimimal. Multiplying this solution by x_r^{-1}, we may also assume that $x_r = 1$. As G is a group, we may assume also that $\sigma_1 = 1_K$, so

$$x_1a_1 + \dots + x_{r-1}a_{r-1} + a_r = 0, \quad x_i \in K^\times,\ r > 1 \text{ (as } a_1 \neq 0).$$

Since $\{a_1, \dots, a_r, a_{r+1}, \dots, a_{n+1}\}$ is K^G-linearly independent, this means that for some integer i, $1 \le i \le r$, we have $x_i \notin K^G$. By definition, there exists an integer k such that $\sigma_k(x_i) \neq x_i$. (So $k > 1$.) The system of equations in (*) has the form

$$(**) \qquad 0 = \sum_{i=1}^{r-1} x_i\sigma_j(a_i) + \sigma_j(a_r) = 0 \text{ for } j = 1, \dots, n.$$

Taking σ_k of each of the equations in (**) yields

$$0 = \sum_{i=1}^{r-1} \sigma_k(x_i)\sigma_k\sigma_j(a_i) + \sigma_k\sigma_j(a_r) = 0 \text{ for } j = 1, \dots, n.$$

As G is a group, $\sigma_k G = G$, hence these equations are the same as

$$(\dagger) \qquad 0 = \sum_{i=1}^{r-1} \sigma_k(x_i)\sigma_j(a_i) + \sigma_j(a_r) = 0 \text{ for } j = 1, \dots, n.$$

Subtracting the appropriate equations in (†) from those in (*) yields

$$0 = \sum_{i=1}^{r-1} [x_i - \sigma_k(x_i)]\sigma_j(a_i) = 0 \text{ for } j = 1, \dots, n.$$

Since $\sigma_k(x_i) \neq x_i$, we have

$$x_1 - \sigma_k(x_1), \dots, x_{r-1} - \sigma_k(x_{r-1}), 0, \dots, 0$$

is a nontrivial solution over K to the system of the equations in (*). This contradicts the minimality of r and establishes (1).

(2): By (1), we have $[K : K^G] = |G|$ is finite. By Corollary 54.9, we have $[K : K^G] \ge |G(K/K^G)|$, so $G(K/F)$ is also finite. Since G is a subgroup of $G(K/K^G)$ by definition, we have $|G| = |G(K/K^G)|$ and is finite, so $G = G(K/K^G)$.

(3) is now immediate. □

We want to interrelate Galois groups and field extensions. An application of Artin's Theorem produces one such result.

Corollary 54.16. *Let K be a field, G_1, G_2 two finite subgroups of $\mathrm{Aut}(K)$, the group of field automorphisms of K, and $F_i = K^{G_i}$ for $i = 1, 2$. Then $F_1 = F_2$ if and only if $G_1 = G_2$.*

Proof. (⇐) is clear.

(⇒): By Artin's Theorem, $G_i = G(K/F_i)$, so this also follows. □

We now see that for finite field extensions, being Galois only depends on a simple cardinality test.

Corollary 54.17. *Suppose that K/F is a finite extension of fields. Then K/F is Galois if and only if $[K : F] = |G(K/F)|$.*

PROOF. ($\Rightarrow$) follows from Artin's Theorem and the special case of Artin's Lemma, Corollary 54.9.

($\Leftarrow$): Let $E = K^{G(K/F)}$. Then we know that $G(K/E) = G(K/F)$. As K/E is Galois, by the proven sufficiency, we have $[K : E] = |G(K/E)|$. Therefore,

$$[K : E] = |G(K/E)| = |G(K/F)| = [K : E][E : F].$$

It follows that $[E : F] = 1$, i.e., $E = F$. $\square$

Exercise 54.18. This exercise computes the Galois group of a non algebraic extension. Using Exercise 48.25(10) show that $G(F(t)/F)$ consists of all F-automorphisms of $F(t)$ mapping t to $(at + b)/(ct + d)$ where $a, b, c, d \in F$ satisfies $ad - bc \neq 0$

55. Computations

Of course, given a finite extension K/F of fields, one would like to determine its Galois group. In general, as expected, this is very difficult. In this section, we do a few computations as well as giving some ideas on how to do such computations. To begin, we discuss the problem by reviewing (and repeating) some of the material that we have already discussed.

Summary 55.1. Let K/F be a finite extension of fields and f a non-constant polynomial in $F[t]$.

1. We know if $K/E/F$ is an intermediate field that
   (a) $G(K/E) \subset G(K/F)$.
   (b) $G(K/F) = G(K/K^{G(K/F)})$.
   (c) $[K : F] \geq |G(K/F)|$ with equality if and only if K/F is Galois, i.e., $F = K^{G(K/F)}$.
2. Let α in K be a root of f and σ an element in $G(K/F)$. Then $\sigma(\alpha)$ is also a root of f. In particular, if $f = g_1 \cdots g_r$ in $F[t]$ with g_i non-constant polynomials in $F[t]$, then $G(K/F)$ takes roots of g_i in K to roots of g_i for all i:

   If $f = \sum a_i t^i$, then
   $$0 = \sigma\big(f(\alpha)\big) = \sigma(\sum a_i \alpha^i) = \sum \sigma(a_i)\sigma(\alpha^i) = \sum a_i \sigma(\alpha)^i = f\big(\sigma(\alpha)\big).$$
3. If K is a splitting field of f and σ an element in $G(K/F)$, then σ is completely determined by what it does to the roots of f:

   If $\alpha_1, \dots, \alpha_r$ are the roots of f, then $K = F(\alpha_1, \dots, \alpha_n)$.
4. Let $S = \{\alpha \mid \ \alpha \text{ is a root of } f\}$. If S is not empty, then the map $G \to \Sigma(S)$ given by $\sigma \mapsto \sigma|_S$ is a group homomorphism. If K is a splitting field of f over F, then this map is monic and we can view $G \subset \Sigma(S)$. [This was Galois' viewpoint.]
5. If α is a root of f in F then $\sigma(\alpha) = \alpha$ for all $\sigma \in G(K/F)$.

6. This is the most useful remark. If K is the splitting field of f over F and g is an irreducible polynomial in $F[t]$ with roots α and β in K, then there exists an F-isomorphism
$$\tau : F(\alpha) \to F(\beta) \text{ satisfying } \alpha \mapsto \beta$$
and this extends to an F-automorphism τ in $G(K/F)$, so $\tau(\alpha) = \beta$. [Make sure that you can prove this.] In particular,
$$G(K/F) \text{ acts transitively on the roots of } g \text{ in } K.$$
We shall later see that, in fact,
$$\begin{gathered}\text{if } g \text{ is irreducible in } F[t] \text{ and}\\ \text{has a root in a splitting field } K \text{ of } f\text{, then } g \text{ splits over } K.\end{gathered}$$
Can you prove this? [You have all the tools to do so.]

To do some of our computations, we shall need a few lemmas.

Lemma 55.2. *Let G be a finite cyclic group of order n and $\mathrm{Aut}(G)$ the automorphism group of G. Then $\mathrm{Aut}(G) \cong (\mathbb{Z}/n\mathbb{Z})^\times$. In particular, $\mathrm{Aut}(G)$ is abelian (and cyclic if n is a prime) of order $\varphi(n)$.*

Proof. Let $G = \langle a\rangle$ and $\sigma \in \mathrm{Aut}(G)$. Then $\sigma(a) = a^i$, for some $1 \le i \le n$ and $\langle a\rangle = \langle a^i\rangle$. It follows that i and n are relatively prime. Conversely, if i and n are relatively prime with $1 \le i \le n$, then $\sigma_i : G \to G$ given by $a^m \mapsto a^{mi}$ is checked to be a group monomorphism, hence an isomorphism as G is finite. The map $(\mathbb{Z}/n\mathbb{Z})^\times \to \mathrm{Aut}(G)$ given by $i \mod n \mapsto \sigma_i$ is checked to be an isomorphism, so we also have $|\mathrm{Aut}(G)| = |(\mathbb{Z}/n\mathbb{Z})^\times| = \varphi(n)$. □

Remark 55.3. In F, we shall show the following: If p is an odd prime then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic. If $p = 2$, this is not true. Indeed if $m \ge 3$, then $(\mathbb{Z}/2^m\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{m-2}\mathbb{Z})$. It then follows, using the Chinese Remainder Theorem, that $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic if and only if $n = 2, 4, p^r$, or $2p^r$ where p is an odd prime.

Proposition 55.4. *Let K be a splitting field of $t^n - 1$ in $F[t]$ over F and*
$$U = \{z \in K \mid z^n = 1\} = \{z \mid z \text{ a root of } t^n - 1 \text{ in } K\}.$$
Then U is a cyclic subgroup of $K^\times$. Suppose, in addition, that either $\mathrm{char}\, F = 0$ or $\mathrm{char}\, F \nmid n$. Then $|U| = n$ and the Galois group $G(K/F)$ is isomorphic to a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. In particular, $G(K/F)$ is abelian and $|G(K/F)| \mid \varphi(n)$.

Proof. As $(z_iz_j)^n = 1 = (z_i^{-1})^n$, for all $z_i, z_j \in U$, we know that U is a group. Since it is a finite subgroup of $K^\times$, it is cyclic by Theorem 34.15. As $K = F(U)$, every σ in $G(K/F)$ is determined by $\sigma(z)$, with $z \in U$, so
$$\varphi : G(K/F) \to \Sigma(U) \text{ given by } \sigma \mapsto \sigma|_U$$
is a group monomorphism. As U is cyclic, say $U = \langle z\rangle$, and every σ in $G(K/F)$ satisfies and $\sigma(z^i) = \sigma(z)^i$, it follows that every such σ is completely determined by $\sigma(z)$ and $\sigma|_U$ lies in the automorphism group $\mathrm{Aut}(U)$ of U. If $\mathrm{char}\, F = 0$ or $\mathrm{char}\, F \nmid n$, we have $(t^n - 1)' = nt^{n-1} \ne 0$ has no root in common with $t^n - 1$ in $F[t]$, i.e., $t^n - 1$ has only simple roots in every field extension of F. In particular, $|U| = n$. □

In the proposition, the element z satisfying $U = \langle z \rangle$ is called a *primitive nth root of unity*. If, in the proposition, $\operatorname{char} F = 0$ or $\operatorname{char} F \not| \, n$, then there exists $\varphi(n)$ primitive roots of unity.

Remark 55.5. A (deep) theorem of Kronecker-Weber says: Suppose a finite extension of fields $K/\mathbb{Q}$ is Galois with abelian Galois group. We call such an extension is an *abelian extension* of $\mathbb{Q}$. Then there exists a primitive nth root of unity ζ, e.g., $e^{2\pi\sqrt{-1}/n}$, for some n, such that K is a subfield of $\mathbb{Q}(\zeta)$. One says that all abelian extensions of $\mathbb{Q}$ are *cyclotomic*. We shall show later this is true in the special case that K/F is quadratic.

Remarks 55.6. The following are left as exercises:

1. If $z \in F$, then $|G(K/F)| = 1 = [K : F]$.
2. Suppose that $|F| = q$ and L/F a field extension satisfying $[L : F] = r$. Let $n = q^r - 1$. Then $x^n = 1$ for all x in L. The field L is a splitting field of $t^n - 1$ over F, hence also of $t^{q^r} - t$ in $F[t]$, and $G(K/F)$ is cyclic of order r.
3. If $F = \mathbb{Q}$, then $[K : \mathbb{Q}] = \varphi(n)$ and $G(K/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$.

We did not investigate the case when $\operatorname{char} F \mid n$ in the above. Can you say anything in this case?

Corollary 55.7. *Let F be a field of characteristic zero or* $\operatorname{char} F \not| \, n$ *satisfying the polynomial $t^n - 1$ in $F[t]$ splits over F with U the set of nth roots of units in F. Suppose that K is a splitting field of the polynomial $t^n - a$ in $F[t]$ over F with a nonzero. Then $G(K/F)$ is a cyclic subgroup isomorphic to a subgroup of U. In particular, $|G(K/F)| \mid n$.*

Proof. Let $U = \{z_1, \dots, z_n\}$, a cyclic subgroup of $F^\times$ of order n. As $(t^n - a)' = nt^{n-1}$ has only zero as a root, $t^n - a$ has n distinct roots in K. Let r in K be a root of $t^n - a$, then $rz_i,\ i = 1, \dots, n,$ are all its roots. As $z_i \in F$ for all i, we know that $K = F(r)$ and $\sigma(rz_i) = \sigma(r)z_i$ for all $\sigma \in G(K/F)$ and all i. Consequently, $\sigma(r)$ determines σ for each $\sigma \in G(K/F)$. If $\sigma(r) = rz_i$, then $\sigma(r)/r = z_i$ lies in U. Therefore, we have a map

$$\varphi : G(K/F) \to U \text{ defined by } \sigma \mapsto \frac{\sigma(r)}{r}.$$

If σ and τ lie in $G(K/F)$, say $\sigma(r) = rz_i$ and $\tau(r) = rz_j$, then

$$\varphi(\sigma\tau) = \frac{\sigma\tau(r)}{r} = \frac{\sigma(r)z_j}{r} = \frac{\sigma(r)z_j}{r} = z_i z_j = \varphi(\sigma)\varphi(\tau),$$

so φ is a well-defined group homomorphism. If

$$\frac{\sigma(r)}{r} = \frac{\tau(r)}{r}, \text{ then } \sigma(r) = \tau(r),$$

hence $\sigma = \tau$ and φ is monic. As U is cyclic of order n, we have $G(K/F)$ is cyclic of order dividing n. $\square$

With these preliminaries, we can make some explicit computations.

Computation 55.8. Let F be a field. [You should fill in all the omitted details.]

1. $G(F/F) = 1$ and F/F is Galois.

2. Suppose that the characteristic of F is not two, d an element in F that is not a square in F. Then $K = F(\theta)$ is a splitting field of $t^2 - d$ in $F[t]$ over F where θ is a root of $t^2 - d$. Then K/F is Galois and $G(K/F) \cong \mathbb{Z}/2\mathbb{Z}$. [Of course, one often writes $\sqrt{d}$ for θ.]
3. Let $f = (t^2 - 2)(t^2 - 3)$ in $\mathbb{Q}[t]$ and $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Then K is a splitting field of f over $\mathbb{Q}$. As $\sqrt{2} \notin \mathbb{Q}(\sqrt{3})$, the polynomial $t^2 - 2$ is irreducible in $\mathbb{Q}(\sqrt{3})[t]$. Let $G(K/\mathbb{Q}(\sqrt{3})) = \langle \sigma \rangle$ were σ maps $\sqrt{3} \mapsto -\sqrt{3}$. Similarly, $G\big(K/\mathbb{Q}(\sqrt{2})\big) = \langle \tau \rangle$ where τ maps $\sqrt{2} \mapsto -\sqrt{2}$. Then $G(K/\mathbb{Q}) = \{1, \sigma, \tau, \sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $K/\mathbb{Q}$ is Galois.
4. Let $f = t^3 - 2$ in $\mathbb{Q}[t]$. It is irreducible, so $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. Let $\zeta = \cos(2\pi/3) + \sqrt{-1}\sin(2\pi/3)$ and $K = \mathbb{Q}(\sqrt[3]{2}, \zeta)$. As $m_{\mathbb{Q}}(\zeta) = t^2 + t + 1$, we have $[Q(\zeta) : \mathbb{Q}] = 2$. The field K is a splitting field of f over $\mathbb{Q}$ so $[K : \mathbb{Q}] = 6$. As K is also the splitting field of f over $\mathbb{Q}(\zeta)$ and of degree three, the Galois group $G\big(K/\mathbb{Q}(\zeta)\big)$ is cyclic and isomorphic to a subgroup of $\mathbb{Z}/3\mathbb{Z}$. Since f remains irreducible in $\mathbb{Q}(\zeta)[t]$ (why?), there exists a $\mathbb{Q}(\zeta)$-automorphism τ of K satisfying $\sqrt[3]{2} \mapsto \zeta\sqrt[3]{2}$. Therefore, $G\big(K/\mathbb{Q}(\zeta)\big)$ is cyclic of order three. The polynomial $m_{\mathbb{Q}}(\zeta)$ has no real roots, so remains irreducible in $\mathbb{Q}(\sqrt[3]{2})[t]$, so K is a splitting field of $t^2 + t + 1$ over $\mathbb{Q}(\sqrt[3]{2})$ and there exists a $\mathbb{Q}(\sqrt[3]{2})$-automorphism $\sigma : K \to K$ such that $\zeta \mapsto \zeta^{-1} = \overline{\zeta}$. Therefore, $G\big(K/\mathbb{Q}(\sqrt[3]{2})\big)$ is cyclic of order two. It follows $G(K/\mathbb{Q})$ contains an element of order two and an element of order three, hence $[K : \mathbb{Q}] = 6 = |G(K/\mathbb{Q})|$ and $K/\mathbb{Q}$ is Galois. As
$$\begin{array}{rcccl} \tau\sigma(\sqrt[3]{2}) & = & \tau(\sqrt[3]{2}) & = & \zeta\sqrt[3]{2} \quad \text{and} \\ \sigma\tau(\sqrt[3]{2}) & = & \sigma(\zeta\sqrt[3]{2}) & = & \zeta^{-1}\sqrt[3]{2}, \end{array}$$
$G(K/\mathbb{Q})$ is not abelian, hence isomorphic to $S_3 \cong D_3$. Any element of $G(K/\mathbb{Q})$ must take $\sqrt[3]{2}$ to a root of $t^3 - 2$ in K, so $Q(\sqrt[3]{2}) \subset \mathbb{R}$ means that the subgroup $G(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = 1$ and $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not Galois. Note also that $G\big(K/\mathbb{Q}(\zeta)\big)$ is normal in $G(K/\mathbb{Q})$ but $G\big(K/\mathbb{Q}(\sqrt[3]{2})\big)$ is not normal in $G(K/\mathbb{Q})$.
The following picture summarizes this:
$$\begin{tikzcd}
& K = \mathbb{Q}(\sqrt[3]{2}, \zeta) \arrow[dl, dash, "2"', "gal"] \arrow[dr, dash, "3", "gal"'] \arrow[dd, dash, "6"] & \\
\mathbb{Q}(\sqrt[3]{2}) \arrow[dr, dash, "3"', "not\ gal"] & & \mathbb{Q}(\zeta) \arrow[dl, dash, "2", "gal"'] \\
& \mathbb{Q} &
\end{tikzcd}$$
5. Let $f = t^5 - 1$ in $\mathbb{Q}[t]$ and $\zeta = \cos(2\pi/5) + \sqrt{-1}\sin(2\pi/5)$ in $\mathbb{C}$. Then $K = \mathbb{Q}(\zeta)$ is the splitting field of f and $[K : \mathbb{Q}] = 4 = \varphi(5)$, as $m_{\mathbb{Q}}(\zeta) = t^4 + t^3 + t^2 + t + 1$. We know that group homomorphism $G(K/\mathbb{Q}) \to (\mathbb{Z}/5\mathbb{Z})^\times$ given by $\sigma \mapsto i \mod 5$ if $\sigma \in G(K/\mathbb{Q})$ satisfies $\sigma(\zeta) = \zeta^i$ is monic, so $|G(K/\mathbb{Q})| = 1, 2,$ or 4. As K is the splitting field of the irreducible polynomial $m_Q(\zeta)$ in $\mathbb{Q}[t]$, the Galois group $G(K/\mathbb{Q})$ acts transitively on its roots, hence $|G(K/\mathbb{Q})| = 4$. Therefore, $K/\mathbb{Q}$ is Galois and $G(K/\mathbb{Q}) \cong (\mathbb{Z}/5\mathbb{Z})^\times$

Let $\sigma_i : K \to K$ be the $\mathbb{Q}$-automorphism determined by $\zeta \mapsto \zeta^i$. Then σ_4 takes ζ to $\zeta^{-1} = \overline{\zeta}$, so has order two. If $H = \langle \sigma_4 \rangle$, a subgroup of index two in $G(K/\mathbb{Q})$, then its fixed field is

$$\begin{aligned} K^H &= \{a + b(\zeta^2 + \zeta^3) \mid a, b \in \mathbb{Q}\} \\ &= \{a + b(\zeta^2 + \overline{\zeta}^2) \mid a, b \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{5}). \end{aligned}$$

[Can you show this?]. We have $[K : K^H] = 2 = [K^H : \mathbb{Q}]$, so $K/\mathbb{Q}$ is a square root tower (which leads to the construction of a regular pentagon).

6. Let K be a splitting field of the irreducible polynomial $f = t^4 - 2$ in $\mathbb{Q}[t]$. The roots of f are: $\sqrt[4]{2}$, $\sqrt{-1}\sqrt[4]{2}$, $-\sqrt[4]{2}$, $-\sqrt{-1}\sqrt[4]{2}$; so $K = \mathbb{Q}(\sqrt[4]{2}, \sqrt{-1})$. As $\sqrt{-1} \notin \mathbb{Q}(\sqrt[4]{2}) \subset \mathbb{R}$, the polynomial $t^2 + 1$ is irreducible in $\mathbb{Q}(\sqrt[4]{2})[t]$ and

$$[K : \mathbb{Q}] = [K : \mathbb{Q}(\sqrt[4]{2})][\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}] = 8 \geq |G(K/\mathbb{Q})|.$$

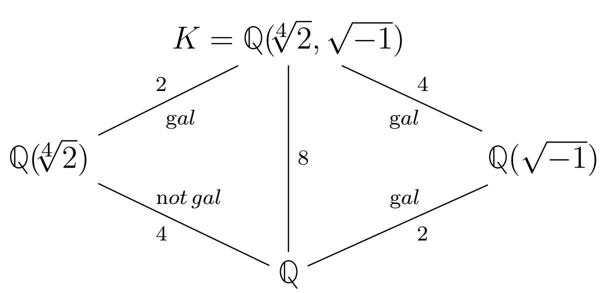
Since $t^4 - 1$ splits over $\mathbb{Q}(\sqrt{-1})$, the group $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ is cyclic of order $1, 2$, or 4 by Corollary 55.7. We show that it is of order 4. We know that $[K : \mathbb{Q}(\sqrt{-1})] = 4$ and K is a splitting field of $m_{\mathbb{Q}}(\sqrt[4]{2})$ over $\mathbb{Q}$. We also know that $\sqrt{2} \notin \mathbb{Q}(\sqrt{-1})$, so f does not factor in $\mathbb{Q}(\sqrt{-1})[t]$ into two quadratic polynomials. It follows that $m_{\mathbb{Q}}(\sqrt[4]{2})$ must remain irreducible in $\mathbb{Q}(\sqrt{-1})[t]$, i.e., $m_{\mathbb{Q}(\sqrt{-1})}(\sqrt[4]{2}) = m_{\mathbb{Q}}(\sqrt[4]{2})$. As $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ acts transitively on the roots of $m_{\mathbb{Q}(\sqrt{-1})}(\sqrt[4]{2})$, we must have $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ is a group of order at least 4, hence 4.
Now let

$$\begin{gathered} \tau : K \to K \text{ be the element in } G\big(K/\mathbb{Q}(\sqrt[4]{2})\big) \\ \text{satisfying } \sqrt{-1} \mapsto -\sqrt{-1} \end{gathered}$$

and

$$\begin{gathered} \sigma : K \to K \text{ be the element in in } G\big(K/\mathbb{Q}(\sqrt{-1})\big) \\ \text{satisfying } (\sqrt[4]{2}) \mapsto \sqrt{-1}\sqrt[4]{2}. \end{gathered}$$

Then $\langle \tau, \sigma \rangle \subset G(K/\mathbb{Q})$. It is easy to see this forces $|G(K/\mathbb{Q})| = 8$. Checking $\tau\sigma\tau^{-1} = \sigma^3$ shows that $G(K/\mathbb{Q}) \cong D_4$. Note also that $K/\mathbb{Q}$ is Galois but $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not as $G(\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}) = \{1, \sigma^2|_{Q(\sqrt[4]{2})}\}$.
The following picture summarizes this:



Note we have another interesting intermediate field of K/F, viz., $E = \mathbb{Q}(\sqrt{-1}, \sqrt{2})$. As $[K : \mathbb{Q}(\sqrt{-1})] = [K : \mathbb{Q}]/[\mathbb{Q}(\sqrt{-1}) : \mathbb{Q}]$ and $\sqrt{2}$ lies in K with $\sqrt[4]{2}$ not in E. $\mathbb{Q} < \mathbb{Q}(\sqrt{-1}) < E < K$ is a square root tower. [Can you say more?]

7. Let p be a prime, $F = (\mathbb{Z}/p\mathbb{Z})(x)$ with x transcendental over $\mathbb{Z}/p\mathbb{Z}$ and $f = t^p - x$ a polynomial in $F[t]$. Let K be a splitting field of f over F and α in K a root of f. Then $t^p - x = t^p - \alpha^p$ in $K[t]$, so f has only one root and $K = F(\alpha)$ with $\alpha \notin F$ (why?). Since $G(K/F)$ takes the roots of f to roots of f, the Galois group $G(K/F) = 1$ and $K = K^{G(K/F)}$. So K/F is not Galois. Check that f is irreducible over F so $[K : F] = p$ and K/F is a splitting field of an irreducible polynomial. Note that K/F is not separable.

Exercises 55.9.

1. Let $K = \mathbb{Q}(r)$ with r a root of $t^3 + t^2 - 2t - 1 \in \mathbb{Q}[t]$. Let $r_1 = r^2 - 2$. Show that r_1 is also a root of this polynomial. Find $G(K/\mathbb{Q})$.
2. Suppose the $|K| = p^n$, p a prime, and $F \subset K$. Show that $|F| = p^m$ for some m with $m \mid n$. Moreover, $G(K/F)$ is generated by the *Frobenius automorphism* $\alpha \mapsto \alpha^{p^m}$. In particular, $G(K/F)$ is cyclic.
3. If F is a finite field, n a positive integer, then there exists an irreducible polynomial $f \in F[t]$ of degree n.
4. Let F be a subfield of the real numbers, f an irreducible quartic over F. Suppose that f has exactly two real roots. Show that the Galois group of f is either S_4 or of order 8.
5. Suppose that K/F is Galois with Galois group $G(K/F) \cong S_n$. Show that K is the splitting field of an irreducible polynomial in $F[t]$ of degree n over F.

56. Galois Extensions

The object of Galois theory is to connect intermediate fields of a finite field extension extension to subgroups of the Galois group $G(K/F)$. The results and computations of the last section indicated a close relation when K/F was a splitting field of a separable polynomial. We begin by pursuing that here.

Definition 56.1. We call a finite field extension K/F *normal* if K is the splitting field over F of some non-constant polynomial in $F[t]$.

We shall see that the important properties of normal extensions are independent of a polynomial for which it is a splitting field except for whether the polynomial is separable or not.

It is useful to generalize the definition of a normal field extension. If $\{f_i\}_I$ is a set of non-constant polynomials in $F[t]$, we call an algebraic extension K of F a *splitting field* of the set of polynomials $\{f_i\}_I$, if every f_i, $i \in I$, splits in K and K is the smallest algebraic extension with this property. Such a splitting field exists and is unique up to an F-isomorphism. (Cf. Exercise 56.22(1)). For example, the splitting field of the set of all minimal polynomials over F is algebraically closed and called an *algebraic closure of* F. If an algebraic extension K of F is the splitting field of a set of non-constant polynomials in $F[t]$, we say that K/F is *normal*. Note that if $\{f_i\}_I$ is a finite set of non-constant polynomials in $F[t]$, then K is a splitting field of this set of polynomials over F if and only if it is the splitting field of $\prod_I f_i$.

The following is a very useful criterion to check if a finite field extension is normal.

Proposition 56.2. *Let K/F be a finite extension of fields. Then K/F is normal if and only if any irreducible polynomial in $F[t]$ having a root in K splits over K.*

PROOF. ($\Rightarrow$): Let K be a splitting field the non-constant polynomial g in $F[t]$ over F and $f \in F[t]$ an irreducible polynomial having a root α in K. Let L/K with $\beta \in L$ a root of f. We must show that β lies in K. As f is irreducible, there exists an F-isomorphism $\tau : F(\alpha) \to F(\beta)$ satisfying $\alpha \mapsto \beta$. As $K(\alpha)$ is a splitting field of g over $F(\alpha)$ and $K(\beta)$ is a splitting field of g over $F(\beta)$, there exists an isomorphism $\sigma : K(\alpha) \to K(\beta)$ lifting τ. In particular, σ is also an F-isomorphism, so we have

$$\begin{aligned}[K(\alpha) : F] &= [K(\alpha) : F(\alpha)][F(\alpha) : F] \\ &= [K(\beta) : F(\beta)][F(\beta) : F] = [K(\beta) : F].\end{aligned}$$

But $\alpha \in K$, so $K(\alpha) = K$, and

$$[K : F] = [K(\alpha) : F] = [K(\beta) : F] = [K(\beta) : K][K : F].$$

Therefore, $[K(\beta) : K] = 1$, i.e., $\beta \in K$, as needed.

($\Leftarrow$): As K/F is finite, there exist $\alpha_1, \dots, \alpha_r$ in K, some r, with $K = F(\alpha_1, \dots, \alpha_r)$ and each α_i algebraic over F. Clearly, K is the splitting field of $f = \prod m_F(\alpha_i)$ over F, as each $m_F(\alpha_i)$ has a root in K, hence splits over K and any splitting field of f over F in K must contain all the α_i's. □

Remark 56.3. 1. $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is not normal as the irreducible polynomial $t^3 - 2$ in $\mathbb{Q}[t]$ does not split over $\mathbb{Q}(\sqrt[3]{2})$

[Note: the general proof above is really the same as the special case done before.]

2. Let p be a positive prime and $F = (\mathbb{Z}/p\mathbb{Z})(x)$ with x transcendental over $\mathbb{Z}/p\mathbb{Z}$. Then $f = t^p - x$ in $F[t]$ is irreducible. (Cf. Exercise 53.10 (3).) Let K/F with $\alpha \in K$ a root of f, i.e., $\alpha^p = x$. Set $E = F(\alpha)$. Then $f = (t - \alpha)^p$ in $E[t]$, hence E is a splitting field of f over F. Therefore, E/F is normal. We saw that $G(E/F) = 1$, hence E/F is not Galois.

Proposition 56.4. *Let K/F be a finite, normal extension and $K/E/F$ an intermediate field. Suppose that $\varphi : E \to K$ is an F-homomorphism. Then there exists an F-automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_E = \varphi$, i.e., every F-homomorphism $E \to K$ arises from an F-automorphism of K by restriction.*

PROOF. Let K be a splitting field of the non-constant polynomial f in $F[t]$, hence a splitting field of f over E. In addition, K is a splitting field of $f = \widetilde{\varphi}(f)$ over $\varphi(E)$, so there exists an automorphism $\sigma : K \to K$ lifting $\varphi : E \to \varphi(E)$, so σ lies in $G(K/F)$. □

Using Zorn's Lemma, one can show that the last two propositions hold if we only assume that K/F is algebraic. (Cf. Exercises 56.22(4), (3), respectively.) Alternatively, one can use the existence of an algebraic closure (Corollary 51.2) and its lifting property (Theorem 51.3), together with a restriction argument giving the uniqueness of splitting fields.

Definition 56.5. Let $K/E_i/F$ be finite field extensions, $i = 1, 2$ with K/F normal. We say that E_1 and E_2 are *conjugate over* F if there exists an automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_{E_1} : E_1 \to E_2$ is an isomorphism. Equivalently, by Proposition 56.4, there exists an F-isomorphism $E_1 \to E_2$.

Remarks 56.6. Let L/F be a finite normal extension, $L/K/E_i/F$ field extensions, $i = 1, 2$, with K/F also normal. Then E_1 and E_2 are conjugate relative to the extensions $K/E_i/F$ if and only if they are conjugate relative to the extensions $L/E_i/F$, as any $\sigma \in G(K/F)$ lifts to some $\widehat{\sigma} \in G(L/F)$ and any $\widehat{\sigma} \in G(L/F)$ restricts to $\widehat{\sigma}|_K \in G(K/F)$. That $\widehat{\sigma}$ is an F-automorphism of K follows from the following:

Corollary 56.7. *Let K/F be a finite normal field extension and $K/E/F$ an intermediate field extension. Then K/E is normal. Moreover, the following are equivalent:*

(1) E/F is normal.
(2) $\sigma(E) = E$ for every $\sigma \in G(K/F)$.
(3) $\sigma|_E \in G(E/F)$ for every $\sigma \in G(K/F)$.
(4) The map $\Phi : G(K/F) \to G(E/F)$ given by $\sigma \mapsto \sigma|_E$ is well-defined and an epimorphism.
(5) E is the only conjugate of E.

Proof. If K is the splitting field of the non-constant polynomial $f \in F[t]$ over F, then it is also the splitting field of f over E, so the first statement follows.

We turn to the equivalences. By Proposition 56.4, any $\varphi \in G(E/F)$ lifts to an element in $G(K/F)$ and if $G(K/F) \to G(E/F)$ by $\sigma \mapsto \sigma|_E$ is well-defined – usually it is not as, in general, $\sigma(E) \not\subset E$ – the equivalences of (2), (3), (4) and (5) become clear. So it suffices to show that (1) and (2) are equivalent.

(1) $\Rightarrow$ (2): Let $x \in E$ and $\sigma \in G(K/F)$. As $x \in E$ and E/F is normal, we know that $m_F(x)$ splits over E. Since $G(K/F)$ takes the roots of $m_F(x)$ to roots of $m_F(x)$, the root $\sigma(x)$ also lies in E. As the argument also works for σ^{-1}, Statement (2) follows.

(2) $\Rightarrow$ (1): Let $f \in F[t]$ be an irreducible polynomial having a root $\alpha \in E$. As K/F is normal, f splits over K. Let $\beta \in K$ be root of f. We must show that $\beta \in E$, i.e., f splits over E. We know that there exists an F-isomorphism $\varphi : F(\alpha) \to F(\beta)$ taking $\alpha \mapsto \beta$. By Proposition 56.4, there exists an automorphism $\sigma \in G(K/F)$ satisfying $\varphi = \sigma|_{F(\alpha)}$, so $\beta = \varphi(\alpha) = \sigma(\alpha)$ lies in E as needed. $\square$

Note in the above if E/F is normal, then $\ker \Phi = G(K/E) \triangleleft G(K/F)$ and $G(E/F) \cong G(K/F)/G(K/E)$.

Definition 56.8. Let K/F be a finite field extension. Then a field extension L of K is called a *normal closure* of K/F if L/F is normal and $[L : K]$ is minimal with respect to this property. If this is the case, we write L/K is a normal closure of K/F.

Question 56.9. How would you define the normal closure of an algebraic extension of F?

Proposition 56.10. *Let K/F be a finite field extension. Then a normal closure of K/F exists and is unique up to a K-isomorphism.*

Proof. Let $K = F(\alpha_1, \dots, \alpha_n)$ with each α_i algebraic over F, $f = \prod m_F(\alpha_i)$ in $F[t]$, and L a splitting field of f over K, hence also a splitting field of f over F. Therefore, L/F is normal. By the uniqueness of splitting fields, L is unique up to a K-isomorphism (relative to f). If E/K is a normal extension, then each $m_F(\alpha_i)$ must split over E, so f must split over E. The result now follows easily. $\square$

If in the above proposition, K/F is also separable, then L/F is also separable by Exercise 53.10(6). We shall also give a proof of this below.

The main technical result needed to establish the field theoretic equivalence of a Galois extension (rather than a group theoretic one) is the following:

Proposition 56.11. *Let K/F be a finite extension of fields of degree n. Suppose that L/K with L/F finite and normal and*

$$\tau_1, \dots, \tau_m : K \to L$$

are all the distinct F-homomorphisms. Then the following are true:

(1) $m \leq n$.
(2) $m = n$ if and only if K/F is separable.
(3) $\tau_1(K), \dots, \tau_m(K)$ are all the conjugates of K over F in L (and independent of the normal extension L/F).
(4) Let $E = F\big(\tau_1(K) \cup \dots \cup \tau_m(K)\big)$. Then E/K is a normal closure of K/F.

Using the exercises in 56.22, the assumption in the proposition that L/K be finite may be dropped.

Proof. We may assume that τ_1 is the inclusion, so $m \geq 1$. Let $S = \{\tau_1, \dots, \tau_m\}$.
(1): By assumption, $F \subset K^S$, so $[K : F] \geq |S| = m$ by Artin's Lemma 54.8.
(2): We induct on n. The case of $n = 1$ is immediate, so we may assume that $n > 1$. Let $\alpha \in K \setminus F$ be chosen with α not separable, i.e., $m_F(\alpha)$ not separable, if K/F is not separable. Let

$$r = [F(\alpha) : F] > 1$$

and

$$\alpha = \alpha_1, \dots, \alpha_{r_0} \text{ the distinct roots of } m_F(\alpha) \text{ in } L.$$

[Note that $m_F(\alpha)$ splits over L as L/F is normal.]
So we have

$$\begin{aligned} r = r_0 &\text{ if and only if } \alpha \text{ is separable over } F \\ &\text{ if and only if } K/F \text{ is separable.} \end{aligned}$$

In particular,

$$r_0 < r \text{ if } K/F \text{ is not separable.}$$

Let

$$\varphi_i : F(\alpha) \to F(\alpha_i) \; [\subset L]$$

be the F-isomorphism satisfying $\alpha \mapsto \alpha_i$, $1 \leq i \leq r_0$.
Since the roots of $m_F(\alpha)$ go to roots of $m_F(\alpha)$, we have

$$\varphi_i, \; 1 \leq i \leq r_0, \text{ are all the } F\text{-homomorphisms } F(\alpha) \to L.$$

As L/F is normal, each φ_i lifts to some

$$\widehat{\varphi}_i \in G(L/F),\ 1 \le i \le r_0.$$

[There can be many distinct lifts of each φ_i — chose one.]
Thus

$$\widehat{\varphi}_i|_{F(\alpha)} \neq \widehat{\varphi}_j|_{F(\alpha)} \text{ if } i \neq j.$$

So we have the following:

(*i*) $[K : F(\alpha)] = \dfrac{n}{r}$.

(*ii*) $K/F(\alpha)$ is separable if K/F is separable.

By induction, we have:

$$\text{There exist } m_0 \text{ distinct } F(\alpha)\text{-homomorphisms}$$
$$\psi_1, \dots, \psi_{m_0} : K \to L$$
$$\text{with } m_0 = \frac{n}{r} \text{ if and only if } K/F(\alpha) \text{ is separable.}$$

[Note if K/F is separable so is $K/F(\alpha)$.]
Let

$$\rho_{ij} = \widehat{\varphi}_i\psi_j : K \to L,\ 1 \le i \le r_0,\ 1 \le j \le m_0.$$

and

$$\mathcal{S} = \{\rho_{ij} \mid 1 \le i \le r_0,\ 1 \le j \le m_0\}.$$

Each ρ_{ij} is an F-homomorphism, so $\mathcal{S} \subset \{\tau_1, \dots, \tau_m\}$.

Claim 56.12. *$\mathcal{S}$ contains all the F-homomorphisms $K \to L$ and $m = |\mathcal{S}| = r_0 m_o$:*

If we establish the claim, then

$$m = r_0 m_0 \le m_0 r \le n.$$

So

$$K/F \text{ is separable if and only if } r = r_0 \text{ and } m_0 = \frac{n}{r}$$
$$\text{if and only if } n = m_0 r_0 = m.$$

In particular, if we prove the claim, (2) follows.

Let $\rho : K \to L$ be an F-homomorphism. Then $\rho(\alpha)$ is a root of $m_F(\alpha)$, so there exists an i satisfying $\rho|_{F(\alpha)} = \varphi_i$, hence $\widehat{\varphi}_i^{-1}\rho : K \to L$ is an $F(\alpha)$-homomorphism. Consequently, there exist a j satisfying $\psi_j = \widehat{\varphi}_i^{-1}\rho$. Therefore, $\rho = \widehat{\varphi}_i\psi_j = \rho_{ij}$ lies in $\mathcal{S}$.

To finish proving the claim, we must show that the ρ_{ij} are distinct. Suppose that $\rho_{ij} = \rho_{kl}$. Then

$$\varphi_i = \rho_{ij}|_{F(\alpha)} = \rho_{kl}|_{F(\alpha)} = \varphi_k,$$

hence $\widehat{\varphi}_i = \widehat{\varphi_k}$, so $i = k$. Since

$$\widehat{\varphi}_i\psi_j = \rho_{ij} = \rho_{kl} = \widehat{\varphi_k}\psi_l,$$

it follows that $\psi_j = \psi_l$ as $\widehat{\varphi}_i$ is an automorphism. Consequently, $j = l$. This shows that the ρ_{ij}'s are distinct and establishes the claim.

(3). Each of the $\tau_i : K \to L$ above lifts to some $\widehat{\tau_i}$ in $G(L/F)$ as before. In particular, $\widehat{\tau_i}(K) = \tau_i(K)$. If $\sigma \in G(L/F)$, there exists a j such that $\sigma|_K = \tau_j$. It follows that $\tau_1(K), \dots, \tau_m(K)$ are all the conjugates of K (repetitions possible).

(4): Suppose that L/F is also a normal closure of K/F. Set $E = F\big(\tau_1(K) \cup \cdots \cup \tau_m(K)\big) \subset L$. Write $K = F(\beta_1, \dots, \beta_s)$, for some β_i, and set $f = \prod_{i=1}^{s} m_F(\beta_i)$ in $F[t]$. Then L is a splitting field of f over F. By the argument proving (2), we know that for any root of $m_F(\beta_i)$, there exist an F-homomorphism $K \to L$ sending β_i to that root. As $\tau_1, \dots, \tau_m : K \to L$ are all the F-homomorphisms and each must take roots of $m_K(\beta_i)$ to roots of $m_F(\beta_i)$, the field E contains all the roots of the $m_F(\beta_i)$. As L/F is a splitting field of f, we must have $E = L$. □

Corollary 56.13. *Let K/F be a extension of fields with α an element in K algebraic over F. Then α is separable over F if and only if $F(\alpha)/F$ is separable.*

Proof. Let $L/F(\alpha)$ be a field extension such that L/F is finite and normal. We know that $m_F(\alpha)$ splits over L. Let $\alpha = \alpha_1, \dots, \alpha_m$ be the distinct roots of $m_F(\alpha)$ in L. We know that there exist F-isomorphisms $\varphi_i : F(\alpha) \to F(\alpha_i)$ satisfying $\alpha \mapsto \alpha_i$ for $i = 1, \dots, m$ and these are all the distinct F-homomorphisms into L. Thus $F(\alpha)/F$ is separable if and only if $[F(\alpha) : F] = \deg m_F(\alpha) = m$ if and only if α is separable over F. □

Proposition 56.14. *Let K be a splitting field of a separable polynomial in $F[t]$. Then K/F is separable and Galois.*

Proof. Suppose that K is a splitting field of the separable polynomial $f \in F[t]$. We may assume that $F < K$. Let $\alpha \in K$ be a root of f not in F. As α is separable over F, the extension $F(\alpha)/F$ is separable by the corollary. Therefore, there exist $m = [F(\alpha) : F]$ F-homomorphisms $\varphi_i : F(\alpha) \to K$, $1 \leq i \leq m$. As K/F is normal, each φ_i lifts to some $\widehat{\varphi_i} \in G(K/F)$. Since K is a splitting field of the separable polynomial f over $F(\alpha)$, by induction on $[K : F]$, we conclude that $K/F(\alpha)$ is separable. As K/K is the normal closure of $K/F(\alpha)$ (as well as K/F), there exist precisely $[K : F(\alpha)]$ distinct $F(\alpha)$-homomorphisms $\psi_j : K \to K$, $1 \leq j \leq [K : F(\alpha)]$. Moreover, the argument to prove Claim 56.12 shows that

$$\mathcal{S} = \{\widehat{\varphi_i}\psi_j \mid 1 \leq i \leq [F(\alpha) : F],\ 1 \leq j \leq [K : F(\alpha)]\}$$

is the set of all F-homomorphisms $K \to K$ and has $[K : F] = [K : F(\alpha)][F(\alpha) : F]$ elements, hence K/F is separable. Since $\mathcal{S} \subset G(K/F)$, we have $|G(K/F)| = [K : F]$ by Artin's Lemma 54.8, so K/F is also Galois. □

Corollary 56.15. *Let K/F be a finite separable field extension and L/K a normal closure of K/F. Then L/F is separable and Galois.*

Proof. Let $K = F(\alpha_1, \dots, \alpha_n)$. Then L is a splitting field of the separable polynomial $\prod m_F(\alpha_i)$ in $F[t]$, hence L/F is separable and Galois. □

Corollary 56.16. *Let $K = F(\alpha_1, \dots, \alpha_n)$ with α_i separable over F for each $i = 1, \dots, n$. Then K/F is separable.*

Proof. Let L/K be a normal closure of K/F, hence the splitting field of the separable polynomial $\prod m_F(\alpha_i)$. Therefore, L/F is separable hence so is K/F. □

Corollary 56.17. *If K/F is finite field extension and $K/E/F$, then K/F is separable if and only if K/E and E/F are separable.*

PROOF. $(\Rightarrow)$ has already been done.
$(\Leftarrow)$: Let L/K be a (finite) normal extension and

$$\varphi_1, \ldots, \varphi_m : E \to L \text{ be all the distinct } F\text{-homomorphisms}$$
$$\psi_1, \ldots, \psi_n : K \to L \text{ be all the distinct } E\text{-homomorphisms.}$$

Let $\widehat{\varphi}_i \in G(L/F)$ lift φ_i for $1 \leq i \leq m$. Then as in the proof of Claim 56.12,

$$|\{\widehat{\varphi}_i \psi_j \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}| = mn.$$

As K/E and E/F are separable, $[K : F] = [K : E][E : F] = mn$, hence K/F is separable by Proposition 56.11. □

We arrive at our goal of finding the field theoretic interpretation for a finite field extension K/F to be Galois.

Theorem 56.18. *Let K/F be a finite extension of fields. Then K/F is Galois if and only if K/F is normal and separable.*

PROOF. $(\Leftarrow)$ has already been done.
$(\Rightarrow)$: Let L/K be a normal closure of K/F. As K/F is Galois, the number of F-homomorphisms $K \to L$ is at least $|G(K/F)| = [K : F]$. It follows by Proposition 56.11 there exist exactly $|G(K/F)| = [K : F]$ such F-homomorphisms and, in addition, K/F must be separable. Since each of these homomorphisms lies in $G(K/F)$, Proposition 56.11(4), shows that $L = F\big(\cup_{G(K/F)} \sigma(K)\big) = K$. □

Corollary 56.19. *Let K/F be a finite extension and $K/E/F$ an intermediate field. Suppose that K/F is Galois. Then*

(1) K/E is Galois.
(2) E/F is Galois if and only if E/F is normal.

PROOF. As K/F is Galois, it is separable and normal, so K/E and E/F are separable and K/E is normal. Therefore, K/E is Galois and E/F is normal if and only if E/F is Galois. □

Let K/F be a finite field extension and $K/E/F$ an intermediate field. Recall that we have seen, in general, K/F being Galois does not imply that E/F is Galois. For example, $K = \mathbb{Q}(\sqrt[3]{2}, e^{2\pi\sqrt{-1}/3})$, $E = \mathbb{Q}(\sqrt[3]{2})$, and $F = \mathbb{Q}$. A question left to the reader is: If K/E and E/F are Galois, is K/F Galois?

Definition 56.20. A field F is called *perfect* if any algebraic extension K/F is separable.

So we have

Remark 56.21. Let F be a field.

(1) If $\operatorname{char} F = 0$, then F is perfect.
(2) If F is a finite field, then F is perfect.

(3) If F is perfect, K/F a finite extension, then K/F is Galois if and only if K/F is normal.

(4) If x is transcendental over $\mathbb{Z}/p\mathbb{Z}$, with $p > 0$ a prime and $F = (\mathbb{Z}/p\mathbb{Z})(x)$, then F is not perfect as $t^p - x$ in $F[t]$ is not separable.

Exercises 56.22.

1. Let F be a field and $\{f_i\}_I$ be a set of non-constant polynomials in $F[t]$. Prove that a splitting field of $\{f_i\}_I$ over F exists and is unique up to an F-isomorphism.
2. Let K/F be an algebraic extension. Show that any F-homomorphism $\sigma : K \to K$ is an F-automorphism.
3. Let K/F be a normal (possibly infinite) extension of fields and $K/E/F$ an intermediate field. Suppose that $\varphi : E \to K$ is an F-homomorphism. Then there exists an F-automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_E = \varphi$, i.e., every F-homomorphism $E \to K$ arises from an F-automorphism of K by restriction.
4. Let K/F be an algebraic extension of fields. Then K/F is normal if and only if any irreducible polynomial in $F[t]$ having a root in K splits over K.
5. Let K/F be an algebraic field extension. Answer Question 56.9 and prove that a normal closure of K/F exists and is unique up to a K-isomorphism.
6. Let F be a field and L and algebraic closure of F. Show that L/F is normal. Is it true that $L^{G(L/F)} = F$, i.e., L/F is (infinite) Galois? If not, give an example when it is not.
7. Show that there exist uncountably many field automorphisms of $\mathbb{C}$.
8. Let K/F be an algebraic extension of fields and $K/E/F$ an intermediate field. Show that K/F is separable if and only if K/E and E/F are separable.
9. Suppose that K/F is Galois and $\alpha \in K$ has precisely r distinct images under $G(K/F)$. Show $[F(\alpha) : F] = r$.
10. Let F be a field of positive characteristic p and $f = t^p - t - a \in F[t]$.
    (a) Show the polynomial f has no multiple roots.
    (b) If α is a root of f, show so is $\alpha + k$ for all $0 \leq k \leq p - 1$.
    (c) Show f is irreducible if and only if f has no root in F.
    (d) Suppose that $a \neq b^p - b$ for any $b \in F$. Find $G(K/F)$ where K is a splitting field of $t^p - t - a \in F[t]$.
11. Let K/F be a finite field extension and $K/E/F$ an intermediate field. Suppose that K/E and E/F are Galois. Is K/F Galois? Either prove or provide a counterexample.
12. Prove if F is not a finite field and u, v are algebraic and separable over F that there exists an element $a \in F$ such that $F(u, v) = F(u + av)$. Is this true if $|K| < \infty$ with $K(u) < K(u, v)$ and $K(v) < K(u, v)$?
13. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, u)$ where $u^2 = (9 - 5\sqrt{3})(2 - \sqrt{2})$. Show that $K/\mathbb{Q}$ is normal and find $G(K/\mathbb{Q})$.
14. Let L/F be a finite normal extension and $E = L^{G(L/F)}$. Show

$$E = \{\alpha \mid \alpha \text{ is purely inseparable over } F\}.$$

(Cf. Exercises 53.10(11) and (14).) In particular, L/E is separable and E/F is purely inseparable, i.e., we have

$$\begin{tikzcd} & L \arrow[dl, dash, "\mathrm{p}\,insep"'] \arrow[dr, dash, "sep"] & \\ L_{sep} \arrow[dr, dash, "sep"'] & & E = L^{G(L/F)} \arrow[dl, dash, "\mathrm{p}\,insep"] \\ & F & \end{tikzcd}$$

57. The Fundamental Theorem of Galois Theory

In this section we put together the pieces that we have developed, i.e., the group theoretic and field theoretic interpretations of Galois extensions. As an application we prove the Fundamental Theorem of Algebra. We begin with the following lemma.

Lemma 57.1. *Let K/F be a finite Galois extension of fields, $H_i \subset G(K/F)$ a subgroup and $E_i = K^{H_i}$ for $i = 1, 2$. If σ is an element of the Galois group $G(K/F)$, then*

$$\sigma|_{E_1} : E_1 \to E_2 \text{ is an isomorphism if and only if } H_2 = \sigma H_1 \sigma^{-1},$$

i.e., E_1 and E_2 are conjugate over F (via σ) if and only if H_1 and H_2 are conjugate in $G(K/F)$ (via σ).

Proof. We know that K/E_i is Galois and $H_i = G(K/E_i)$ for $i = 1, 2$. In particular, $[K : E_i] = |H_i|$ for $i = 1, 2$, so under either condition, we have

$$|H_1| = |H_2| = [K : E_1] = [K : E_2].$$

($\Rightarrow$): Let $\tau \in H_1$ and $y \in E_2$. Then $\sigma^{-1}(y)$ lies in E_1, so $\sigma\tau\sigma^{-1}(y) = \sigma\sigma^{-1}(y) = y$ as $\tau|_{E_1} = 1_{E_1}$. Therefore, we have $\sigma H_1 \sigma^{-1} \subset H_2 = G(K/E_2)$, hence $\sigma H_1 \sigma^{-1} = H_2$ as $|\sigma H_1 \sigma^{-1}| = |H_1| = |H_2| < \infty$.
($\Leftarrow$): Let $x \in E_1$ and $\tau \in H_1$. Then $\sigma\tau\sigma^{-1}(\sigma(x) = \sigma\tau(x) = \sigma(x)$, as $\tau|_{E_1} = 1|_{E_1}$. Consequently, $\sigma(x)$ lies in $K^{\sigma H_1 \sigma^{-1}} = K^{H_2} = E_2$, and $\sigma|_{E_1} : E_1 \to E_2$ and is a homomorphism. As $[K : E_1] = [K : E_2]$ and $[E_1 : F] = [E_2 : F]$, the map σ_{E_1} is an isomorphism (since a linear monomorphism of finite dimensional vector spaces of the same dimension). □

Notation 57.2. Let K/F be an algebraic extension of fields. We set

$$\mathcal{F}(K/F) := \{E \mid K/E/F \text{ is an intermediate field}\}$$
$$\mathcal{G}(K/F) := \{H \mid H \subset G(K/F) \text{ is a subgroup}\}.$$

Putting together all our results leads to:

Theorem 57.3. (The Fundamental Theorem of Galois Theory) *Suppose that K/F is a finite Galois extension. Then*

$$i : \mathcal{G}(K/F) \to \mathcal{F}(K/F) \text{ given by } H \mapsto K^H$$

is a bijection of sets with inverse

$$j : \mathcal{F}(K/F) \to \mathcal{G}(K/F) \text{ given by } E \mapsto G(K/E).$$

Moreover,

1. *The bijection i is order-reversing, i.e., $H_1 \subset H_2$ if and only if $K^{H_1} \supset K^{H_2}$.*
2. *If $E \in \mathcal{F}(K/F)$, then E/F is normal (hence Galois) if and only if $G(K/E) \triangleleft G(K/F)$.*
3. *If $E \in \mathcal{F}(K/F)$ with E/F normal, then the canonical epimorphism*
$$^{-} : G(K/F) \to G(E/F) \ \text{ given by } \ \sigma \mapsto \sigma|_E$$
induces an isomorphism
$$G(K/F)/G(K/E) \to G(E/F) \ \text{ given by } \ \sigma G(K/E) \mapsto \sigma|_E.$$
4. *If $H \in \mathcal{G}(K/F)$, then $|H| = [K : K^H]$ and K/K^H is Galois with $H = G(K/K^H)$.*
5. *If $H \in \mathcal{G}(K/F)$, then $[G(K/F) : H] = [K^H : F]$.*

The bijection $i : \mathcal{G}(K/F) \to \mathcal{F}(K/F)$ given by $H \mapsto K^H$ and its inverse is called the *Galois Correspondence*. We have the following picture:

$$
\begin{array}{ccccc}
 & K & \text{———} & G(K/K) = 1 & \\
 & | & & | & \\
| = \cup & E & \text{———} & G(K/E) & | = \cap \\
 & \| & & \| & \\
 & F & \text{———} & G(K/F) &
\end{array}
$$

with the identified verticals having the same degree (index) and if E/F is normal, then $G(K/F)/G(K/E) \cong G(E/F)$.

Proof. We already know that $H_1 = H_2$ in $\mathcal{G}(K/F)$ if and only if $K^{H_1} = K^{H_2}$ in $\mathcal{F}(K/F)$, so $i : \mathcal{G}(K/F) \to \mathcal{F}(K/F)$ is injective. Let $E \in \mathcal{F}(K/F)$. Then K/E is Galois, so $G(K/E) \mapsto K^{G(K/E)} = E$. Hence $i : \mathcal{G}(K/F) \to \mathcal{F}(K/F)$ is bijective.

(1) is clear.

(2): We know that if $E \in \mathcal{F}(K/F)$, then E/F is normal if and only if $\sigma(E) = E$ for all $\sigma \in G(K/F)$. By Lemma 57.1, this is true if and only if $\sigma G(K/E)\sigma^{-1} = G(K/E)$ for all $\sigma \in G(K/F)$, i.e., if and only if $G(K/E) \triangleleft G(K/F)$.

We have seen if E/F is normal then $\Phi : G(K/F) \to G(E/F)$ by $\sigma \mapsto \sigma|_E$ is a well-defined group epimorphism. As
$$\ker \Phi = \{\sigma \in G(K/F) \mid \sigma|_E = 1_E\} = G(K/E),$$
Φ induces an isomorphism $G(K/F)/G(K/E) \cong G(E/F)$.

(4): As K/K^H is Galois, $H = G(K/K^H)$, so $H = [K : K^H]$ by Artin's Theorem 54.15.

(5): By Artin's Theorem 54.15,
$$[K : K^H][K^H : F] = [K : F] = |G(K/F)| = [G(K/F) : H]|H|,$$
so (4) $\Rightarrow$ (5). □

Corollary 57.4. *Let K/F be a finite Galois extension of fields. Then $\mathcal{F}(K/F)$ is finite, i.e., there exist only finitely many intermediate fields E with $K/E/F$.*

Proof. By the Fundamental Theorem of Galois Theory, we have $|\mathcal{F}(K/F)| = |\mathcal{G}(K/F)|$, hence $\mathcal{F}(K/F)$ is a finite set. $\square$

Corollary 57.5. *Let K/F be a finite separable extension. Then $\mathcal{F}(K/F)$ is finite.*

Proof. Let L/K be a normal closure of K/F. Then L/F is finite, separable, and normal, so Galois. Thus $|\mathcal{F}(K/F)| \leq |\mathcal{F}(L/F)|$ is finite. $\square$

We give a few examples, before deriving significant consequences of this fundamental theorem.

Example 57.6. 1. We know that $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $(t^2-2)(t^2-3)$ in $\mathbb{C}$ over $\mathbb{Q}$. Then we have seen that $G(K/\mathbb{Q}) = \{1, \sigma, \tau, \sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ with

$$\sigma : K \to K \text{ the } \mathbb{Q}(\sqrt{3})\text{-automorphism given by } \sqrt{2} \mapsto -\sqrt{2},$$
$$\tau : K \to K \text{ the } \mathbb{Q}(\sqrt{2})\text{-automorphism given by } \sqrt{3} \mapsto -\sqrt{3}.$$

Intermediate fields and corresponding subgroups are given by the following diagrams:

$$\begin{tikzcd}
& K \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbb{Q}(\sqrt{2}) \arrow[dr, dash] & \mathbb{Q}(\sqrt{6}) \arrow[d, dash] & \mathbb{Q}(\sqrt{3}) \arrow[dl, dash] \\
& \mathbb{Q} &
\end{tikzcd}
\quad \text{and} \quad
\begin{tikzcd}
& 1 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\langle\tau\rangle \arrow[dr, dash] & \langle\sigma\tau\rangle \arrow[d, dash] & \langle\sigma\rangle \arrow[dl, dash] \\
& G(K/\mathbb{Q}). &
\end{tikzcd}$$

2. Let K be a splitting field of $t^4 - 2$ over $\mathbb{Q}$. Then $K = \mathbb{Q}(\sqrt[4]{2}, \sqrt{-1})$ and $[K : \mathbb{Q}] = 8$. If

$$\sigma : K \to K \text{ the } \mathbb{Q}(\sqrt[4]{2})\text{-automorphism given by } \sqrt{-1} \mapsto -\sqrt{-1},$$
$$\tau : K \to K \text{ the } \mathbb{Q}(\sqrt{-1})\text{-automorphism given by } \sqrt[4]{2} \mapsto \sqrt{-1}\sqrt[4]{2},$$

then $|\langle\sigma\rangle| = 2,\ |\langle\tau\rangle| = 4,\ \sigma\tau\sigma^{-1} = \tau^3,\ G(K/\mathbb{Q}) = \langle\sigma, \tau\rangle \cong D_4$.
Intermediate fields and corresponding subgroups are given by the following diagrams with $\theta = \sqrt[4]{2}$, $i = \sqrt{-1}$:

$$\begin{tikzcd}
& & K \arrow[dll, dash] \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] \arrow[drr, dash] & & \\
\mathbb{Q}(\theta) \arrow[dr, dash] & \mathbb{Q}(i\theta) \arrow[d, dash] & \mathbb{Q}(i, \sqrt{2}) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \mathbb{Q}\big((1+i)\theta\big) \arrow[d, dash] & \mathbb{Q}\big((1-i)\theta\big) \arrow[dl, dash] \\
& \mathbb{Q}(\sqrt{2}) \arrow[dr, dash] & \mathbb{Q}(i) \arrow[d, dash] & \mathbb{Q})(\sqrt{-2}) \arrow[dl, dash] & \\
& & \mathbb{Q} & &
\end{tikzcd}$$

$$\begin{tikzcd}
& & 1 \arrow[dll, dash] \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] \arrow[drr, dash] & & \\
\langle\sigma\rangle \arrow[dr, dash] & \langle\tau^2\sigma\rangle \arrow[d, dash] & \langle\tau^2\rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \langle\tau\sigma\rangle \arrow[d, dash] & \langle\tau^3\sigma\rangle \arrow[dl, dash] \\
& \{1,\tau^2,\sigma,\sigma\tau^2\} \arrow[dr, dash] & \langle\tau\rangle \arrow[d, dash] & \{1,\tau\sigma,\tau^3\sigma,\tau^2\} \arrow[dl, dash] & \\
& & \mathbb{Q}. & &
\end{tikzcd}$$

3. Let $L = F(t_1, \dots, t_n)$, the quotient field of $F[t_1, \dots, t_n]$. The symmetric group S_n acts on L as F-automorphisms by permuting the t_i, i.e.,
$$\sigma(t_i) = t_{\sigma(i)} \text{ for all permutations } \sigma \in S_n \text{ and } i = 1, \dots, n.$$
We view $S_n \subset G(L/F)$. Let
$$s_j = s_j(t_1, \dots, t_n) := \sum_{1 \le i_1 < \cdots < i_j \le n} t_{i_1} \cdots t_{i_j}$$
be the jth *elementary symmetric polynomial* of $t_1, \dots, t_n$ for $j = 1, \dots, n$ and set $s_0 = 1$. We know that $t_1, \dots, t_n$ are the roots of the polynomial
$$f := t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n \text{ in } F[s_1, \dots, s_n][t],$$
as $f = (t - t_1) \cdots (t - t_n)$ in $L[t]$. Let $K = L^{S_n}$, so L/K is Galois. Let $E := F(s_1, \dots, s_n) \subset K$. Clearly, L is a splitting field of f over E, and by our theory, we know that
$$[L : E] \le (\deg f)! = n! \text{ and } [L : K] = |G(L/K)| = [S_n| = n!.$$
So $K = E = F(s_1, \dots, s_n)$ and $G(L/K) = G\big(L/F(s_1, \dots, s_n)\big) \cong S_n$. Note that f must be irreducible in $E[t]$, lest $f = f_1 f_2$ in $E[t]$ with $0 < \deg f_1 < n$ for $i = 1, 2$, and we would have $[L : E] \le (\deg f_1)!(\deg f_2)! < (\deg f)! = [L : E]$. Note further that $s_1, \dots, s_n$ are algebraically independent over F as $t_1, \dots, t_n$ are and K/E is finite. (Cf. Proposition 49.4.)

As an application of this computation, we also get a weak form of the Fundamental Theorem of Symmetric Polynomials. (Cf. Theorem 72.4 below.) Let F be a field and S_n act on $F[t_1, \dots, t_n]$ by restriction of the above action. Then we have with s_j the jth elementary symmetric function in $t_1, \dots, t_n$:

Corollary 57.7. *Let F be a field and f be a symmetric polynomial in $F[t_1, \dots, t_n]$. Then f is a rational function in the elementary symmetric functions in $t_1, \dots, t_n$, i.e., $f \in F(s_1, \dots, s_n)$.*

The Fundamental Theorem of Symmetric Functions [cf. Theorem 72.4 below] implies that a symmetric polynomial in $F[t_1, \dots, t_n]$ actually lies in $F[s_1, \dots, s_n]$, even with F replaced by a commutative ring. [Cf. Theorem 72.4 for the proof.]

We continue with the Examples 57.6

4. Every finite group G is isomorphic to some Galois group, i.e., there exists a Galois field extension L/K with $G \cong G(L/K)$:

   Let G be a finite group of order n and F be an arbitrary field. Set $L = F(t_1, \dots, t_n)$. If $E = F(s_1, \dots, s_n)$ as in previous example. Then L/E is Galois with $G(L/E) \cong S_n$. By Cayley's Theorem 12.5, there exists a group monomorphism $\lambda : G \to S_n$, so $G \cong \lambda(G) \subset S_n$. Let $K = L^{\lambda(G)}$, then L/K is Galois with $G(L/K) \cong G$.

Open Problem 57.8. (Inverse Galois Problem) Let H be a finite group. Does there exist a Galois extension L of $\mathbb{Q}$ with $G(L/\mathbb{Q}) \cong H$?

The answer to this question is known for some types of groups, e.g., cyclic groups, abelian groups, solvable groups, S_n, A_n.

We now turn to some significant consequences of the Fundamental Theorem of Galois Theory.

Theorem 57.9. (Primitive Element Theorem) *Let K/F be a finite extension of fields. If $\mathcal{F}(K/F)$ is finite, then there exists an element α in K such that $K = F(\alpha)$. In particular, this is true if K/F is separable.*

Proof. **Case 1**. F is finite:
K must also be a finite field, so $K^\times = \langle \alpha \rangle$ for some $\alpha \in K$. Therefore, $K = F(\alpha)$ and $\mathcal{F}(K/F)$ is finite.

Case 2. F is infinite:
Choose $\alpha \in K$ with $[F(\alpha) : F]$ maximal, so $[F(\alpha) : F] \leq [K : F]$. Suppose that $F(\alpha) < K$, then there exists a $\beta \in K \setminus F(\alpha)$. For each $a \in F$ consider the subfield $F(\alpha + a\beta) < F(\alpha, \beta) \subset K$. As $\mathcal{F}(F(\alpha, \beta)/F)$ is a subset of the finite set $\mathcal{F}(K/F)$ and F is infinite, there exists elements c and d in F with $c \neq d$ satisfying $F(\alpha + c\beta) = F(\alpha + d\beta)$. Since $c - d$ lies in $F^\times$ and $(c - d)\beta$ lies in $F(\alpha + c\beta)$, we must have $\beta \in F(\alpha + c\beta)$ and hence that α also lies in $F(\alpha + c\beta)$. It follows that $F(\alpha, \beta) \subset F(\alpha + c\beta)$, contradicting the maximality of $F(\alpha)$. $\square$

By Exercise 53.10(9b), we know this is not always true if K/F is not separable.

Remark 57.10. Let F be an infinite field and K/F a finite extension with $\mathcal{F}(K/F)$ finite. Suppose that α and β lie in K and S is an infinite subset of F. Then the proof above shows that there exists an element c in S such that $F(\alpha, \beta) = F(\alpha + c\beta)$. In particular, if $K = F(\alpha_1, \dots, \alpha_n)$ and S is an infinite subset of F, then there exist infinitely many n-tuples $c_1, \dots, c_n$ in S^n satisfying $K = F(c_1\alpha_1 + \cdots + c_n\alpha_n)$.

We finally prove the loose end in the constructibility of regular n-gons by straight-edge and compass (cf. Theorem 52.9).

Theorem 57.11. (Square Tower Theorem) *Let F be a field of characteristic different from two and K/F a finite normal field extension. Then K/F is a square root tower if and only if $[K : F] = 2^e$ for some positive integer e.*

PROOF. ($\Rightarrow$) has previously been established.

($\Leftarrow$): Suppose that $[K : F] = 2^e$, for some integer e. Let $K = F(\alpha_1, \dots, \alpha_n)$. We know that $\deg m_F(\alpha_i) = [F(\alpha_i : F] \mid [K : F] = 2^e$ for each i. As $\operatorname{char} F \neq 2$ each $m_F(\alpha_i)$ is separable, hence the extension K/F is separable by Corollary 56.16. Consequently, K/F is normal and separable, hence Galois, so $|G(K/F)| = [K : F] = 2^e$. We know that there exists a subgroup H of the Sylow 2-group $G(K/F)$ of index two. Therefore, H is a normal subgroup. It follows that K^H/F is normal of degree two. By Remark 52.8(6), K^H/F is a square root tower, as $\operatorname{char} F$ is not two. The field extension K/K^H is Galois of degree 2^{e-1}, so a square root tower by induction on e. Therefore, K/F is a square root tower by Remark 52.8(4). $\square$

Next we give a proof of the Fundamental Theorem of Algebra that uses a minimal amount of analysis. Indeed the only analysis that we need is the Intermediate Value Theorem from Calculus (which is really the completeness of the real line).

Theorem 57.12. (Fundamental Theorem of Algebra) *The field of complex numbers is algebraically closed.*

PROOF. **Claim.** $\mathbb{C} = \mathbb{C}^2 := \{x^2 \mid x \in \mathbb{C}\}$:

Let a be a positive real number. The polynomial $t^2 - a$ in $\mathbb{R}[t]$ has a real root by the Intermediate Value Theorem, since $f(1 + a) > 0$ and $f(a) < 0$. Therefore, every positive real number is a square in $\mathbb{R}$ hence in $\mathbb{C}$. Since $-a = (\sqrt{-1})^2 a$, it follows that every real number is a square in $\mathbb{C}$. Let $\alpha = a + b\sqrt{-1}$ with a and b real numbers. We must show α is a square in $\mathbb{C}$. We may assume that $b \neq 0$. Let $x = (a + \sqrt{a^2 + b^2})/2$ in $\mathbb{R}$. Then x is positive (why?), so there exists a real number c such that $x = c^2$. Similarly, $y = (-a + \sqrt{a^2 + b^2})/2$ in $\mathbb{R}$ is positive, so there exists a real number d such that $y = d^2$. Then $\alpha = (c + d\sqrt{-1})^2$, establishing the claim.
[The above arises from $\sqrt{\alpha} = \sqrt{|\alpha|}e^{\sqrt{-1}\,\theta/2}$ with θ the angle satisfying $\tan\theta = b/a$ and the half angle formula in trigonometry – which we do not need!]

Now let f be a non-constant polynomial in $\mathbb{C}[t]$. We must show that f splits over $\mathbb{C}$. Let $K/\mathbb{C}$ be a splitting field of f and L/K a normal closure of $K/\mathbb{R}$. If we show that $L = \mathbb{C}$, we are done. We know that $L/\mathbb{R}$ is Galois and $2 \mid [\mathbb{C} : \mathbb{R}]$. Therefore, there exists a Sylow 2-subgroup H of $G(K/\mathbb{R})$. In particular, $[L^H : \mathbb{R}] = [G(L/\mathbb{R}) : H]$ is odd. Since $\mathbb{R}$ is a field of characteristic zero, it is perfect, so there exists an element $\alpha \in L^H$ satisfying $L^H = \mathbb{R}(\alpha)$ by the Primitive Element Theorem 57.9. The element α in L^H is a root of the irreducible polynomial $m_{\mathbb{R}}(\alpha)$ in $\mathbb{R}[t]$ of odd degree. By the Intermediate Value Theorem, any real polynomial of odd degree has a real root. Therefore, α is real, so $L^H = \mathbb{R}(\alpha) = \mathbb{R}$, Therefore, $G(L/\mathbb{R}) = G(L/L^H)$ is a 2-group. As $L/\mathbb{R}$ is Galois and $\operatorname{char}\mathbb{R} \neq 2$, $L/\mathbb{R}$ is a square root tower by the Square Root Tower Theorem. It follows that $L(\mathbb{C})/\mathbb{C}$ is also a square root tower by Remark 52.8(2). By the claim, $\mathbb{C}$ has no proper square root towers over it, so $L(\mathbb{C}) = \mathbb{C}$ as needed. $\square$

Exercises 57.13.

1. Show that the lattices of subgroups and subfields in Example 57.6(2) are correct.

2. Let F be a perfect field and K/F a finite Galois extension. Show if $G(K/F) \cong S_n$, then K is the splitting field of an irreducible polynomial of degree n. Is this still true if $G(K/F) \cong A_n$?
3. Let K be a splitting field of $f \in \mathbb{Q}[t]$. Find K, $G(K/\mathbb{Q})$, and all intermediate fields if:
   (i) $f = t^4 - t^2 - 6$.
   (ii) $f = t^3 - 3$.
4. Let K be a splitting field of $t^5 - 2 \in \mathbb{Q}[t]$.
   (i) Find $G(K/\mathbb{Q})$.
   (ii) Show that there exists a group monomorphism $G(K/\mathbb{Q}) \to S_5$.
   (iii) Find all subgroups of $G(K/\mathbb{Q})$ and the corresponding fields.
5. Suppose that L/F is a finite Galois extension and $L/K/F$ an intermediate field. Show that $G(K/F) = N_{G(L/F)}\big(G(L/K)\big)/G(L/K)$, the quotient of the normalizer of $G(L/K)$ in $G(L/F)$ modulo $G(L/K)$.
6. Suppose that K/F is a finite Galois extension of fields. Let $F \subset E \subset K$ and L the smallest subfield of K containing E such that L/F is normal. Show $G(K/L) = \bigcap_{\sigma\in G(K/F)} \sigma G(K/E)\sigma^{-1}$, the core of $G(K/E)$ in $G(K/F)$.
7. Let L/F be a finite Galois extension, $L/K_i/F$ intermediate fields, and $H_i = G(K_i/F)$ for $i = 1, 2$. Show $H_1 \cap H_2 = G\big(L/K_1(K_2)\big)$ and the fixed field of the smallest group in $G(L/F)$ containing H_1 and H_2 is $K_1(K_2)$
8. Let K/F be a finite Galois extension. Suppose that p is a prime satisfying $p^r \mid [K : F]$ but $p^{r+1} \nmid [K : F]$. Show that there exist fields L_i, $1 \le i \le r$, satisfying $F \subseteq L_r < L_{r-1} < \cdots < L_1 < L_0 = K$ with L_i/L_{i+1} is normal, $[L_i : L_{i+1}] = p$ and $p \nmid [L_r : F]$ for each i.
9. Let f be an irreducible quartic over a field F of characteristic zero, G the Galois group of f (i.e., the Galois group of K/F with K a splitting field of f over F), u a root of f. Show that there is no field properly between F and $F(u)$ if and only if $G = A_4$ or $G = S_4$ and that there exist such irreducible polynomials with Galois group A_4 and S_4. [This will explode the myth that there must be an intermediate field when the dimension is not prime.]
10. Let L/F and K/F be finite extensions of fields with L/F Galois and K, L lying in some extension field F. Let $LK = K(L) = L(K)$. Show that LK/K and $L/L \cap K$ are Galois and the restriction map $\varphi : G(LK/K) \to G(L/F)$ given by $\sigma \mapsto \sigma|_L$ a monomorphism with image $G(L/L \cap K)$. In particular, if $L \cap K = F$, then $G(LK/K) \cong G(L/F)$.
11. Suppose that M/F be a finite Galois extension and K and L two intermediate fields of M/F with K/F Galois. Let $KL = L(K)$, $N = G(M/K) \triangleleft G(M/F)$, and $H = G(M/L)$. Then KL/L is Galois by Exercise 10. Show, in addition, that all of the following are true:
    (a) $G(M/KL) = H \cap N \triangleleft H$.
    (b) $G(KL/L) \cong H/H \cap N \cong HN/N$.
    (c) $G(K/K \cap L) \cong HN/N$.
12. Let E/F be an extension of fields and $E/K_i/F$ be intermediate fields, $i = 1, 2$ with K_i/F finite Galois for $i = 1, 2$. Let $K_1K_2 = K_1(K_2)$. Show K_1K_2/F is a Galois

extension and the restriction map $\varphi : G(K_1K_2/F) \to G(K_1/F) \times G(K_2/F)$ given by $\sigma \mapsto (\sigma|_{K_1}, \sigma|_{K_2})$ is a monomorphism. In particular, it is an isomorphism if $K_1 \cap K_2 = F$.

13. Let L/F be a finite Galois extension. Suppose that $G(L/F) = G_1 \times \cdots \times G_n$, a direct product of subgroups. Let K_i be the fixed field of $G_1 \times \cdots \times \{1\} \times \cdots \times G_n$ where the group $\{1\}$ occurs in the ith place. Show that K_i/F is Galois for every i, $K_i \cap F(\cup_{j\neq i} K_j) = F$, and $L = F(\cup_i K_i)$.
14. Prove Remark 57.10.
15. Let F be a field having no nontrivial field extensions of odd degree and K/F a finite field extension. Show if K has no field extensions of degree two, then F is perfect and K is algebraically closed. (Cf. Exercise 53.10(11)).

58. Addendum: Infinite Galois Theory

In this section, we indicate how Galois Theory for finite extensions extends to arbitrary algebraic extensions. In particular, normal extensions need not be finite and an algebraic extension K of F is called *Galois* if $F = K^{G(K/F)}$. We must generalize some of our proofs. The first is a proof of Exercise 56.22(2). The key is to reduce our proofs to finite Galois extensions. We can do this as we often deal with finitely many algebraic elements if K/F is a Galois extension.

Definition 58.1. Let L/F be a field extension and $L/E_i/F$ intermediate fields with $i \in I$. The *compositum* K of the E_i, $i \in I$ is the smallest intermediate field $L/K/F$ containing all E_i, $i \in I$, i.e., $K = F(\cup_I E_i)$. When we talk about compositums, we shall always mean all fields that are mentioned lie in some larger field L. If we are restricting ourselves to algebraic extensions, we can always assume that $L = \tilde{F}$, an algebraic closure of F. We usually do without comment.

In particular, fixing such an algebraic closure $\tilde{F}$ of F, we would like to take compositums of finite extensions of F. For example the algebraic closure of F is a compositum of the splitting field of all polynomials in $F[t]$.

Proposition 58.2. *Let K/F be an algebraic extension and $\sigma : K \to K$ an F-homomorphism, then σ lies in $G(K/F)$, i.e., σ is onto.*

Proof. Let α lie in K. Define E to be the intermediate field $K/E/F$ obtained by adjoining all the roots of $m_F(\alpha)$ in K to F. Therefore E/F is a finite extension. Since the F-homomorphism σ must permute the roots of $m_F(\alpha)$ in K, we must have $\sigma(E) \subset E$. As σ is a injective linear operator on E and E is a finite dimensional F-vector space, σ must be onto. So there exists a $\beta \in E$, such that $\sigma(\beta) = \sigma|_E(\beta) = \alpha$ as needed. $\square$

Proposition 58.3. *Let K/F be an algebraic extension. If K/F is normal and $K/E/F$ is an intermediate field, then any F-homomorphism $\sigma : E \to K$ lifts to an element of $G(K/F)$.*

Proof. This is Exercise 56.22(3), which follows by a Zorn Lemma argument and the finite extension case (and the previous proposition). [Alternatively, one can descend from an algebraic closure of F.] $\square$

Therefore, we can generalize our characterization of finite normal extensions (proving Exercise 56.22(4)). [As noted above, we assume that all algebraic extensions of the field F lie in a fixed algebraic closure of F.]

Corollary 58.4. *Let K/F be an algebraic extension. Then the following are equivalent:*

(1) K/F is normal.

(2) If $f \in F[t]$ is irreducible and has a root in K, then f splits over K.

(3) K is the compositum of all the splitting fields of $m_F(\alpha)$ with $\alpha \in K$.

(4) Let L/F be an algebraic extension with L/F normal and $L/K/F$. If $\sigma : K \to L$ is an F-homomorphism, then σ defines an element in $G(K/F)$.

In particular, any normal extension of F is a compositum of finite normal extensions of F.

Proof. (1) $\Rightarrow$ (2): As K/F is normal, let K be a splitting field of X over F, $X = \{f_i \mid i \in I\}$, and $f \in F[t]$ irreducible with a root α in K. We must show that f splits over K. Since α is algebraic over F, there exist a finite subset $Y \subset X$, such that α lies in a splitting field E of Y over F with $K/E/F$. Then $F(Y)/F$ is finite normal with $\alpha \in E$. Since E/F is finite normal, f splits over E hence over K.

(2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (1) are immediate.

(2) $\Rightarrow$ (4): Let $\sigma : K \to L$ be an F-homomorphism and $f \in F[t]$ irreducible with f having a root α in K. If α lies in K, then $m_F(\alpha)$ splits over K by assumption, i.e., K contains all the roots of $m_F(\alpha)$. Hence $\sigma(\alpha)$ lies in K. Therefore, $\sigma(K) = K$. By Proposition 58.2, we must have σ lies in $G(K/F)$.

(4) $\Rightarrow$ (2): Let $L/K/F$ with $L/F/K$ normal and $f \in F[t]$ an irreducible polynomial. Let α and β be two roots of f in L. There exists an F-homomorphism $\sigma : F(\alpha) \to \mathrm{F}(\beta)$ such that $\alpha \mapsto \beta$. This map lifts to $\widehat{\sigma} : L \to L$ by Proposition 58.3. Since $\widehat{\sigma}|_K : K \to L$ is an F-homomorphism, by (4) we have $\widehat{\sigma}|_K \in G(K/F)$. In particular, if $\alpha \in K$, then $\widehat{\sigma}|_K(\alpha) = \beta$. It follows that $\beta \in K$, i.e., f splits over K. $\square$

We can now show that the field theoretic characterization of finite Galois extensions extends to the algebraic Galois case.

Theorem 58.5. *Let K/F be an algebraic extension. If K/F is normal, then K/F is Galois if and only if K/F is separable and normal.*

Proof. ($\Leftarrow$): Let α lie in $K^{G(K/F)}$. As above there exists an intermediate field $K/E/F$ with E a splitting field of $m_F(\alpha)$ over F. As K/F is separable, so is E/F, hence E/F is finite normal and separable so Galois. Let σ lie in $G(E/F)$, By the Proposition 58.3, there exists an extension $\widehat{\sigma}$ of σ in $G(K/F)$, i.e., $\widehat{\sigma}|_E = \sigma$. So by assumption $\sigma(\alpha) = \widehat{\sigma}(\alpha) = \alpha$. Therefore, α lies in $E^{G(E/F)} = F$.

($\Rightarrow$): Let α lie in F. Every element of $S = \{\sigma(\alpha) \mid \sigma \in G(K/F)\}$ is a root of $m_F(\alpha)$ (which a priori may have other and/or multiple roots), so S is a finite set. Let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_n$ be the finite distinct elements of S and $f = \prod_{i=1}^n (t - \alpha_i)$ in $K[t]$. As $G(K/F)$ permutes $\alpha_1, \ldots, \alpha_n$, we must have $f = \widetilde{\sigma} f (:= \prod_{i=1}^n (t - \sigma(\alpha_i))$ for every $\sigma \in G(K/F)$, i.e., f lies in $F[t]$. But then $f \mid m_f(\alpha)$ in $F[t]$, so $f = m_F(\alpha)$ is separable and $F = m_F(\alpha)$ splits over K. It follows that K/F is separable and normal. $\square$

We turn to the group theoretic characterization of Galois extensions in the arbitrary algebraic case. Here we have a problem that will force us to modify our Galois Correspondence in the finite case.

Remark 58.6. Let K/F be an algebraic Galois extension of fields, H a subgroup of $G(K/F)$, and $E = K^H$. Suppose that σ and τ are elements of $G(K/F)$. If $\tau^{-1}\sigma$ lies in H, then for every $x \in E$, we have $\tau^{-1}\sigma(x) = x$, so $\sigma|_E = \tau|_E$. However, if we only know that $(\tau^{-1}\sigma)|_E = 1_E$, we can only conclude that $\tau^{-1}\sigma$ lies in $G(K/F)$. It does not necessarily follow that $\tau^{-1}\sigma$ lies in H. Therefore, we only know that $H \subset G(K/E) = G(K/K^H)$. In the case that K/F is a finite extension, we get equality by counting using Artin's Theorem, but this does not work in the arbitrary algebraic case.

We do, however, have the following:

Lemma 58.7. *Let K/F be an algebraic and Galois extension of fields and $K/E/F$ an intermediate field. If $[E : F]$ is finite, then the index $[G(K/F) : G(K/E)]$ is finite and $[G(K/F) : G(K/E)] = [E : F]$.*

Proof. Suppose that $\sigma_1, \dots, \sigma_n : E \to K$ are all the distinct F-homomorphisms. These lift to $\widehat{\sigma}_1, \dots, \widehat{\sigma}_n$ in $G(K/F)$, i.e., $\widehat{\sigma}_i|_E = \sigma_i$ for each i, by Proposition 58.3 and, as E/F is finite separable, we have $n = [E : F]$ by Proposition 56.11. Suppose that $\widehat{\sigma}$ is an element of $G(K/F)$. Then there exists an i such that $\widehat{\sigma}|_E = \sigma_i$. Therefore, $(\widehat{\sigma}_i)^{-1}\widehat{\sigma}$ lies in $G(K/E)$. It follows that

$$G(K/F) = \bigvee_{i=1}^{n} \widehat{\sigma}_i G(K/E), \tag{58.8}$$

and the result follows. □

The major new result that we need about infinite algebraic Galois extensions involves some knowledge of point set topology, much of which the reader may have seen. For the convenience of the reader, we summarize some of the concepts that we need.

A *topology* $\mathcal{T}$ on a space X is a collection of sets, called *open sets* closed under arbitrary unions and finite intersections. X is then called a *topological space* (via $\mathcal{T}$). Let X have a topology $\mathcal{T}$. A *closed set* in X is one that is the complement of an open set in $\mathcal{T}$. If x is an element in X, an open set containing x is called an *(open) neighborhood* of x, e.g., if $X = \mathbb{R}$, an open interval containing x. A subcollection of $\mathcal{T}$ consisting of neighborhoods of x such that every neighborhood of x contains an element of this subcollection is called a *fundamental system of neighborhoods* or a *base* for x, e.g., if $X = \mathbb{R}$, all the open intervals with rational end points containing x forms a base for X. A collection of all the fundamental neighborhoods of all elements of X together with the empty set (the empty intersection of fundamental neighborhoods) is called a *base* for the topology of X. In particular, $\mathcal{B}$ is a base for the topology if and only if every open set is a union of elements in the base. (One says that a topological space X is the *coarsest* topology generated by $\mathcal{B}$.) A collection of open sets $\mathcal{S}$ in X is called a *subbase* for X if all finite intersections of elements in $\mathcal{S}$ together with the empty set form a base for X.

Let X be a topological space via $\mathcal{T}$. X is called a *Hausdorff* space if given any two points $x, y \in X$ there exist disjoint open sets U and V in X with $x \in U$ and

$y \in V$. Let S be a subset of X. S becomes a *topological subspace* via the *induced topology* $\mathcal{T}_S := \{U \cap S \mid U \in \mathcal{T}\}$. S is called *compact* if for any $\{U_i\}_I$, an *open cover* of the topological subspace S, i.e., $S \subset \bigcup_I U_i$ with each U_i an open set in $\mathcal{T}_S$, there exists a subcollection $\{U_i\}_J \subset \{U_i\}_I$ with J finite and $\{U_i\}_J$ an open cover of S. We say every open cover has a finite subcover. If X is itself a compact set, then every closed subset S of X is compact; for if $\{U_i \cap S\}_I$, $U_i \in \mathcal{T}$, is an open cover of the topological subspace S, then $\{U_i\}_I \cup (X \setminus S)$ is an open cover of X. A subset S of X is called *connected* if S is not the disjoint union of open sets in $\mathcal{T}_S$ and a *connected component* of X if it is a maximal connected set. The space X is called *totally disconnected* if the connected components of X are precisely the points of X.

A map $f : X \to Y$ of topological spaces is called *continuous* if the preimage of an open set in Y under f is open in X.

We shall also need Tychonoff's Theorem which we do not prove. Let $\{X_i\}_I$ be a collection of topological spaces and $\pi_j : \prod_I X_i \to X_j$ by $(x_i)_I \mapsto x_j$ for each $j \in I$. Then the set of preimages $\{(\pi_j)^{-1}(U) \mid U \in \mathcal{T}_{X_j},\ j \in I\}$ forms a subbase for $\prod_I X_i$. The resulting topology is called the *product topology* for $\prod_I X_i$. [Note for this to make sense, we need to assume the Axiom of Choice.] Tychonff's Theorem says that a product of compact spaces is compact. [Tychonoff's Theorem is, in fact, equivalent to the Axiom of Choice.]

We now state and prove the theorem that we are after.

Theorem 58.9. (Krull) *Suppose that K/F is an algebraic Galois extension of fields. Let*

$$\mathcal{N}(K/F) := \{G(K/E) \mid K/E/F \textit{ with } E/F \textit{ finite Galois}\}.$$

Then there exists a topology on G compatible with the group structure of G (i.e., the group operation on G is continuous relative to this topology) *and which has $\mathcal{N}(K/F)$ as a fundamental system of neighborhoods of the identity 1_K in $G(K/F)$. With this topology, $G(K/F)$ is a Hausdorff, totally disconnected, topological group.*

The topology in Krull's Theorem is called the Krull Topology on $G(K/F0$. It is an example of a *profinite topology*.

Proof. Let

$$m : G(K/F) \times G(K/F) \to G(K/F) \text{ by } (\sigma,\tau) \mapsto \sigma\tau \text{ and}$$
$$i : G(K/F) \to G(K/F) \text{ by } \sigma \mapsto \sigma^{-1}$$

define the group operations. If $\sigma, \tau \in G(K/F)$, we have $m^{-1}(\sigma\tau) = \sigma\tau\mathcal{N}(K/F)$, which contains the open neighborhood $\sigma\mathcal{N}(K/F) \times \tau\mathcal{N}(K/F)$ and $i^{-1}(\sigma^{-1}) = \sigma\mathcal{N}(K/F)$, which contains the open neighborhood $\sigma\mathcal{N}(K/F)$. It follows that this induces a topological group structure on $G(K/F)$ with a base of open sets $\{\sigma\mathcal{N}(K/F) \mid \sigma \in G(K/F)\}$ by translation as m and i are continuous in this topology. This shows that $G(K/F)$ is a topological group in this topology. Moreover, every element in this basis is both open and closed called a *clopen* set by Equation 58.8.

If K/F is a finite Galois extension, then $G(K/F)$ has the discrete topology, i.e., all subsets are clopen. As $G(K/F)$ is finite, it is a compact group. The result then follows in this case. So we may assume that K/F is infinite. We set up some notation for the rest of the proof as follows:

$$G := G(K/F).$$
$$\mathcal{F}_{fg} := \{E \mid K/E/F \text{ with } E/F \text{ finite Galois}\}.$$
$$\mathcal{G}_f := \{G(E/F) \mid E \in \mathcal{F}_{fg}\}.$$

So we have $K = \bigcup_{\mathcal{F}_{fg}} E$. Suppose that $\sigma, \tau \in G$ satisfy $\sigma \neq \tau$. Then there exists an $E \in \mathcal{F}_{fg}$ satisfying $\sigma G(K/E) \neq \tau G(K/E)$. Therefore, $\sigma G(K/E) \cap \tau G(K/E) = \emptyset$ by Equation 58.8. As elements of the base are clopen, G is Hausdorff and G is totally disconnected. To prove that G is compact, we use the group homomorphism

$$\Phi : G \to \prod_{\mathcal{G}_f} G(E/F) \text{ by } \sigma \mapsto \prod_{\mathcal{F}_{fg}} \sigma|_E.$$

We first show that this map is a homeomorphism. As $\prod_{\mathcal{G}_f} G(E/F)$ is a product of compact sets, it is compact by Tychonoff's Theorem. Since $\Phi(\sigma) = 1$ if and only if $\sigma_E = 1_E$ for all $E \in \mathcal{F}_{fg}$, we have Φ is injective. For each $E_0 \in \mathcal{F}_{fg}$ and $\overline{\sigma} \in G(E_0/F)$, let

$$U_{E_0,\overline{\sigma}} = \prod_{\substack{\mathcal{F}_{fg} \\ E \neq E_0}} (G(E/F) \times \{\overline{\sigma}\}.$$

Then $\{U_{E_0,\overline{\sigma}} \mid E_0 \in \mathcal{F}_{fg},\ \overline{\sigma} \in G(E_0/F)\}$ forms a subbase for the topology of $\prod_{\mathcal{G}_f} G(E/F)$. Let $\overline{\sigma} \in G(E_0/F)$ with $E_0 \in \mathcal{F}_{fg}$ and $\sigma \in \Phi^{-1}(U_{E_0,\overline{\sigma}})$. Then $\Phi^{-1}(U_{E_0,\overline{\sigma}}) = \sigma\big(G(K/E_0)$. Therefore, Φ is continuous. Moreover, $\Phi\big(\sigma G(K/F)\big) = \Phi\big(G(K/F)\big) \cap U_{E_0,\overline{\sigma}}$. Therefore, Φ is also an open map, hence it is a homeomorphism.

We now show $G(K/F)$ is compact. To do so, it suffices to show that $\operatorname{im}\Phi$ is closed in the compact space $\prod_{fg} G(E/F)$. Suppose that $E_1, E_2, \in \mathcal{F}_{fg}$ with E_2/E_1. Set

$$I_{E_2/E_1} := \{\prod_{\mathcal{F}_{fg}} \sigma_E \in \prod_{\mathcal{F}_{fg}} G(E/F) \mid \sigma_{E_2}|_{E_1} = \sigma_{E_1}\}.$$

Then we have $\Phi(G) = \bigcap_{\substack{\mathcal{F}_{fg} \\ E_2/E_1}} I_{E_2/E_1}$. So it suffices to show that each finite set I_{E_2/E_1} is closed. Let $G(E_1/F) = \{\sigma_1, \dots, \sigma_n\}$ and $J_i = \{\sigma_i \in G(E_1/F) \mid \sigma_i \text{ extends to } E_2\}$. Then

$$I_{E_2/E_1} = \bigcup_{i=1}^{n} \big(\prod_{E \neq E_1, E_2} G(E/F) \times J_i \times \{\sigma_i\}\big).$$

It follows that I_{E_2/E_1} is closed. Hence G is compact. $\square$

Let K/F be a Galois extension and H a subgroup of $G(K/F)$. Denote by $\overline{H}$ the *closure* of H in $G(K/F)$ in this topology, i.e., the smallest closed set in $G(K/F)$ containing H. This means that if σ in $G(K/F)$ lies in $\overline{H}$, then every open neighborhood of σ intersects H nontrivially. Because of the topology on $G(K/F)$, this means that $H \cap \sigma H'$ is nonempty for every H' in $\mathcal{N}(K/F)$.

The key to obtaining a Galois Correspondence generalizing that in the finite Galois case is the following:

Lemma 58.10. *Let K/F be an algebraic Galois extension of fields, H a subgroup of $G(K/F)$, and $E = K^H$. Then $G(K/E) = \overline{H}$, the closure of H in $G(K/F)$.*

Proof. $G(K/E) \subset \overline{H}$: Let σ lie in $G(K/E)$. We must show that $\sigma \in \overline{H}$. As remarked above, since $\sigma\mathcal{N}(K/F)$ is a fundamental system of neighborhoods for σ, it suffices to show that $H \cap \sigma H'$ is not empty for all H' in $\mathcal{N}(K/F)$. Suppose that $H' \in \mathcal{N}(K/F)$. As $K^{H'}/F$ is finite Galois, it is finite separable, so $K^{H'} = F(\alpha)$ for some α in K by the Primitive Element Theorem 57.9. Let $K/L/E$ be an intermediate field such that L/E is finite normal with $\alpha \in L$. Since L/E is finite, normal, and separable, it is finite Galois. As $H \subset G(K/E)$ and L/E is Galois, $\tau|_E$ lies in $G(K/E)$ for all $\tau \in H$, i.e, we have a group homomorphism $\Psi : H \to G(L/E)$ given by $\tau \mapsto \tau|_L$ and it has $G(L/L^{\operatorname{im}\Psi})$ as (its finite) image. By the definition of H, we have $E = L^{\operatorname{im}\Psi}$, so Ψ is onto. [Note, a priori, we only know that $E \subset K^{G(K/E)}$.] Thus there exists a τ in H with $\tau|_L = \sigma|_L$. In particular, as α lies in L, we have $\sigma(\alpha) = \tau(\alpha)$, so $\sigma^{-1}\tau$ lies in H'. This shows that $\tau \in \sigma H' \cap H$.

$\overline{H} \subset G(K/E)$: Let σ lie in $\overline{H}$. We must show that σ lies in $G(K/E)$, i.e., if α is an element of E, then $\sigma(\alpha) = \alpha$. Let $K/L/F(\alpha)$ with L/F finite, normal, and separable, hence finite Galois. Set $H' = G(K/L)$. By assumption, $H \cap \sigma H'$ is not empty, so there exists an element τ in $H \cap \sigma H'$. In particular, $\tau|_E = 1_E$, since $E = K^H$, and $\sigma(\alpha) = \tau(\alpha)$, since $F(\alpha) \subset K^{H'}$. Therefore, $\sigma(\alpha) = \tau(\alpha) = \alpha$. □

The Fundamental Theorem of Galois Theory in this more general setting becomes:

Theorem 58.11. *Let K/F be an algebraic Galois extension of fields and*

$$\mathcal{G}_c(K/F) := \{H \mid H \text{ is a closed subgroup of } G(K/F)\},$$

then

$$i : \mathcal{G}_c(K/F) \to \mathcal{F}(K/F) \text{ given by } H \mapsto K^H$$

is an order reversing bijection.

Proof. By the lemma, we know that

$$\mathcal{G}_c(K/F) = \{G(K/E) \mid E \in \mathcal{F}(K/F)\}.$$

i **is injective**: If $K^{H_1} = K^{H_2}$, then by the lemma,

$$H_1 = \overline{H_1} = G(K/K^{H_1}) = G(K/^{H_2}) = \overline{H_2} = H_2.$$

i **is surjective**: Let E lie in $\mathcal{F}(K/F)$, so $E \subset K^{G(K/E)}$. We must show that $G(K/E)$ moves $K \setminus E$. Let $\alpha \in K \setminus E$. As before, let $K/L/E$ with L/E finite normal, hence Galois, and $\alpha \in L$. As α does not lie in E, there exist an element σ in $G(L/E)$ satisfying $\sigma(\alpha) \neq \alpha$. By Proposition 58.3, there exists a lift $\widehat{\sigma}$ in $G(K/F)$ of σ. So $\widehat{\sigma}(\alpha) = \sigma(\alpha) \neq \alpha$. □

Remark 58.12. Let K/F be an algebraic Galois extension of fields, $E \in \mathcal{F}(K/F)$. Then

(i) E/F is normal if and only if $G(K/E) \triangleleft G(K/F)$.

(ii) If E/F is normal, then $G(E/F) \cong G(K/F)/G(K/E)$.

We leave the proofs of these as exercises.

We give an example to show that the cardinalities of $\mathcal{G}(K/F)$ and $\mathcal{G}_c(K/F$, when K/F is Galois can be different, showing the extent of the reduction if the groups in which the Galois Correspondence holds.

Example 58.13. Let $S = \{\sqrt{a} \mid a > 0 \text{ a prime or } a = -1\} \subset \mathbb{C}$ and $K = \mathbb{Q}(S)$ the splitting field of $\{t - p^2 \mid p > 0 \text{ a prime}\} \cup \{t^2 + 1\}$ over $\mathbb{Q}$ in $\mathbb{C}$. Therefore, $K/\mathbb{Q}$ is Galois. As $\mathbb{Z}$ is a UFD, it is easy to see that K is the *compositum* of all quadratic extensions of $\mathbb{Q}$ in $\mathbb{C}$, i.e., $K = \mathbb{Q}\big(\bigcup_{d \in \mathbb{Z}} \mathbb{Q}(\sqrt{d})\big)$, the number of which is countable. If σ and τ lie in $G(K/\mathbb{Q})$, then we have $\sigma^2 = 1_K$ and $\sigma\tau = \tau\sigma$, i.e., $G(K/\mathbb{Q})$ is abelian and all non-identity elements are of order two. It follows that $G(K/\mathbb{Q})$ is a vector space over $\mathbb{Z}/2\mathbb{Z}$, hence has a $\mathbb{Z}/2\mathbb{Z}$-basis by Proposition 28.6, say $\mathcal{B}$.
Let T be a subset of S and define $\sigma : S \to S$ by

$$\sigma(x) = \begin{cases} x & \text{if } x \in T \\ -x & \text{if } x \notin T. \end{cases} j$$

Then σ induces an element in $G(K/\mathbb{Q})$. In particular, $G(K/\mathbb{Q})$ is uncountable, so cannot have a countable $\mathbb{Z}/2\mathbb{Z}$-basis. Therefore, for all but countably many $x \in \mathcal{B}$, the subgroup $H_x := \langle \mathcal{B} \cup \{x\} \rangle$ is a normal subgroup of index two in $G(K/\mathbb{Q})$ that does cannot correspond to a quadratic extension of $\mathbb{Q}$, i.e., only countable many H_x are closed (or open) in $G(K/\mathbb{Q})$.

Exercises 58.14.

1. Show the two items in Remark 58.12 are true.
2. Here is a way to prove Tychonoff's Theorem:
   (i) Let X be a topological space with a base $\mathcal{B}$. Show that X is compact if and only if every cover of X by elements of $\mathcal{B}$ has a finite subcover.
   (ii) Let X be a topological space with a subbase $\mathcal{B}$. Show that X is compact if and only if every cover of X by elements of $\mathcal{B}$ has a finite subcover.
   (iii) Prove Tychonoff's Theorem.

59. Roots of Unity

Recall that an *arithmetic function* $f : \mathbb{Z}^+ \to \mathbb{C}$ is called a *multiplicative function* if $f(1)$ is nonzero and $f(mn) = f(m)f(n)$, whenever m and n are relatively prime. For example, the Euler phi-function is multiplicative (by the Chinese Remainder Theorem). Recall also that if f is a multiplicative function, then $f(1) = 1$. Let $I : \mathbb{Z}^+ \to \mathbb{C}$ be the Identity multiplicative function defined by $I(n) = 1$ and $I(n) = [\frac{1}{n}] = 0$ for all $n > 1$. The set $\mathcal{A} := \{f \mid f : \mathbb{Z}^+ \to \mathbb{C} \text{ with } f(1) \neq 0\}$ is an abelian monoid with unity I under the *Dirichlet product* $\star : \mathcal{A} \times \mathcal{A} \to \mathbb{C}$ defined by

$$(f \star g)(n) := \sum_{d \mid n} f(d) g(\frac{n}{d})$$

(cf. Exercise 59.22(1)). This group arises from the study of Dirichlet series, i.e., absolutely convergent series of the form $\displaystyle\sum_n \frac{f(n)}{n^s}$, f a multiplicative function and s a complex variable. It can be shown that $\mathcal{M} := \{f \mid f \in \mathcal{A} \text{ multiplicative}\}$ is a subgroup. Basic to this is Möbius inversion that we now investigate. Recall

Definition 59.1. The *Möbius function* $\mu : \mathbb{Z}^+ \to \mathbb{C}$ is defined by $\mu(1) = 1$ and if $n = p_1^{e_1} \cdots p_r^{e_r}$ is a standard factorization of $n > 1$, then

$$\mu(n) := \begin{cases} (-1)^r & \text{if } n \text{ is square-free, i.e., } e_1 = \cdots = e_r = 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $U : \mathbb{Z}^+ \to \mathbb{C}$ be the multiplicative function defined by $U(n) = 1$ for all n.

Lemma 59.2. *The Möbius function μ is multiplicative and satisfies:*

(1) $I = \mu \star U$ i.e.,

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

(2) If $n = p_1^{e_1} \cdots p_r^{e_r}$ is a standard factorization of $n > 1$, then $\sum_{d|n} |\mu(d)| = 2^r$.

PROOF. Let m and n be relatively prime positive integers. Clearly, either m or n is not square free if and only if mn is not square free. It follows easily that μ is multiplicative. Let $\epsilon = \mu$ or $|\mu|$. If $n > 1$ is not square-free, then $\epsilon(n) = 0$. In particular, we may assume that $n = p_1 \cdots p_r$ is a standard factorization of $n > 1$. Then

$$\begin{aligned} \sum_{d|n} \epsilon(d) = \epsilon(1) + \sum_i \epsilon(p_i) + \sum_{i_1 < i_2} \epsilon(p_{i_1} p_{i_2}) + \\ \cdots + \sum_{i_1 < \cdots < i_j} \epsilon(p_{i_1} \cdots p_{i_j}) + \cdots + \epsilon(p_1 \cdots p_r) \\ = 1 + \epsilon(p_1)\binom{r}{1} + \epsilon(p_1 p_2)\binom{r}{2} + \cdots + \epsilon(p_1 \cdots p_r)\binom{r}{r}. \end{aligned}$$

If $\epsilon = \mu$, this is $1 - \binom{r}{1} + \binom{r}{2} + \cdots + (-1)^r\binom{r}{r} = (1-1)^r = 0$ and if $\epsilon = |\mu|$, this is $1 + \binom{r}{1} + \binom{r}{2} + \cdots + +\binom{r}{r} = (1+1)^r = 2^r$. □

Proposition 59.3. (Möbius Inversion Formula) *Let f and g be arithmetic functions not zero at* 1. *Then*

$$f(n) = \sum_{d|n} g(d) \text{ if and only if } g(n) = \sum_{d|n} f(d)\mu(\frac{n}{d}).$$

PROOF. The proposition is equivalent to $f = g \star U$ if and only if $g = f \star \mu$. By Exercise 8.5(15), the Dirichlet product is associative with I a unity, so $f = g \star U$ if and only if $f \star \mu = (g \star U) \star \mu = g \star (U \star \mu) = g \star I = g$. □

For each n, let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$. Define the nth *cyclotomic polynomial* $\Phi_n(t)$ in $\mathbb{C}[t]$ by $\Phi_n = \prod_{\langle \zeta_n \rangle}(t - \zeta)$ in $\mathbb{C}[t]$, a polynomial of degree $\varphi(n)$. Clearly,

$$t^n - 1 = \prod_{d|n} \Phi_d,$$

so $\log(t^n - 1) = \sum_{d|n} \log \Phi_d$. By the Möbius Inversion Formula (viewing the cyclotomic polynomials as polynomial functions on the integers or use Exercise 59.22(6)), we see that

$$\Phi_n = \prod_{d|n} (t^d - 1)^{\mu(\frac{n}{d})} = \prod_{d|n} (t^{\frac{n}{d}} - 1)^{\mu(d)} \text{ lies in } \mathbb{C}[t] \cap \mathbb{Q}(t) = \mathbb{Q}[t]$$

and is monic. By induction Φ_d lies in $\mathbb{Z}[t]$ for $d < n$, hence so does $\Phi(n)$ by Corollary 35.6 (the corollary to Gauss's Lemma).

Example 59.4. Using the formula above, we have

$$\begin{aligned}\Phi_{12} &= \prod_{d|12} (t^{\frac{12}{d}} - 1)^{\mu(d)} \\ &= (t^{12} - 1)^{\mu(1)}(t^6 - 1)^{\mu(2)}(t^4 - 1)^{\mu(3)}(t^3 - 1)^{\mu(4)} \\ &\qquad\qquad (t^2 - 1)^{\mu(6)}(t - 1)^{\mu(12)} \\ &= \frac{(t^{12} - 1)(t^2 - 1)}{(t^6 - 1)(t^4 - 1)} = t^4 - t^2 + 1.\end{aligned}$$

Definition 59.5. Let K/F be a finite Galois extension of fields. We say that K/F is an *abelian* (respectively, *cyclic*) *extension* if $G(K/F)$ is abelian (respectively, cyclic).

Theorem 59.6. *Let ζ be a primitive n root of unity in $\mathbb{C}$. Then $\mathbb{Q}(\zeta)/\mathbb{Q}$ is an abelian extension of degree $\varphi(n)$.*

Proof. We must show that Φ_n is irreducible in $\mathbb{Q}[t]$, equivalently in $\mathbb{Z}[t]$. Let ζ be a primitive nth root of unity. As $\mathbb{Z}[t]$ is a UFD, we see that there exist polynomials f and g in $\mathbb{Z}[t]$ satisfying $\Phi_n = fg$ in $\mathbb{Z}[t]$ with f irreducible in $\mathbb{Z}[t]$ and having ζ as a root. As Φ_n is monic, we may assume that both f and g are monic. [Note that this implies that $f = m_{\mathbb{Q}}(\zeta)$.] Let p be a prime such that $p \not| n$.

Claim: ζ^p is a root of f.

If we prove the claim, it would follow that ζ^m, with m relatively prime to n, is also root of f, hence $f = \Phi_n$ establishing the theorem.

Suppose the claim is false. We can write $t^n - 1 = fh$ in $\mathbb{Z}[t]$ with $h \in \mathbb{Z}[t]$ monic by Corollary 35.6 (the corollary to Gauss's Lemma). As ζ^p is not a root of f, it must be a root of h. This implies that ζ is a root of the monic polynomial $h_p := h(t^p)$ in $\mathbb{Z}[t]$. It follows that $h_p = fg$ in $\mathbb{Z}[t]$ for some monic g in $\mathbb{Z}[t]$ by Corollary 35.6. Let $^{-} : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism; and, if f lies in $\mathbb{Z}[t]$, let $\overline{f}$ be the image of f in $(\mathbb{Z}/p\mathbb{Z})[t]$. Applying the Frobenius homomorphism shows that

$$\overline{h}^p = \overline{h}(t^p) = \overline{h_p} = \overline{f}\overline{g}$$

and has $\overline{\zeta}$ as a root, hence so does $\overline{h}$. Therefore, $t^n - 1 = \overline{f}\,\overline{h}$ has a multiple root over $\mathbb{Z}/p\mathbb{Z}$. As $p \not| n$, this is impossible. This establishes the claim and the theorem. □

Remark 59.7. Let p be an odd prime and ζ a primitive p^rth root of unity. Then $G\big(\mathbb{Q}(\zeta)/\mathbb{Q}\big) \cong (\mathbb{Z}/p^r\mathbb{Z})^\times$ is cyclic of order $p^{r-1}(p-1)$. In particular, $\mathbb{Q}(\zeta)$ contains a unique quadratic extension of $\mathbb{Q}$ using Proposition F.4.

Corollary 59.8. *Let ζ_m, ζ_n, and ζ_{mn} be primitive mth, nth, and (mn)th roots of unity over $\mathbb{Q}$ with m and n relatively prime positive integers. Then $\mathbb{Q}(\zeta_{mn}) = \mathbb{Q}(\zeta_m, \zeta_n)$ and $\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q}$.*

PROOF. As m and n are relatively prime, $\zeta_m\zeta_n$ is a primitive (mn)th root of unity. Since the Euler phi-function is multiplicative, the result follows from the theorem. □

By the results in §F, we know $G(\mathbb{Q}(\zeta_n)/\mathbb{Q})$. Indeed if $n = 2^e p_1^{e_1} \cdots p_r^{e_r}$ is a factorization with $2 < p_1 < \cdots < p_r$ primes, with $e \geq 0$ and $e_i > 0$ for all i, then

$$G(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/2^e\mathbb{Z})^\times \times (\mathbb{Z}/{p_1}^{e_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/{p_r}^{e_r}\mathbb{Z})^\times$$

by the Chinese Remainder Theorem. We know that $(\mathbb{Z}/{p_i}^{e_i}\mathbb{Z})^\times$ is cyclic of order $\varphi(p_i^{e_i}) = p_i^{e_i-1}(p_i - 1)$ for every i and if n is even $(\mathbb{Z}/2^e\mathbb{Z})^\times$ is cyclic only if $e = 1, 2$, otherwise $(\mathbb{Z}/2^e\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{e-2}\mathbb{Z})$ if $e \geq 3$.

We next wish to prove a special case of the following well-known result that uses the same idea as the proof of the above as well as our previous idea in proving the infinitude of primes.

Theorem 59.9. (Dirichlet's Theorem on Primes in an Arithmetic Progression) *Let m and n be relatively prime integers. Then there exist infinitely many primes p satisfying $p \equiv m \mod n$.*

We shall not prove this general theorem. The standard proof uses complex analysis, although there is an elementary proof, i.e., one that uses no complex analysis. The special case that we shall prove is the following:

Proposition 59.10. *Let $n > 1$ be an integer. Then there exist infinitely many primes p satisfying $p \equiv 1 \mod n$.*

To prove this, we first establish the following lemma:

Lemma 59.11. *Let a be a positive integer and p a (positive) prime not dividing n. If $p \mid \Phi_n(a)$ in $\mathbb{Z}$, then $p \equiv 1 \mod n$.*

PROOF. Let $\bar{\ } : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism. We know that $t^n - 1 = \prod_{d|n} \Phi_d$ is a factorization of $t^n - 1$ into irreducible polynomials in $\mathbb{Z}[t]$. In particular, $a^n \equiv 1 \mod p$. Let m be the order of $\overline{a}$ in $\mathbb{Z}/p\mathbb{Z}$. So $m \mid n$. If $m < n$, then

$$t^n - 1 = \prod_{d|n} \Phi_d = \Phi_n \prod_{\substack{d|n \\ d<n}} \Phi_d = \Phi_n(t^m - 1) \prod_{\substack{d|n \\ d<n \\ d\nmid m}} \Phi_d.$$

It follows that $\overline{a}$ is a multiple root of $t^n - 1$ in $\mathbb{Z}/p\mathbb{Z}$. But $p \nmid n$, so $t^n - 1$ has no multiple roots in $\mathbb{Z}/p\mathbb{Z}$. It follows that $m = n$, i.e., n is the order of $\overline{a}$, so $n \mid p - 1$ by Fermat's Little Theorem. The result follows. □

PROOF. (of the Proposition) (Cf. Exercise 1.13(4).) Suppose that the result is false, and $p_1, \ldots, p_r$ are all the primes p satisfying $p \equiv 1 \mod n$. As Φ_n is monic, we can choose an integer $N > 1$ satisfying $\Phi_n(M) > 1$ for all integers $M > N$ (by calculus). Let

$M = p_1 \cdots p_r nN > N$ and p a prime such that $p \mid \Phi_n(p_1 \cdots p_r nN) > 1$. (Replace $p_1 \cdots p_r$ by 1 if $r = 0$.) Since $t^n - 1 = \prod_{d|n} \Phi_d$ in $\mathbb{Z}[t]$, the constant term of Φ_d is ± 1, so

$$\Phi_n(p_1 \cdots p_r nN) \equiv \begin{cases} \pm 1 \mod p_i \text{ for } 1 \leq i \leq r \\ \pm 1 \mod n. \end{cases}$$

In particular, $p \neq p_1, \dots, p_r$ and $p \not| n$. It follows that $p \equiv 1 \mod n$ by the lemma, a contradiction. □

The proof of the full theorem shows that primes are "equally distributed" among the integers m relatively prime to n with $1 \leq m \leq n$.

We next turn to a special case of another important theorem whose proof we shall also omit, viz.,

Theorem 59.12. (Kronecker-Weber Theorem) *Let $K/\mathbb{Q}$ be a abelian extension of fields. Then there exists a root of unity ζ in $\mathbb{C}$ such that K is a subfield of $\mathbb{Q}(\zeta)$*

We shall prove this in the case that K is a quadratic extension of $\mathbb{Q}$. The general theorem shows that abelian extensions of $\mathbb{Q}$ are determined by the unit circle. A similar geometric interpretation is true when $\mathbb{Q}$ is replaced by an imaginary quadratic extension of $\mathbb{Q}$, i.e., by $\mathbb{Q}(\sqrt{-d})$, with $d = 1$ or $d > 1$ a square-free integer, and the circle replaced by an appropriate "elliptic curve". Unfortunately, this does not extend further, although a major triumph in number theory was the determination of abelian extensions L of K when $K/\mathbb{Q}$ is finite. There just is no geometric formulation in this general case.

If a is an integer not divisible by p, recall we defined the *Legendre symbol* in Definition 32.8 to be

$$\Big(\frac{a}{p}\Big) := \begin{cases} +1 \text{ if } a \bmod p \text{ is a square.} \\ -1 \text{ if } a \bmod p \text{ is not a square.} \end{cases}$$

It is convenient to define $\left(\frac{a}{p}\right) = 0$ if $p \mid a$. Clearly, the Legendre symbol $\left(\frac{a}{p}\right)$ only depends on $a \mod p$. Let p be any prime and consider the squaring map $f : (\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times$ given by $x \mapsto x^2$, Then f is a group homomorphism with $\ker f = \{\pm 1\}$ with image $\big((\mathbb{Z}/p\mathbb{Z})^\times\big)^2$. In particular, if p is odd, then $|\operatorname{im} f| = (p-1)/2$, i.e., half of $(\mathbb{Z}/p\mathbb{Z})^\times$ are squares and half non-squares.

Lemma 59.13. *Let p be an odd prime. Then for all integers a and b not divisible by p, we have the following:*

(1) $\Big(\dfrac{a}{p}\Big) = \Big(\dfrac{b}{p}\Big)$ *if* $a \equiv b \mod p$.

(2) $\Big(\dfrac{ab}{p}\Big) = \Big(\dfrac{a}{p}\Big)\Big(\dfrac{b}{p}\Big)$.

(3) (Euler's Criterion) *If* $p \not| a$, *then* $\Big(\dfrac{a}{p}\Big) \equiv (a)^{\frac{p-1}{2}} \mod p$. *In particular,* $\Big(\dfrac{-1}{p}\Big) = (-1)^{\frac{p-1}{2}}$.

(4) $\displaystyle\sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big) = 0$.

Proof. We have already observed (1), and (4) is the fact that half of $(\mathbb{Z}/p\mathbb{Z})^\times$ are squares and half not.

(3): We know that $x^{p-1} \equiv 1 \mod p$ if $p \not| x$, so if a is a square modulo p, then $a^{\frac{p-1}{2}} \equiv 1 \mod p$. Conversely, suppose that a is a square modulo p. As $(\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$, we have $(\mathbb{Z}/p\mathbb{Z})^\times = \langle x \rangle$ for some x. Suppose that $a \equiv x^n \mod p$. Then n must be even lest $x^{\frac{p-1}{2}} \equiv 1 \mod p$, hence a is a square modulo p. The case of $a = -1$ follows easily as $\left(\frac{-1}{p}\right) - (-1)^{\frac{p-1}{2}}$ is -2, 0, or 2. (In fact, we have previously shown this case.)

(2) follows from (3). $\square$

Proposition 59.14. *Let p be an odd prime and ζ a primitive pth root of unity. Set*

$$S := \sum_{a=1}^{p-1} \left(\frac{a}{p}\right)\zeta^a$$

in $\mathbb{Q}(\zeta)$ (even in $\mathbb{Z}[\zeta]$). Then

$$S^2 = \left(\frac{-1}{p}\right)p = (-1)^{\frac{p-1}{2}}p \text{ in } \mathbb{Q} \text{ (even in } \mathbb{Z}).$$

Moreover, $\mathbb{Q}(S)$ is the unique quadratic extension of $\mathbb{Q}$ lying in $\mathbb{Q}(\zeta)$.

The sum S in the proposition is called a *Gauss sum*.

Proof. We have

$$S^2 = \sum_{a=1,\, b=1}^{p-1} \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)\zeta^{a+b} = \sum_{a=1,\, b=1}^{p-1} \left(\frac{ab}{p}\right)\zeta^{a+b}.$$

As a ranges over $1, \dots, p-1 \mod p$, so does ab for $b = 1, \dots, p-1 \mod p$, so upon replacing a by ab, we see that

$$(*)\qquad \begin{aligned} S^2 &= \sum_{a=1,\, b=1}^{p-1} \left(\frac{ab^2}{p}\right)\zeta^{b(a+1)} = \sum_{a=1,\, b=1}^{p-1} \left(\frac{a}{p}\right)\zeta^{b(a+1)} \\ &= \sum_{b=1}^{p-1} \left(\frac{-1}{p}\right)\zeta^{bp} + \sum_{a=1}^{p-2} \left(\frac{a}{p}\right) \sum_{b=1}^{p-1} \zeta^{b(a+1)}, \end{aligned}$$

where the first term on the second line arises from the term $a = p-1$. As ζ^r is also a primitive pth root of unity for $r = 1, \dots, p-1$, we have $1 + \zeta^r + (\zeta^r)^2 + \cdots + (\zeta^r)^{p-1} = 0$ for such r. In particular, we have $\sum_{b=1}^{p-1} \zeta^{b(a+1)} = -1$ if $a \neq p-1$ in (*). Therefore, we have, using Lemma 59.13(4),

$$\begin{aligned} S^2 &= \sum_{b=1}^{p-1} \left(\frac{-1}{p}\right) - \sum_{a=1}^{p-2} \left(\frac{a}{p}\right) = (p-1)\left(\frac{-1}{p}\right) - \sum_{a=1}^{p-2} \left(\frac{a}{p}\right) \\ &= p\left(\frac{-1}{p}\right) - \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) = p\left(\frac{-1}{p}\right) \end{aligned}$$

in $\mathbb{Z}$. Consequently, as $S \subset \mathbb{Q}(\zeta)$, we have $\sqrt{p\left(\frac{-1}{p}\right)}$ lies in $\mathbb{Q}(\zeta)$. Finally as $G(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$, there exists a unique subgroup H of $G(\mathbb{Q}(\zeta)/\mathbb{Q})$ of index two. By the Fundamental Theorem of Galois Theory, we must have $\mathbb{Q}(S) = \mathbb{Q}(\zeta)^H$. □

Remark 59.15. . What is unclear is the sign of S above. This caused Gauss a lot of difficulty. The answer is

$$S = \sqrt{p} \text{ if } p \equiv 1 \mod 4$$
$$S = \sqrt{-p} \text{ if } p \equiv 3 \mod 4,$$

where we have taken the positive square root. We shall assume this. A proof will be given in Appendix G

Remark 59.16. Let p be an odd prime and ζ a primitive p^rth root of unity in $\mathbb{C}$. Then $\mathbb{Q}\left(\sqrt{(-1)^{\frac{p-1}{2}}p}\right)$ is the unique quadratic extension of $\mathbb{Q}$ in $\mathbb{Q}(\zeta)$ by Remark 59.7 and Remark 59.15.

Corollary 59.17. *Let p be a (positive) prime, ζ_p a primitive pth root of unity, and ζ_{4p} a primitive $4p$ th root of unity. Then $\sqrt{p}$ and $\sqrt{-p}$ lie in $\mathbb{Q}(\zeta_{4p})$.*

Proof. Let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$ for each n.

Case 1. p an odd prime:

If $p \equiv 1 \mod 4$, then by the proposition and Remark 59.15, we know that $\sqrt{p}$ lies in $\mathbb{Q}(\zeta_p) \subset \mathbb{Q}(\zeta_{4p})$ and if $p \equiv 3 \mod 4$, then $\sqrt{-p}$ lies in $\mathbb{Q}(\zeta_p)$. As $\sqrt{-1} = \zeta_4$ and $\mathbb{Q}(\zeta_4, \zeta_p) = \mathbb{Q}(\zeta_{4p})$, this case follows.

Case 2. $p = 2$:

Let $y = \zeta_8 + \zeta_8^{-1}$. Then ζ_8 is the point $(1 + \sqrt{-1})/2$ on the unit circle in the complex plane, i.e., the intersection with the 45^o ray with the real axis and ζ_8^{-1} is its complex conjugate. Computation shows $y^2 = 2$. It follows that $\sqrt{2}$ lies in $\mathbb{Q}(\zeta_8)$. As $\sqrt{-1}$ lies in $\mathbb{Q}(\zeta_4) \subset \mathbb{Q}(\zeta_8)$, it follows that $\sqrt{2}$ and $\sqrt{-2}$ lie in $\mathbb{Q}(\zeta_8)$. □

Theorem 59.18. *Let n be a nonzero integer. Then $\sqrt{n}$ lies in $\mathbb{Q}(\zeta_{4n})$, where ζ_{4n} is a primitive $4n$ th root of unity. In particular, any quadratic extension of $\mathbb{Q}$ lies in $\mathbb{Q}(\zeta)$ for some root of unity ζ in $\mathbb{C}$.*

Proof. Let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$ for each n. We may assume that n is square-free. We know the result if $n = -1$, so we may assume that $n = \pm p_1 \cdots p_r$ with $p_1 < \cdots < p_r$ positive primes with $r \geq 1$. Then $\sqrt{n}$ is an element of $\mathbb{Q}(\zeta_4, \zeta_{p_1}, \ldots \zeta_{p_r}) = \mathbb{Q}(\zeta_{4n})$ if n is odd and $\sqrt{n}$ is an element of $\mathbb{Q}(\zeta_8, \zeta_{p_2}, \ldots \zeta_{p_r}) = \mathbb{Q}(\zeta_{4n})$ if n is even. □

We turn to a proof of Quadratic Reciprocity, one of hundreds!

Theorem 59.19. (Law of Quadratic Reciprocity) *Let p and q be distinct odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

In addition, we have

(Supplement #1) $$\Big(\frac{-1}{p}\Big) = (-1)^{\frac{p-1}{2}}.$$

(Supplement #2) $$\Big(\frac{2}{p}\Big) = (-1)^{\frac{p^2-1}{8}}.$$

Proof. Let ζ be a primitive pth root of unity in $\mathbb{C}$ and

$$S := \sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big)\zeta^a \text{ in } \mathbb{Z}[\zeta].$$

By Proposition 59.14 and Euler's Criterion, we have

$$S^q = S(S^2)^{\frac{q-1}{2}} = S\Big(\frac{-1}{p}\Big)^{\frac{q-1}{2}} p^{\frac{q-1}{2}} = S(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}}.$$

Multiplying this equation by S, we see that

$$(*) \qquad \begin{aligned} S^{q+1} &= S^2(S^2)^{\frac{q-1}{2}} = S^2(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}} \\ &= p^{\frac{q-1}{2}}\Big(\frac{-1}{p}\Big)(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}}, \end{aligned}$$

hence is an integer. By the Binomial Theorem (actually the Multinomial Theorem), since q is an odd prime, we see upon multiplying out and coalescing terms that

$$S^q = \Big(\sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big)\zeta^a\Big)^q = \sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big)^q \zeta^{aq} + qr = \sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big)\zeta^{aq} + qr$$

for some r in $\mathbb{Z}[\zeta]$. As aq, $a = 1, \dots, p-1$, runs over all the nonzero residue classes modulo p, we have

$$\begin{aligned} S^q &= \sum_{a=1}^{p-1} \Big(\frac{q^2 a}{p}\Big)\zeta^{aq} + qr = \Big(\frac{q}{p}\Big) \sum_{a=1}^{p-1} \Big(\frac{aq}{p}\Big)\zeta^{aq} + qr \\ &= \Big(\frac{q}{p}\Big) \sum_{a=1}^{p-1} \Big(\frac{a}{p}\Big)\zeta^{a} + qr = \Big(\frac{q}{p}\Big) S + qr. \end{aligned}$$

Multiplying this by S then yields

$$S^{q+1} = \Big(\frac{q}{p}\Big)S^2 + qr = \Big(\frac{q}{p}\Big)\Big(\frac{-1}{p}\Big)p + qrS$$

in $\mathbb{Z}[\zeta]$.

Check 59.20. $q\mathbb{Z}[\zeta_p] \cap \mathbb{Z} = q\mathbb{Z}$.

So we have, using (*),

$$qrS = S^{q+1} - \Big(\frac{q}{p}\Big)\Big(\frac{-1}{p}\Big)p \text{ lies in } q\mathbb{Z},$$

which implies

$$S^{q+1} \equiv \Big(\frac{q}{p}\Big)\Big(\frac{-1}{p}\Big)p \mod q.$$

Using (*) again, now shows that

$$p\Big(\frac{-1}{p}\Big)(-1)^{\frac{p-1}{2}\frac{q-1}{2}}p^{\frac{q-1}{2}} \equiv \Big(\frac{q}{p}\Big)\Big(\frac{-1}{p}\Big)p \mod q.$$

Since $p \mod q$ is a unit in $\mathbb{Z}/q\mathbb{Z}$ and $p^{\frac{p-1}{2}} \equiv \big(\frac{p}{q}\big) \mod q$ by Euler's Criterion, we have

$$\Big(\frac{p}{q}\Big) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\Big(\frac{q}{p}\Big) \mod q.$$

Finally, as $\big(\frac{p}{q}\big) - (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\big(\frac{q}{p}\big)$ is $-2, 0$, or 2, the result follows.

Since we have already proven Supplements #1, to finish we need only establish Supplement #2. We work in $\mathbb{Z}/p\mathbb{Z}$ with p an odd prime. Let ζ be a primitive 8th root of unity in an algebraic extension of $\mathbb{Z}/p\mathbb{Z}$, i.e., a root of t^8-1 over $\mathbb{Z}/p\mathbb{Z}$ generating the 8th roots of unity and set $y = \zeta + \zeta^{-1}$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$. As $\zeta^8 = 1$ and ζ has order 8, we have $\zeta^4 = -1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$ with $1 \neq -1$ since p is odd. Multiplying this last equation by ζ^{-2}, we see that $\zeta^2 + \zeta^{-2} = 0$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $y^2 = (\zeta+\zeta^{-1})^2 = \zeta^2 + 2 + \zeta^{-2} = 2$. [Cf. this to the case in characteristic zero.] Let $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism. By Euler's Criterion, $\Big(\dfrac{2}{p}\Big) = \overline{2}^{\frac{p-1}{2}} = y^{p-1}$ in $\mathbb{Z}/p\mathbb{Z}$. [Note that $y \notin \mathbb{Z}/p\mathbb{Z}$, so y^{p-1} is not necessarily one.] By the Children's Binomial Theorem, $y^p = \zeta^p + \zeta^{-p}$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$.

Case 1. $p \equiv \pm 1 \mod 8$:

As $\zeta^8 = 1$, we have

$$y^p = \zeta^p + \zeta^{-p} = \zeta + \zeta^{-1} = y \text{ in } (\mathbb{Z}/p\mathbb{Z})(\zeta),$$

so $\Big(\dfrac{2}{p}\Big) = y^{p-1} = 1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $\Big(\dfrac{2}{p}\Big) = 1$ in $\mathbb{Z}$.

Case 2. $p \equiv \pm 5 \mod 8$:

As $\zeta^4 = -1$, we have

$$y^p = \zeta^p + \zeta^{-p} = \zeta^5 + \zeta^{-5} = -\zeta - \zeta^{-1} = -y \text{ in } (\mathbb{Z}/p\mathbb{Z})(\zeta),$$

so $\Big(\dfrac{2}{p}\Big) = y^{p-1} = -1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $\Big(\dfrac{2}{p}\Big) = -1$ in $\mathbb{Z}$.

This completes the proof. $\square$

Examples 59.21. 1. Quadratic Reciprocity and the properties of the Legendre symbol allows computing Legendre symbols quite efficiently. For example,

$$\begin{aligned}\Big(\frac{29}{43}\Big) &= (-1)^{\frac{29-1}{2}\frac{43-1}{2}}\Big(\frac{43}{29}\Big) = \Big(\frac{43}{29}\Big) = \Big(\frac{14}{29}\Big) = \Big(\frac{2}{29}\Big)\Big(\frac{7}{29}\Big)\\ &= (-1)^{\frac{(29-1)(29+1)}{8}}\Big(\frac{7}{29}\Big) = -\Big(\frac{7}{29}\Big) = (-1)^{\frac{7-1}{2}\frac{29-1}{2}}\Big(\frac{29}{7}\Big)\\ &= -\Big(\frac{29}{7}\Big) = -\Big(\frac{1}{7}\Big).\end{aligned}$$

So 29 is not a square modulo 43.

2. The Legendre Symbol can also be use to study Diophantine equations. Let x and y be variables. We show that the Diophantine equation $y^2 = x^3 + 45$ has no solution, i.e., no solution in integers. Suppose that $(x_0, y_0) \in \mathbb{Z}^2$ is a solution. We first show that we cannot have $3 \mid x_0$. If $x_0 = 3x_1$, then $y_0 = 3y_1$ for some integers x_1, y_1. It follows that $3^2y_1^2 = 3^3x_1^3 + 3^2 5$ or $y_1^2 \equiv 5 \equiv 2 \mod 3$ which is impossible. If x_0 is even, then $y_0^2 \equiv 45 \equiv 5 \mod 8$ which is impossible and if $x_0 \equiv 1 \mod 4$, then $y_0^2 \equiv 46 \equiv 2 \mod 4$, which is also impossible. Therefore, we are reduced to the case $x_0 \equiv 3 \mod 4$, equivalently, $x_0 \equiv 3, 7 \mod 8$.

   Case. $x_0 \equiv 7 \equiv -1 \mod 8$.

   As $y_0^2 = x_0^3 + 45$, we have $y_0^2 - 2 \cdot 3^2 \equiv x_0^3 + 27 = (x_0+3)(x_0^2 - 3x + 9)$ and $x_0^2 - 3x_0 + 9 \equiv -3 \mod 8$. Therefore, there exists a prime p satisfying $p \equiv \pm 3 \mod 8$ and $p \mid x_0^2 - 3x_0 + 9$. This implies that $y_0^2 = 2 \cdot 3^2 \mod p$. If $p \mid y_0$, then $p \mid x_0$ which is impossible. So $p \neq 3$, hence $(\dfrac{2 \cdot 3^2}{p}) = (\dfrac{2}{p}) = -1$, contradicting the Second Supplement.

   Case. $x_0 \equiv 3 \mod 8$.

   We have $y_0^2 - 2 \cdot 6^2 = x_0^3 - 27 = (x_0 - 3)(x_0^2 + 3x_0 + 9)$, so there exists a prime p, satisfying $p \equiv \pm 3 \mod 8$ and $p \mid x_0^2 + 3x_0 + 9$, as $x_0^2 + 3x_0 + 9 \equiv 3 \mod 8$. Therefore, $y_0^2 \equiv 2 \cdot 6^2 \mod p$ has a solution. Since $p \neq 3$, hence $(\dfrac{2 \cdot 6^2}{p}) = (\dfrac{2}{p}) = -1$, contradicting the Second Supplement.

Let $K/\mathbb{Q}$ be a finite field extension. Number theory studies the set $\mathbb{Z}_K := \{x \in K \mid x \text{ is a root of a monic polynomial } f \text{ in } \mathbb{Z}[t]\}$. We mention some of the facts about this here, and study this subject in Chapter XV. The set $\mathbb{Z}_K$ is a ring, hence a domain. It is called the *ring of algebraic integers* in K. Although it is not a UFD, it has an analogous property about ideals, viz, that every proper ideal in $\mathbb{Z}_K$ is a product of prime ideals, unique up to order. Such a domain is called a *Dedekind domain*. It turns out that every nonzero prime ideal $\mathfrak{p}$ in $\mathbb{Z}_K$ is maximal, so $\mathbb{Z}_K/\mathfrak{p}$ is a field, in fact a finite field so a finite extension of $\mathbb{Z}/p\mathbb{Z}$ where $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ for some prime p. Let $f_\mathfrak{p}$ denote the degree of this extension. One studies the factorization of $p\mathbb{Z}$, with p a prime integer, in $\mathbb{Z}_K$. Suppose that this factorization is $p\mathbb{Z}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, with $\mathfrak{p}_i$ distinct prime ideals in $\mathbb{Z}_K$ and e_i positive integers. A basic result is that $[K : \mathbb{Q}] = \sum_i^r e_i f_{\mathfrak{p}_i}$. If, in addition, $K/\mathbb{Q}$ is Galois then all the e_i's are equal, say equal e, and all the $\mathfrak{p}_i$'s are equal, say equal f, so $[K : \mathbb{Q}] = efr$. The Legendre symbol provides information about the case that $K/\mathbb{Q}$ is a quadratic extension (hence Galois), say $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer or -1. In this case $\mathbb{Z}_K = \mathbb{Z}[D]$ where $D = d$ if $d \equiv 2$ or $3 \mod 4$ and $D = (-1 + \sqrt{d})/2$ if $d \equiv 1 \mod 4$. Let p be a prime integer. Three possibilities can occur (called the *splitting type* of p).

(i) $r = 2$, i.e., $p\mathbb{Z}_K$ factors as a product of two distinct primes – we say that p *splits completely* in $\mathbb{Z}_K$.

(ii) $f = 1$, i.e., $p\mathbb{Z}_k$ is a prime ideal in $\mathbb{Z}_K$ – we say that p is *inert* in $\mathbb{Z}_K$.

(iii) $e = 1$, i.e., $p\mathbb{Z}_K$ is the the square of a prime ideal in $\mathbb{Z}_K$ – we say that p *ramifies* in $\mathbb{Z}_K$.

A prime p ramifies in $\mathbb{Z}_K$ if and only if $p \mid d_K$, where $d_k = 4D$ if $D \equiv 2$ or $3 \mod 4$ and $d_K = D$ if $d \equiv 1 \mod 4$. Let p be an odd prime. Then p ramifies if and only if $p \mid D$, so assume this is not the case. Then $f = 1$ or 2 depending on whether d is a square or not modulo p, i.e., on $\left(\dfrac{d}{p}\right)$. If $p\mathbb{Z}$ splits completely and a is a square modulo p, then $p\mathbb{Z}_K = (p, a + \sqrt{d})(p, a - \sqrt{d})$. Note that if p and q are odd primes, the Law of Quadratic Reciprocity tells the relation of the splitting type of p in $\mathbb{Z}_{\mathbb{Z}[\sqrt{q}]}$ and the splitting type of q in $\mathbb{Z}_{\mathbb{Z}[\sqrt{p}]}$. The splitting type of 2 is more complicated.

Exercises 59.22.

1. Show that the set of all arithmetic functions satisfies the associative law under the Dirichlet product. In particular, the set $\mathcal{A}$ of arithmetic functions nonzero at 1 is an abelian monoid.
2. Let N, τ, and σ be the arithmetic function functions defined by
$$\begin{aligned} N(n) &= n, \\ \tau(n) &= \text{ the number of divisors of } n, \text{ i.e., } \tau(n) = \sum_{d|n} 1, \text{ and} \\ \sigma(n) &= \text{ the sum of the divisors of } n, \text{ i.e., } \sigma(n) = \sum_{d|n} d \end{aligned}$$
for all n, respectively. Show that $N = \varphi \star U$, $\varphi = N \star \mu$, $\sigma = \varphi \star \tau$, and $\varphi \star \sigma = N \star N$.
3. Let $\zeta(s) := \sum \frac{1}{n^s}$, s a real (or complex) variable. Show the following:
$$\begin{aligned} \frac{\zeta(s-1)}{\zeta(s)} &= \sum_1^\infty \frac{\varphi(n)}{n^s} \text{ if (the real part of) } s > 2. \\ \zeta(s)^2 &= \sum_1^\infty \frac{\tau(n)}{n^s} \text{ if (the real part of) } s > 1. \\ \zeta(s-1)\zeta(s) &= \sum_1^\infty \frac{\sigma(n)}{n^s} \text{ if (the real part of) } s > 2. \end{aligned}$$
4. Show all of the following:
   (i) The Dirichlet product of two multiplicative functions is multiplicative.
   (ii) If f and g are arithmetic functions not zero at one with g and $f \star g$ multiplicative, then f is multiplicative.
5. The set of multiplicative functions is an abelian group under the Dirichlet product.
6. Prove the following versions of the Möbius Inversion Theorem.
   (a) If $f, g : \mathbb{Z}^+ \to G$ with G an additive group, then
$$f(n) = \sum_{d|n} g(d) \text{ if and only if } g(n) = \sum_{d|n} f(d)\mu(\frac{n}{d})$$
   for all $n \in \mathbb{Z}^+$, where μ is viewed as a function to G.

(b) If $f, g : \mathbb{Z}^+ \to G$ with G a multiplicative group, then

$$f(n) = \prod_{d|n} g(d) \text{ if and only if } g(n) = \prod_{d|n} f(d)\mu(\frac{n}{d})$$

for all $n \in \mathbb{Z}^+$, where μ is viewed as a function to G.

7. Let F be any field and d a positive integer. Show that $t^d - 1 \mid t^n - 1$ in $F[t]$ if and only if $d \mid n$.
8. Show that every finite cyclic group occurs as a Galois group $G(K/\mathbb{Q})$ for some Galois extension $K/\mathbb{Q}$.
9. Prove that every finite abelian group occurs as a Galois group over the rational numbers. (Hint: Use Proposition 59.10.)
10. Prove Check 59.20.
11. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathbb{Z}_K = (p, a + \sqrt{d})(p, a - \sqrt{d})$ in $\mathbb{Z}_K$ if p does not divide d and a in $\mathbb{Z}$ is a square modulo p.
12. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathbb{Z}_K$ is a prime ideal if $\mathbb{Z}_K$ if a in $\mathbb{Z}$ is not a square modulo p.
13. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathbb{Z}_K = (p, \sqrt{d})^2$ in $\mathbb{Z}_K$ if $p \mid d$.

60. Radical Extensions

In this section, we generalize the notion of square root towers and solve the problem of when a formula for roots of a polynomial involving only addition, multiplication (in which we include taking inverses) and extraction of nth roots of elements for various n over the rational numbers exists. The key is if the Galois group of a polynomial is a solvable group.

Definition 60.1. An extension of fields K/F is called a *radical extension* if there exist elements u_i in K for $1 \le i \le m$ (some m) such that $K = F(u_1, \ldots, u_m)$ and for each i there exists a positive integer n_i such that $u_1^{n_1}$ lies in F and $u_i^{n_i} \in F(u_1, \ldots, u_{i-1})$ for each $i > 1$.

As one would expect, radical extensions have properties similar to square root towers, and they do.

Remark 60.2. Let K/F and L/F be field extensions with K and L lying in an extension field of K.

1. If K/F is a square root tower, then K/F is radical.
2. If K/F is the splitting field of $t^n - 1$ in $F[t]$, then K/F is radical.
3. If K/F is the splitting field of $t^n - a$ in $F[t]$, then K/F is radical, as $K = F(\omega, \theta)$ with ω a primitive nth root of unity in K and θ an element in K such that $\theta^n = a$ in K.
4. If K/F is radical, then so is $L(K)/L$.
5. If L/K and K/F are both radical, then so is L/F.

6. If L/F is radical and L/K, then L/K is radical.
7. If $E/L_i/F$ are intermediate fields with each L_i/F radical for $i = 1, \dots, n$, then $F(L_1 \cup \cdots \cup L_n)/F$ is radical.
8. If K/F is radical and $\sigma : K \to L$ is a (field) homomorphism, then $\sigma(K)/\sigma(F)$ is radical.
9. If K/F is radical and L/K is a normal closure of K/F, then L/F is radical. Indeed this follows from the last two remarks, as $L = F\big(\cup_{G(L/F)} \sigma(K)\big)$.
10. If K/F is a square root tower and L/K is a normal closure of K/F, then L/F is a square root tower.

The last remark allows us to refine the Constructibility Criterion 52.10.

Theorem 60.3. (Constructibility Criterion (Refined Form)) *Let z be a complex number and $z_1(=0), z_2(=1), \dots, z_n$ other complex numbers with $n \geq 2$. Set $F = \mathbb{Q}(z_1, z_2, \dots, z_n, \overline{z_1}, \dots, \overline{z_n})$. Then the following are equivalent:*

(1) The complex number z is constructible from $z_1, \dots, z_n$.
(2) The complex number z is algebraic over F and the normal closure of $F(z)/F$ in $\mathbb{C}$ is a square root tower.
(3) The complex number z is algebraic over F and if $E/F(z)$ is the normal closure of $F(z)/F$ in $\mathbb{C}$, then $[E : F] = 2^e$ for some e.

Proof. (2) if and only if (3) follows from the Square Root Tower Theorem 57.11.
(2) $\Rightarrow$ (1) follows from the original Constructibility Criterion 52.10.
(1) $\Rightarrow$ (2): By the original Constructibility Criterion 52.10, there exists a square root tower K/F with $z \in K$. By Remark (10) above, we may assume that K/F is normal. Let $K/E/F(z)/F$ with $E/F(z)$ the normal closure of $F(z)/F$ in K. Since $[E : F] \mid [K : F] = 2^e$, some e, we are done by the Square Root Tower Theorem 57.11. $\square$

We wish to find the group theoretic criterion for a finite extension to be a radical extension. For simplicity we shall establish this in the case that the ground field is of characteristic zero.

Let K/F be a finite Galois extension of fields. Recall that we call it an *abelian extension* (respectively, *cyclic extension*) if $G(K/F)$ is abelian (respectively, cyclic).

Example 60.4. 1. Any cyclic extension of fields is abelian.

2. Suppose n is a positive integer and either $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$. If K is a splitting field of $t^n - 1$ over F, then K/F is abelian (and, in fact, cyclic if n is $2, 4, p^r, 2p^r$, with p an odd prime, cf. F.6 below).
3. If K/F is abelian (respectively, cyclic) and $K/E/F$ is an intermediate field, then both K/E and E/F are abelian (respectively, cyclic).
4. Suppose n is a positive integer and either $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$. Suppose that $t^n - 1$ splits over F. If K is a splitting field of $t^n - a$ over F, then K/F is cyclic.

We review our study about solvable groups.

Review 60.5. A group G is called *solvable* (respectively, *polycyclic*) if there exist a finite sequence of subgroups of G:

$$1 = N_0 \subset N_1 \subset \cdots \subset N_r = G$$

satisfying $N_i \triangleleft N_{i+1}$ and N_{i+1}/N_i is abelian (respectively, cyclic) for every $i = 1, \dots, r-1$, i.e., G has a finite subnormal series with abelian (respectively cyclic) factors. Let G be a group. We know the following facts about solvable and polycyclic groups:

1. Abelian groups are solvable.
2. If G is solvable and H a subgroup of G, then H is solvable.
3. If N is a normal subgroup of G, then G is solvable if and only if both N and G/N are solvable.
4. If G is polycyclic, then G is solvable.
5. If G is finite and solvable, then G is polycyclic.
6. The alternating group A_n with $n \geq 5$ is a nonabelian simple group, hence not solvable.
7. The symmetric group S_n with $n \geq 5$ is not solvable.

Applying the above to field theory we have:

Remarks 60.6. Let K/F be a finite extension of fields with $K/E/F$ an intermediate field.

1. If $G(K/F)$ is solvable, then $G(K/E)$ is solvable.
2. If K/F and E/F are Galois, then $G(K/F)$ is solvable if and only if both $G(K/E)$ and $G(E/F)$ are solvable, since, by the Fundamental Theorem of Galois Theory, $G(E/F) \cong G(K/F)/G(K/E)$.

Given a field extension, it is often useful to extend the base field by adjoining appropriate roots of unity. In investigating radical extensions, this is most useful. Indeed using the last remark we easily establish the following lemma.

Lemma 60.7. *Let n be a positive integer and F a field satisfying $\operatorname{char} F = 0$ or $\operatorname{char} F \not| \, n$. Suppose that K/F is a splitting field of a separable polynomial f over F and L/F is a a splitting field of $t^n - 1$ over K. Then $G(L/F)$ is Galois. Moreover, $G(L/F)$ is solvable if and only if $G(K/F)$ is solvable.*

PROOF. We know that L is a splitting field of the separable polynomial $(t^n - 1)f$ over F, so L/F is Galois. As L/K is abelian, by Remark 60.6, the group $G(L/F)$ is solvable if and only if the group $G(K/F)$ is solvable. $\square$

The key result is the following theorem.

Theorem 60.8. *Let F be a field of characteristic zero and K/F a radical extension. If $K/E/F$ is an intermediate field, then $G(E/F)$ is solvable.*

PROOF. We begin with two reductions.

Reduction 1. We may assume that E/F is Galois:

Let $F_0 = E^{G(E/F)} \supset F$. Then K/F_0 is also radical, $G(E/F) = G(E/F_0)$, and E/F_0 is Galois. Replacing F by F_0 establishes the reduction.

Reduction 2. We may assume that K/F is Galois:

Let L/K be a normal closure of of K/F. Since K/F is radical, so is L/F. Replacing K by L establishes the second reduction.

So we may now assume that K/F is a Galois and a radical extension. By Remark 60.6, it now suffices to show that $G(K/F)$ is solvable as E/F is Galois. Therefore, we have reduced to proving the following:

Claim. Let F be a field of characteristic zero and K/F a finite extension that is both radical and Galois. Then $G(K/F)$ is solvable:

Let $K = F(u_1, \dots, u_m)$ and n_i positive integers, $i = 1, \dots, m$, satisfying $u_1^{n_1}$ lies in F and $u_i^{n_i} \in F(u_1, \dots, u_{i-1})$ for each $i > 1$. We prove the claim by induction.

Induction Hypothesis. Let L/E be a finite, radical, Galois extension of fields with E a field of characteristic zero. Suppose that $L = E(v_1, \dots, v_{m-1})$ and r_i positive integers, $i = 1, \dots m-1$, satisfying $v_1^{r_1}$ lies in E and $v_i^{r_i} \in E(v_1, \dots, v_{i-1})$ for each $i > 1$. Then $G(L/E)$ is solvable:

Note: The $m = 0$ case is trivial (and the $m = 1$ case is included in the proof below).

With $K = F(u_1, \dots, u_m)$ as above, let ζ be a primitive n_1st root of unity, so a primitive root of unity of $t^{n_1} - 1$ over K. We have the following picture with the explanation to follow:

$$\begin{tikzcd}
 & K(\zeta) \arrow[dl, dash, "ab"'] \arrow[ddd, dash, "gal"] \arrow[dr, dash, "rad"] & \\
 & & F(\zeta), u_1) \arrow[d, dash, "cyc"] \\
K \arrow[dr, dash, "{gal, rad}"'] & & F(\zeta). \arrow[dl, dash, "ab"] \\
 & F &
\end{tikzcd}$$

$K(\zeta)/K$ and $F(\zeta)/F$ are Galois and abelian, since a splitting field over $t^{n_1} - 1$ over K and F, respectively. The extension $K(\zeta)/F$ is Galois by the lemma, hence $K(\zeta)/F(\zeta)$ and $K(\zeta)/F(\zeta, u_{n_1})$ are Galois. Since $F(\zeta, u_1)$ is a splitting field of $t^{n_1} - u_1^{n_1}$ over $F(\zeta)$, the extension $F(\zeta, u_1)/F$ is Galois, as it is a splitting field of the separable polynomial $(t^{n_1} - u_1^{n_1})(t^{n_1} - 1)$ over F. We know that $F(\zeta, u_1)/F(\zeta)$ is cyclic and $F(\zeta)/F$ is abelian, so $G(F(\zeta, u_1)/F)$ is solvable by Remark 60.6. In particular, all the extensions in the picture above are Galois. We have $K(\zeta) = F(\zeta, u_1)(u_2, \dots, u_m)$ with $K(\zeta)/F(\zeta, u_1)$ radical and Galois. Consequently, by the induction hypothesis, $G\big(K(\zeta)/F(\zeta, u_1)\big)$ is solvable. As $K(\zeta)/F$ is Galois and $G\big(K(\zeta)/F(\zeta, u_1)\big)$ is solvable, the group $G(K(\zeta)/F)$ is solvable by Remark 60.6. As K/F is Galois, the group $G(K/F)$ is solvable again by Remark 60.6. $\square$

The theorem shows that if f is an irreducible polynomial in $F[t]$, e.g., $F = \mathbb{Q}$, then a formula for a root of f exists that consists of addition, multiplication, and extraction of nth roots for various n, if and only if there exists a radical extension K of F such that f has a root in K. As the conjugate of a radical extension is radical, if such a formula for one of the roots of irreducible f_i, such a formula holds for each root of f_i in the appropriate conjugate.

Let F be a field and f be a non-constant polynomial in $F[t]$. We say that f is *solvable by radicals* if for each irreducible factor f_i of f in $F[t]$, there exists a formula for a root of f_i involving addition, multiplication, and extraction of nth roots for various n, i.e., there exists a radical extension K_i of F such that f_i has a root in K_i for each i.

We analyze what we have shown. Let F be a field of characteristic zero. Suppose that f in $F[t]$ is solvable by radicals and f_i an irreducible factor of f in $F[t]$. Then there exists a radical extension K_i/F such that f_i has a root in K. Assume that each such K_i lies in a common extension field $\widetilde{F}$, e.g., an algebraic closure of F. If there exist m distinct (non-associative) factors f_i of f, let $L = F(K_1 \cup \cdots \cup K_m)$, a radical extension of F. The normal closure $\widetilde{L}/L$ in $\widetilde{F}$ of L/F is also a radical extension, so there exists a radical Galois extension $\widetilde{L}/F$ such that f splits over $\widetilde{L}$. Let $\widetilde{L}/E/F$ with E a splitting field of f over F. Although E/F may not be radical, by the theorem, we know that $G(E/F)$ is solvable. Recall if f is a polynomial in $F[t]$ and K/F is a splitting field of f, then $G(K/F)$ is called the *Galois group of* f. In general it is only unique up to isomorphism, but if we have fixed an algebraic closure of F, as we have done this time, it is unique. Therefore, we have shown the following:

Theorem 60.9. *Let F be a field of characteristic zero and f an non-constant polynomial in $F[t]$. If f is solvable by radicals, then the Galois group of f is solvable.*

We wish to show that there exist polynomials in $\mathbb{Q}[t]$ whose Galois groups are not solvable by radicals. We need the following group theory result and its application to field theory.

Lemma 60.10. *Let p be a prime and G a subgroup of the symmetric group S_p containing a p-cycle and a transposition. Then $G = S_p$.*

This is not true if p is not a prime.

To prove the lemma, we recall the following fact: For any n-cycle $(a_1 \cdots a_n)$, we have

$$\sigma(a_1 \cdots a_n)\sigma^{-1} = \big(\sigma(a_1) \cdots \sigma(a_n)\big)$$

for all permutations σ in S_N with $n \leq N$.

PROOF. (of the lemma) By changing notation, we may assume that $(12) \in G$. Let $\sigma = (a_1 \cdots a_p)$ lie in G. As p is a prime, σ^i, $1 \leq i < p$, is also a p-cycle (cf. Exercise 24.24(1), so we may assume that $\sigma = (12 \cdots p)$. As $\sigma(1\,2)\sigma^{-1} = (2\,3)$, $\sigma(2\,3)\sigma^{-1} = (3\,4)$, etc., we see that $(i\,i+1)$ lies in G for all i. Therefore $(1\,i+1) = (1\,i)(i\,i+1)(1\,i)$ lies in G for all i by induction, hence $G = S_p$ by Proposition 24.7.

$\square$

Proposition 60.11. *Let f be an irreducible polynomial in $\mathbb{Q}[t]$ of degree p with p a prime. Suppose that f has precisely two nonreal roots in the splitting field K in $\mathbb{C}$ of f over $\mathbb{Q}$. Then $G(K/\mathbb{Q}) \cong S_p$.*

PROOF. Let α in K be a root of f, so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg f = p$. As $\operatorname{char} K = 0$, we know that $K/\mathbb{Q}$ is Galois, hence $|G(K/\mathbb{Q}) = [K : \mathbb{Q}] \leq p!$. As $G(K/\mathbb{Q})$ acts as permutations on the roots of f and K is the splitting field of f, we may assume that $G(K/\mathbb{Q}) \subset S_p$. Since f has precisely two nonreal roots, the restriction of complex conjugation to K is a nontrivial ($\mathbb{Q}$-) automorphism of K and it must interchange these two nonreal roots and fix all the other roots of f, i.e., it corresponds to a transposition in $G(K/\mathbb{Q})$. As $p \mid [K : \mathbb{Q}] = |G(K/\mathbb{Q})|$, the group $G(K/\mathbb{Q})$ must contain an element of order p by Cauchy's Theorem 21.22. But the only elements in S_p of order p are p-cycles. Therefore, $G(K/\mathbb{Q}) = S_p$ by the lemma. □

As a consequence, we obtain the following famous theorem:

Theorem 60.12. (Abel-Ruffini) *In general, there is no formula to determine the roots of a general fifth degree polynomial in $\mathbb{Q}[t]$ that involves only addition, multiplication, and extraction of nth roots for various n.*

Proof. It suffices to give an example of a polynomial f in $\mathbb{Q}[t]$ that has a nonsolvable Galois group. Let $f = t^5 - 6t + 3$. We know that f is irreducible by Eisenstein's Criterion, and by calculus it has three real roots in $\mathbb{R}$, since $f' = 5t^4 - 6$ has only two real roots, $f(-1) > 0$, and $f(1) < 0$. By the proposition, the Galois group of f is isomorphic to S_5 which is not solvable. □

Corollary 60.13. *In general, there is no formula to determine the roots of a general nth degree polynomial in $\mathbb{Q}[t]$ that involves only addition, multiplication, and extraction of nth roots for various n.*

Proof. Apply the theorem to any nth degree polynomial with $n \geq 5$ divisible by a fifth degree polynomial not solvable by radicals. □

Remark 60.14. 1. For each positive integer n, we have constructed fields F and K_n/F Galois with $G(K_n/F) \cong S_n$ in Example 57.6(3), so, in general, no formula can exist.

2. Hilbert showed that for each positive integer n, there exists a Galois extension $K_n/\mathbb{Q}$ with $G(K_n/\mathbb{Q}) \cong S_n$. This is not easy. It uses his Irreducibility Theorem. For a proof see Chapter 69. By the Primitive Element Theorem, $K_n = \mathbb{Q}(u)$ for some u, hence the Galois group of $m_{\mathbb{Q}}(u)$ is isomorphic to S_n, so is not solvable for $n \geq 5$.
3. There exist fields of characteristic zero such that every algebraic extension is radical, hence every non-constant polynomial over it is solvable by radicals, e.g., $\mathbb{R}$, $\mathbb{C}$. There are others, e.g., the so called local fields of characteristic zero that arise in number theory.
4. If F is a finite field, then any finite extension of F is cyclic so radical. In particular, the Galois group of any polynomial over f is solvable by radicals.

We want a converse to our result. To do so needs additional concepts.

Definition 60.15. Let K/F be a finite Galois extension of fields and x an element of K. Define the *norm* of x to be

$$\mathrm{N}_{K/F}(x) := \prod_{G(K/F)} \sigma(x)$$

and the *trace* of x to be

$$\mathrm{Tr}_{K/F}(x) := \sum_{G(K/F)} \sigma(x).$$

Examples 60.16. Let F be a field of characteristic different from two and d an element in F that is not a square. Let $K = F(\sqrt{d})$. If $x = a + b\sqrt{d}$ with a, b in F, then $\mathrm{N}_{K/F}(x) = a^2 - b^2 d$ and $\mathrm{Tr}_{K/F}(x) = 2a$. Both of these lie in F. If, in addition, x does not lie in F, i.e., b is not zero, then $K = F(x)$ and $m_F(x) = t^2 - \mathrm{Tr}_{K/F}(x)t + \mathrm{N}_{K/F}(x) = \big(t - (a + b\sqrt{d})\big)\big(t - (a - b\sqrt{d})\big)$ lies in $F[t]$.

Properties 60.17. Let K/F be a finite Galois extension of fields, $G = G(K/F)$, and x, y elements of K.

1. If τ lies in G, then $\tau G = G = G\tau$, so
$$\tau\big(\mathrm{N}_{K/F}(x)\big) = \mathrm{N}_{K/F}(x) = \mathrm{N}_{K/F}\big(\tau(x)\big)$$
$$\tau\big(\mathrm{Tr}_{K/F}(x)\big) = \mathrm{Tr}_{K/F}(x) = \mathrm{Tr}_{K/F}\big(\tau(x)\big).$$
In particular,
$$\mathrm{N}_{K/F}(x) \text{ and } \mathrm{Tr}_{K/F}(x) \text{ lie in } K^G = F:$$
We have
$$\tau(\prod_G \sigma(x)) = \prod_G \tau\sigma(x) = \prod_G \sigma(x) = \prod_G \sigma\tau(x)$$
$$\tau(\sum_G \sigma(x)) = \sum_G \tau\sigma(x) = \sum_G \sigma(x) = \sum_G \sigma\tau(x).$$
2. Suppose that $G = \{\sigma_1, \dots, \sigma_n\}$. As before let s_j denote the jth elementary symmetric function $\sum_{1 \le i_1 < \cdots < i_j \le n} t_{i_1} \cdots t_{i_j}$ (and $s_0 = 1$). Then
$$m_F(x) = \prod \big(t - \sigma_i(x)\big) = \sum (-1)^{n-i} s_{n-i}\big(\sigma_1(x), \dots, \sigma_n(x)\big)t^i$$
in $F[t]$. In particular, if $K = F(x)$, then
$$\mathrm{N}_{K/F}(x) = s_n\big(\sigma_1(x), \dots, \sigma_n(x)\big)$$
$$\mathrm{Tr}_{K/F}(x) = s_1\big(\sigma_1(x), \dots, \sigma_n(x)\big).$$
What can you say if $F(x) < K$?
3. If x lies in F, then $\sigma(x) = x$ for each σ in G. It follows that if $K/E/F$ then
$$\mathrm{N}_{E/F}(x) = x^{[E:F]} \text{ and } \mathrm{Tr}_{E/F}(x) = [E:F]x,$$
as every F-embedding of E into K lifts to an element of G.
4. The norm is multiplicative, i.e., $\mathrm{N}_{K/F}(xy) = \mathrm{N}_{K/F}(x)\,\mathrm{N}_{K/F}(y)$ and, if x is nonzero, $\mathrm{N}_{K/F}(x^{-1}) = \big(\mathrm{N}_{K/F}(x)\big)^{-1}$ (as $\mathrm{N}_{K/F}(1) = 1$). In particular,
$$\mathrm{N}_{K/F} : K^\times \to F^\times \text{ is a group homomorphism.}$$
5. The map $\mathrm{Tr}_{K/F} : K \to F$ is an F-linear functional of F-vector spaces. By Dedekind's Lemma 54.3, $\mathrm{Tr}_{K/F}$ is not the trivial map as it is the sum of distinct F-homomorphisms. In particular, $\dim \ker \mathrm{Tr}_{K/F} = [K:F] - 1$.

We can generalize the norm and trace as follows:

Remark 60.18. Let E/F be a finite separable extension of fields of degree n with K/E a normal closure of E/F. Hence K/F is Galois. Let

$$\tau_1, \dots, \tau_m : E \to K$$

be all the distinct F-homomorphisms. We know that $m = n$ as E/F is finite and separable. Let x be an element in E. Define the *norm* of x to be

$$\mathrm{N}_{E/F}(x) := \prod_i \tau_i(x)$$

and the *trace* of x to be

$$\mathrm{Tr}_{E/F}(x) := \sum_i \tau_i(x).$$

Since K/F is Galois, if $\sigma \in G(K/F)$, then

$$\sigma\tau_1, \dots, \sigma\tau_n : E \to K$$

are distinct F-homomorphisms, so these are just a permutation of $\tau_1, \dots, \tau_n$, hence

$$\mathrm{N}_{E/F}(x) \text{ and } \mathrm{Tr}_{E/F}(x) \text{ lie in } K^G = F$$

just as in the Galois case. [Note we only needed $K/E/F$ with K/F finite and Galois to deduce this i.e., is independent of the finite Galois extension of F containing E.] Moreover, Properties 60.17 hold for E/F with $\{\tau_1, \dots, \tau_m\}$ replacing the σ in Property 60.17(1) and $\{\sigma_1, \dots, \sigma_n\}$ in Property 60.17(2) This generalization is useful in applications, especially in studying towers of finite separable extensions of fields. (Cf. the exercises at the end of this section). This can be further generalized to include inseparable extensions, but we shall not pursue this here.

In his book on algebraic number theory, Hilbert had the following result as his 90th Satz (theorem):

Lemma 60.19. (Hilbert Theorem 90) *Let K/F be a Galois extension of fields of degree n. Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $x \in K$. Then*

$$\mathrm{N}_{K/F}(x) = 1 \textit{ if and only if there exists a } y \in K^\times \textit{ satisfying } x = \frac{y}{\sigma(y)},$$

i.e., we have an exact sequence

$$1 \to K^\times \xrightarrow{\psi} K^\times \xrightarrow{\mathrm{N}_{K/F}} F^\times$$

with $\psi(y) = y/\sigma(y)$.

Proof. ($\Leftarrow$): If $x = y/\sigma(y)$, then

$$\mathrm{N}_{K/F}(x) = \frac{\mathrm{N}_{K/F}(y)}{\mathrm{N}_{K/F}\big(\sigma(y)\big)} = 1.$$

($\Rightarrow$): Suppose that

$$1 = \mathrm{N}_{K/F}(x) = x\sigma(x)\cdots\sigma^{n-1}(x).$$

In particular, x is nonzero. By Dedekind's Lemma 54.3, we know that the F-automorphisms $1, \sigma, \dots, \sigma^{n-1}$ are independent, so if $a_0, \dots, a_{n-1}$ lie in K, we have $\sum_{i=0}^{n-1} a_i\sigma^i(z) = 0$ for all z in K if and only if all the a_i are zero. Set $a_0 = 1$ and $a_i = x\sigma(x)\cdots\sigma^{i-1}(x)$ in $K^\times$ for $i = 1, \dots, n-1$. Then

$$\begin{aligned} g &= 1 + a_1\sigma + \cdots + a_{n-1}\sigma^{n-1} \\ &= 1 + x\sigma + x\sigma(x)\sigma^2 + \cdots + x\sigma(x)\cdots\sigma^{n-2}(x)\sigma^{n-1} \end{aligned}$$

is not trivial, so there exists an element u in K satisfying

$$0 \neq y := g(u) = u + x\sigma(u) + \cdots + x\sigma(x)\cdots\sigma^{n-2}(x)\sigma^{n-1}(u).$$

Since $\sigma^n = 1_K$ and $1 = \mathrm{N}_{K/F}(x) = x\sigma(x)\cdots\sigma^{n-1}(x)$, we see that

$$x\sigma(y) = x\sigma(u) + x\sigma(x)\sigma^2(u) + \cdots + x\sigma(x)\cdots\sigma^{n-1}(x)\sigma^n(u) = y,$$

i.e., $x = y/\sigma(y)$ as needed. □

Although the computational proof reveals little, generalizations of the lemma above have resulted in very deep theorems in algebra. Hilbert established it to prove the following important result.

Theorem 60.20. *Let K/F be a Galois extension of fields of degree n with the characteristic of F either zero or* $\operatorname{char} F \nmid n$. *Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $t^n - 1$ splits over F. Then there exists an element u in K satisfying $K = F(u)$ and $m_F(u) = t^n - a$ in $F[t]$ for some a in F. In particular, K/F is radical.*

Proof. Let ζ be a primitive nth root of unity in F. By our previous work, we know that $U = \langle\zeta\rangle$ is a cyclic group satisfying $|U| = n$, so $|U| = [K : F]$. As ζ^{-1} lies in F and $\mathrm{N}_{K/F}(\zeta^{-1}) = \zeta^{-n} = 1$, an application of Hilbert Theorem 90 shows that there exists a nonzero element u in K satisfying $\zeta = \sigma(u)/u$, i.e., $\sigma(u) = \zeta u$. Since $\sigma(u^n) = \sigma(u)^n = \zeta^n u^n = u^n$, we have $\sigma^i(u^n) = u^n$ for all i. Therefore, $a := u^n$ lies in $K^{G(K/F)} = F$ as K/F is Galois. Let $f := t^n - a$, a polynomial lying in $F[t]$. This polynomial has n distinct roots: $\zeta^i u$ for $i = 1, \dots, n$. In particular, $F(u)$ is a splitting field of f over F. To finish, we need to show that $K = F(u)$, as then $\deg f = \deg m_F(u)$. We know that

$$\sigma^i|_{F(u)} : F(u) \to K \text{ maps } u \mapsto \zeta^i u$$

comprise n distinct F-homomorphisms, so by Proposition 56.11, we have

$$[F(u) : F] \geq n = |U| = [K : F],$$

hence $K = F(u)$. □

Let K/F be a Galois extension of fields of degree n with the characteristic of F either zero or $\operatorname{char} F \nmid n$. Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $t^n - 1$ splits over F. Then there exists an element u in K satisfying $K = F(u)$ and $m_F(u) = t^n - a$ in $F[t]$ for some a in F. In particular, K/F is radical.

Corollary 60.21. *Let K/F be a cyclic extension of fields of degree n with the characteristic of F either zero or* $\operatorname{char} F \nmid n$. *Suppose that K/F is cyclic and $t^n - 1$ splits over F. Then there exists an $x \in K$ such that $\{\sigma(x) \mid \sigma \in G(K/F)\}$ is an F-basis for K. In particular, this is true if F is a finite field.*

Proof. Let $G(K/F) = \langle\sigma\rangle$. By the proof of Theorem 60.20, $K = F(u)$ with $u^i = \sigma(u^i)$ for $i = 0, \dots, n-1$. The first statement follows. If F is a finite finite field, then L/F is a cyclic extension with $\operatorname{char} F \nmid |F|$ by Exercise 55.9(2). $\square$

We shall show that this corollary generalizes to the case that K/F is a finite Galois extension in Theorem 63.7 below. Theorem 60.20 allows us to deduce the desired appropriate converse of Theorem 60.8.

Theorem 60.22. *Suppose that F is a field of characteristic zero and K/F is a finite Galois extension with $G(K/F)$ solvable. Then there exists a finite extension L/K such that L/F is Galois and radical.*

Proof. Let $n = [K : F] = |G(K/F)|$.

Case 1. The polynomial $t^n - 1$ splits over F:

We show in this case that K/F itself is radical and Galois. As the group $G(K/F)$ is finite and solvable, it is polycyclic, i.e., has a cyclic series say

$$1 = N_1 \lhd N_2 \lhd \cdots \lhd N_r = G$$

with each N_i/N_{i-1} cyclic. Let $F_i = K^{N_i}$. By the Fundamental Theorem of Galois Theory 57.3 and induction on i, we conclude that

$$K/F_{i+1} \text{ is Galois with } G(K/F_{i+1}) = N_{i+1},$$

and, as $N_i \lhd N_{i+1}$,

$$F_i/F_{i+1} \text{ is normal with } G(F_i/F_{i+1}) \cong N_{i+1}/N_i \text{ cyclic.}$$

Let $n_i = |N_{i+1}/N_i| = [F_i : F_{i+1}]$. Since $[F_i : F_{i+1}] \mid [K : F] = |G(K/F)|$, we know that $n_i \mid n$, hence $t^{n_i} - 1 \mid t^n - 1$ in $F[t]$ (why?). Thus $t^{n_i} - 1$ in $F[t] \subset F_{i+1}[t]$ splits over F_i for each i. It follows that K/F is radical and Galois by Theorem 60.20.

Case 2. $t^n - 1$ in $F[t]$ does not split over F:

Let E be a splitting field of $t^n - 1$ over K and ζ a primitive nth root of unity in E.

Claim: $K(\zeta)/F$ is a radical and Galois extension (and, therefore, we are done).

We have already seen that $K(\zeta)/F$ is Galois as K/F is the splitting field of a separable polynomial over F. So we need only show that $K(\zeta)/F$ is a radical extension. As $K(\zeta)/F(\zeta)$ is Galois and $F(\zeta)/F$ is radical and Galois, it suffices to show that $K(\zeta)/F(\zeta)$ is radical.

We reduce to Case 1 by showing:

(i) $[K(\zeta) : F(\zeta)] = |G(K/F(\zeta))| \mid n$.
(ii) $G(K(\zeta)/F(\zeta))$ is a solvable group.

Indeed, suppose we have shown (i) and (ii). Then $F(\zeta)$ contains a primitive $|G(K(\zeta)/F(\zeta))|$th primitive root of unity, so by Case 1, the extension $K(\zeta)/F(\zeta)$ is Galois and radical. As $F(\zeta)/F$ is radical, it then follows that $K(\zeta)/F$ is radical. So we need only show (i) and (ii). Let $\sigma \in G(K(\zeta)/F(\zeta))$. Then σ fixes F (and ζ), hence $\sigma|_K : K \to K(\zeta)$ is an

F-homomorphism. By Proposition 56.11, there exist at most $[K : F] = |G(K/F)|$ F-homomorphisms $K \to K(\zeta)$. As each element of $G(K/F)$ is such an F-homomorphism, we must have $\sigma|_K$ lies in $G(K/F)$. Therefore,

$$\psi : G\big(K(\zeta)/F(\zeta)\big) \to G(K/F) \text{ given by the restriction } \sigma \mapsto \sigma|_K$$

is a well-defined group homomorphism. But ψ is monic, as $\sigma|_K = 1_K$ means that σ fixes K and ζ hence $K(\zeta)$, i.e., $\sigma = 1_{K(\zeta)}$. Therefore,

$$|G\big(K(\zeta)/f(\zeta)\big)| \mid |G(K/F)| = n$$

and $G(K\big(\zeta)/F(\zeta)\big)$ is isomorphic to a subgroup of the solvable group $G(K/F)$, so also solvable. □

Corollary 60.23. *Let F be a field of characteristic zero and f a non-constant polynomial in $F[t]$. Then f is solvable by radicals if and only if the Galois group of f is solvable.*

Proof. ($\Rightarrow$) has already been done.

($\Leftarrow$): Let K be a splitting field of f over F. By hypothesis, $G(K/F)$ is solvable, so by the theorem there exists an extension L/K with L/F Galois and radical. Since f splits over L, it is solvable by radicals. □

Corollary 60.24. *Let F be a field of characteristic zero and f a non-constant polynomial in $F[t]$ of degree at most four. Then f is solvable by radicals.*

Proof. The Galois group of f is isomorphic to a subgroup of S_4, a solvable group. □

Remarks 60.25. The quadratic formula for the general polynomial of degree two is well-known and gives a root for any quadratic polynomial (and works over any field of characteristic not two). The corollary says if F is a field of characteristic zero and f the general polynomial of degree three (respectively, four), then there exists a formula involving only field operations and taking various nth roots giving a root, hence this formula is applicable to all polynomials of this degree. This is the formula of Ferro-Tartaglia-Cardano (respectively, Ferrari). Thus all the roots of polynomials of degree at most four over a field of characteristic zero can be found by such formulae.

Exercises 60.26.

1. Let L/F be a field extension and $L/K_i/F$ with K_i/F abelian. Show that $K_1K_2 = K_1(K_2)$ is abelian.
2. Show that Lemma 60.10 is false if p is not a prime.
3. Let K/F be a finite separable extension of fields. Show that $\mathrm{N}_{K/F} = \mathrm{N}_{E/F} \circ \mathrm{N}_{K/E}$ and $\mathrm{Tr}_{K/F} = \mathrm{Tr}_{E/F} \circ \mathrm{Tr}_{K/E}$ for any intermediate field $K/E/F$.
4. Let K/F be a finite separable extension of fields, x an element of K, and $\lambda_x : F(x) \to F(x)$ the F-linear transformation given by $y \mapsto xy$. Show that the characteristic polynomial of λ_x is $\big(m_F(x)\big)^{[K:F(x)]}$. In particular, if $m_F(x) = t^r + a_{r-1}t^{r-1} + \cdots + a_0$ in $F[t]$, then $\mathrm{N}_{K/F}(x) = (-1)^r a_0^{[K:F(x)]}$ and $\mathrm{Tr}_{K/F}(x) = -[K : F(x)]a_{r-1}$

5. Suppose that K/F is a finite separable extension of fields, $K^* := \mathrm{Hom}_F(K, F)$, the F-linear dual space of K. Show that the map $T : K \to K^*$ defined by $x \mapsto \mathrm{tr}_x : y \mapsto \mathrm{Tr}_{K/F}(xy)$ is an F-linear isomorphism. In particular, if $\{x_1, \dots, x_n\}$ is an F-basis for K, then there exists a basis $\{y_1, \dots, y_n\}$ for K satisfying $\mathrm{Tr}_{K/F}(x_i y_j) = \delta_{ij}$ (the Kronecker delta).
6. Let K/F be a finite separable extension of fields, x an element of K, and $\lambda_x : F(x) \to F(x)$ the F-linear transformation given by $y \mapsto xy$. If $\mathcal{B}$ is an F-basis for K, show that $\mathrm{Tr}_{K/F}(x) = \mathrm{trace}[\lambda_x]_{\mathcal{B}}$ and $\mathrm{N}_{K/F} = \det[\lambda_x]_{\mathcal{B}}$.
7. Let K/F be a Galois extension of fields of degree n. Suppose that K/F is cyclic with $G(K/F) = \langle \sigma \rangle$ and $x \in K$. Show $\mathrm{Tr}_{K/F}(x) = 0$ if and only if there exists a $y \in K^\times$ satisfying $x = y - \sigma(y)$.
8. Let F be a field of characteristic zero and K/F an abelian extension. Let n be a positive integer such that $\sigma^n = 1$ for every $\sigma \in G(K/F)$ and F contains a primitive nth root of unity. Show that there exist $\alpha_1, \dots, \alpha_r$ in K such that $K = F(\alpha_1, \dots, \alpha_r)$ with $\alpha_i^{n_i} \in F$ for some $n_i \mid n$ for $1 = 1, \dots, r$.

61. Addendum: On Hilbert Theorem 90

In this section we use the proof of Hilbert Theorem 90 (Theorem 60.19) to state a more modern version whose generalizations has become very important in algebra and number theory.

Let K be a field and G a group of automorphisms of K. We call $S_G := \{x_\sigma \in K \mid \sigma \in G\}$ a *solution set to the Noether equations* if $0 \notin S_G$ and

(*) $$x_\sigma \sigma(x_\tau) = x_{\sigma\tau}, \text{ for all } \sigma, \tau \in G.$$

If $x_\sigma = 0$ satisfies (*), then $S_G = \{0\}$, since $\tau G = G$ for all $\tau \in G$. This is the reason that we exclude $0 \in S_G$.

Theorem 61.1. *Let K be a field and G a finite group of automorphisms of K. Then a set $S = \{x_\sigma \in K \mid \sigma \in G\}$ is a solution set to the Noether equations if and only if there exists an element y in K satisfying $x_\sigma = y/\sigma(y)$ for all $\sigma \in G$.*

The proof is almost identical to the proof of Hilbert 90 60.19.

Proof. If y in $K^\times$ satisfies $x_\sigma = y/\sigma(y)$ for all $\sigma \in G$, then

$$\frac{y}{\sigma(y)}\,\sigma\Big(\frac{y}{\tau(y)}\Big) = \frac{y}{\sigma(y)}\,\frac{\sigma(y)}{\sigma\tau(y)} = \frac{y}{\sigma\tau(y)}$$

for all $\sigma, \tau \in G$. So $S = \{y/\sigma(y) \mid \sigma \in G\}$ is a solution set to the Noether equations.

Conversely, suppose that S is a solution set to the Noether equations. By Dedekind's Lemma 54.3, G is an independent set, so there exists an element $z \in K^\times$ satisfying $0 \neq y = \sum_{\tau \in G} x_\tau \tau(z)$. Hence we have $\sigma(y) = \sum_{\tau \in G} \sigma(x_\tau)\sigma\tau(z)$. Multiply this equation by x_σ. Using (*) then yields

$$x_\sigma \sigma(y) = \sum_{\tau \in G} x_\sigma \sigma(x_\tau)\sigma\tau(z) = \sum_{\tau \in G} x_\tau \tau(z) = y.$$

Therefore, $x_\sigma = y/\sigma(y)$ for all $\sigma \in G$. $\square$

We now construct the general setup where the Noether equations constitute a special case.

Construction 61.2. Let G be a multiplicative group, A a (multiplicative) abelian group, and $G \to \mathrm{Aut}(A)$ a group homomorphism. This defines a G-action on A by evaluation that we write by $\sigma(a)$ for $\sigma \in G$ and $a \in A$. We call a map $f : G \to A$ a *crossed homomorphism*, if

$$f(\sigma\tau) = f(\sigma)\,\sigma\bigl(f(\tau)\bigr)$$

for all $\sigma, \tau \in G$. In particular, if $G \to \mathrm{Aut}(A)$ is the trivial map, the set of crossed homomorphisms is just $\mathrm{Hom}(G, A)$, the group of group homomorphism $G \to A$. A crossed homomorphism $f : G \to A$ is called a *principal crossed homomorphism* if there exists a $y \in A$ satisfying $f(\sigma) = \sigma(y)/y$ for all $\sigma \in G$.

Set

$$\begin{aligned} Z^1(G, A) &:= \{f : G \to A \mid f \text{ a crossed homomorphism}\} \\ B^1(G, A) &:= \{f : G \to A \mid f \text{ a principal crossed homomorphism}\}. \end{aligned}$$

Then $Z^1(G, A)$ is an abelian group and $B^1(G, A)$ is a subgroup (whose proof we leave as an exercise). The quotient group

$$H^1(G, A) = Z^1(G, A)/B^1(G, A)$$

is called the 1*st cohomomology group of* G *with coefficients in* A.

Corollary 61.3. (Hilbert Theorem 90) *Let* L/F *be a finite Galois extension. Then*

$$H^1\bigl(G(L/F), L^\times\bigr) = 1.$$

Proof. This is just a reformulation of Theorem 61.1. □

It can be shown, if L/F is any Galois extension, that

$$H^1_{cont}\bigl((G(L/F), L^\times\bigr) = 1,$$

where all the crossed homomorphisms defined above must also be continuous relative to the topology introduced in Section 58. In particular, if L/F is finite, $G(L/F)$ has the discrete topology and this agrees with the corollary.

Exercises 61.4.

1. Let L/F be a finite Galois extension. Let $y \in K$ satisfy $x_\sigma = y/\sigma(y)$ for all $x \in L^\times$. Let r be the least common multiple of the orders of the elements of $G(L/F)$. The y^r lies in F.
2. Let G be a group and A a (multiplicative) abelian group. Show that $Z^1(G, A)$ and $B^1(G, A)$ are abelian groups.
3. Let L/F be a finite Galois extension and suppose that the group μ_n of nth roots of unity lies in F and consists of n elements. Then show that $B^1(G(L/F), \mu_n) = 1$ and $Z^1(G(L/F), \mu_n) = \mathrm{Hom}(G(L/F), \mu_n)$, i.e., $H^1(G(L/F), L^\times) = \mathrm{Hom}(G(L/F), \mu_n)$.

62. Addendum: Kummer Theory

In Section 60, we proved Theorem 60.20 classifying finite cyclic field extensions, i.e., finite Galois extensions with cyclic Galois group, of degree n when the base field contained all the nth roots of unity and the characteristic of the field was zero or did not divide n. Exercise 60.26(8) extends this to finite abelian extension of a field. In this section, we generalize this to arbitrary abelian extensions of a field, i.e., not necessarily finite. To do this, we shall need results from Section 58.

We begin with a preliminary discussion of pairing of groups. Let A, B, and C be abelian multiplicative groups. A *pairing* $\langle\ ,\ \rangle : A \times B \to C$ is a map that is a group homomorphism in each variable. Set

$$\begin{aligned} A^\perp &:= \{a \in A \mid \langle a, x\rangle = e_C \ \text{ for all } x \in B\} \\ B^\perp &:= \{b \in B \mid \langle y, b\rangle = e_C \ \text{ for all } y \in A\}. \end{aligned}$$

This pairing induces a homomorphism $A \to \operatorname{Hom}(B, C)$ by $a \mapsto \langle a,\quad \rangle$, hence a monomorphism $A/A^\perp \to \operatorname{Hom}(B/B^\perp, C)$. Similarly, this pairing induces a homomorphism $B \to \operatorname{Hom}(A, C)$ by $b \mapsto \langle\quad, b\rangle$, hence a monomorphism $B/B^\perp \to \operatorname{Hom}(A/A^\perp, C)$.

Suppose m is a positive integer such that both $A/A^\perp$ and $B/B^\perp$ satisfy $x^m = e_A$ and $y^m = e_B$ for all $x \in A$ and for all $y \in B$, and $C = \mu_m$ is the cyclic group of order m, i.e., m divides the exponent of these three groups, where the *exponent* N of a (multiplicative) group G is the least positive integer N such that $g^N = e_G$ for all $g \in G$ (if such exists). Then in the above,

$$A/A^\perp \ \text{ is finite if and only if } \ B/B^\perp \ \text{ is finite,} \tag{62.1}$$

and if this is the case, then we have an isomorphism:

$$A/A^\perp \xrightarrow{\sim} \widehat{B/B^\perp} := \operatorname{Hom}(B/B^\perp, \mu_m). \tag{62.2}$$

We shall apply this when μ_m is the group of mth roots of unity in a field F of characteristic zero or characteristic not dividing m.

We set up notation to be used in the rest of this section. Let F be a field. Denote by $\widetilde{F}$ a fixed algebraic closure of F and assume all algebraic extensions of F lie in it. So the group of mth roots of unity μ_m lie $\widetilde{F}$. We shall assume that $|\mu_m| = m$, so the characteristic of F is zero or does not divide m. We shall also assume that $\mu_m \subset F$. Therefore, if K/F is a splitting field of $t^m - a \in F[t]$, then $K = F(\alpha)$ with $\alpha^m = a$ and the m distinct roots of $t^m - a$ are given by $\zeta\alpha$ for $\zeta \in \mu_m$ by Theorem 60.20. We write this symbolically as $K = F(a^{\frac{1}{m}})$. Let $(F^\times)^m = \{x^m \mid x \in F\}$ and if $(F^\times)^m \subset P \subset F^\times$ is a subgroup, let $F_P = F(P^{\frac{1}{m}})$ be the field generated by the $F(a^{\frac{1}{m}})$ with $a \in P$. The (possibly infinite) field extension F_P/F is normal and separable, hence is Galois by Theorem 58.5.

Definition 62.3. Let K/F be a (possibly infinite) Galois extension of fields. We say that K/F has *exponent* m if $G(K/F)$ has finite exponent dividing m. Suppose, in addition, that $\mu_m \subset F$ and $|\mu_m| = m$. We say that K/F is a *Kummer extension of exponent* m if K/F is abelian of exponent m, i.e., Galois with abelian Galois group.

Let F be a field containing the mth roots of unity μ_m with $|\mu_m| = m$. Set

$$\mathcal{F}_m := \{K \mid K/F \text{ is a Kummer extension of exponent } m\}$$
$$\mathcal{G}_m^{ab} := \{P \mid (F^\times)^m \subset P \subset F^\times \text{ is a subgroup}\}.$$

Theorem 62.4. *Let F be a field containing the mth roots of unity μ_m with $|\mu_m| = m$. Then the map $i : \mathcal{G}_m \to \mathcal{F}_m$ given by $P \mapsto F_P$ is a bijection. Moreover, F_P/F is a finite extension if and only if $P/(F^\times)^m$ is a finite group, and if finite, then $[F : F_P] = [P : (F^\times)^m]$.*

PROOF. Let $P \in \mathcal{G}_m$. We first show that $F_P \in \mathcal{F}_m$. Let $a \in P$, so $a = \alpha^m$, some $\alpha \in \widetilde{F}$. In the proof of Theorem 60.20 using Hilbert Theorem 90 (Lemma 60.19), we saw that if $\sigma \in G(F_P/F)$, then there exists a unique $\zeta_\sigma \in \mu_m$ satisfying $\sigma(\alpha) = \zeta_\sigma\alpha$. Check that ζ_σ is independent of the choice of α. In particular, $\sigma^m(\alpha) = \alpha$ and if $\tau \in G(F_P/F)$, then

$$\sigma\tau(\alpha) = \zeta_\sigma\zeta_\tau = \tau\sigma(\alpha).$$

It follows that F_P/F is an abelian extension. Since $F_P = F(\{\alpha \in \widetilde{F} \mid \alpha^m \in P\}$, we have $F_P \in \mathcal{F}_m$.

Next let $K \in \mathcal{F}_m$. Set

$$N := \{a \in F^\times \mid \text{there exists } \ \alpha \in K \text{ such that } \alpha^m = a\} \in \mathcal{G}_m.$$

Let $G = G(K/F)$. Define

$$f : N \to \widehat{G} := \mathrm{Hom}(G, \mu_m)$$

as follows: Let $a \in N$. Choose $\alpha \in K$ satisfying $\alpha^m = a$. If $\sigma \in G$, then, as above, there exists a unique $\zeta_\sigma \in \mu_m$ satisfying $\sigma(\alpha) = \zeta_\sigma\alpha$. Define $f_a : N \to \widehat{G} = \mathrm{Hom}(G, \mu_m)$ by $\sigma \mapsto \sigma(\alpha)/\alpha = \zeta_\sigma$.

Claim 62.5. *f_a is a well-defined homomorphism.*

If f_a is well-defined, it is clearly a homomorphism. So we must show that f_a is independent of the choice of α. We leave its proof as an exercise. By the claim we obtain a map $f : N \to \widehat{G}$ given by $a \mapsto f_a$.

We apply this construction to $K = F_P$ with $P \in \mathcal{G}_m$. So we have a pairing

$$\langle\ ,\ \rangle : G \times P \to \mu_m$$

defined by $\langle \sigma, a\rangle = f_a(\sigma) = \sigma(\alpha)/\alpha$ where $a = \alpha^m$.

Claim 62.6. *$G^\perp = 1$ and $P^\perp = (F^\times)^m$.*

That $G^\perp = 1$ is easy, so we turn to $P^\perp$. Suppose that $\alpha \in K = F_P$ satisfies $\alpha^m = \alpha$ and $\langle \sigma, a\rangle = \sigma(\alpha)/\alpha = 1$ for all $\sigma \in G$, but $\alpha \notin F$. Then there exists a $\tau \in G$ such that $\tau(\alpha) \neq \alpha$. Hence $\langle \tau, a\rangle = \tau(\alpha)/\alpha \neq 1$, a contradiction. The Claim now follows easily.

We also conclude by (62.1) that G is finite, i.e., F_P/F is finite, if and only if $P/(F^\times)^m$ is finite. If this is the case then by (62.2), we have

$$[F : F_P] = [P : (F^\times)^m] \ \text{ and } \ P/(F^\times)^m/(F^\times)^m \cong \widehat{G(F_P/F)}. \tag{62.7}$$

This is the crucial observation.

Claim 62.8. *$P_1 \subset P_2$ in $\mathcal{G}_m$ if and only if $F_{P_1} \subset F_{P_2}$. In particular, $i : \mathcal{G}_m \to \mathcal{F}_m$ is injective.*

The only if statement is clear, so suppose that $F_{P_1} \subset F_{P_2}$. Let $\alpha \in P_1$. Then α lies in F_{P_2}, hence there exists a finitely generated extension $K/F(F^m)$ with F_{P_2}/K and $\alpha \in K$. It follows that we may assume that $F_{P_2}/F(F^m)$ is finite generated, hence a finite field extension. Let P_3 be the group generated by α and P_2. Then $F(P_2) = F(P_3)$, and $[F_{P_2} : F^m] = [F_{P_3} : F^m]$ by (62.7). Hence $P_2 = P_3$. This proves the claim.

So we are left to show that $i : \mathcal{G}_m \to \mathcal{F}_m$ is surjective. Let $K \in \mathcal{F}_m$, say $K = F(P)$ with $P \in \mathcal{G}_m$. For any element x in K, there exists a finite subgroup $P_0 \in \mathcal{G}_m$ such that x lies in $F(P_0)$, i.e., K is the compositum of finite abelian extensions. Therefore, we may assume that K/F is finite abelian extension with $G = G(K/F)$ of finite exponent dividing m. Since G is a finite abelian group of finite exponent dviding m, it is a product of finitely many cyclic groups of exponent dividing m by Theorem 44.6 (or Proposition 14.9). Consequently, $G \cong \widehat{G}$ by Exercise 44.24(10). By Exercise 57.13(13), this reduces us to the case that K/F is finite cyclic which has been done. □

Remark 62.9. Let F contain μ_m with $|\mu_m| = m$ as in the theorem. In the notation of the theorem, if we let $P \in \mathcal{G}_m$ and $G = G(F_P/F)$, then the map

$$f : P \to \widehat{G} \;\text{ given by }\; a \mapsto (f_a : \sigma \to \sigma(\alpha)/\alpha) \;\text{ with }\; \alpha^m = a \tag{62.10}$$

induces a monomorphism $\overline{f} : F/(F^\times)^m \to \widehat{G}$. If we take continuous homomorphisms $G \to \mu_m$ in the Krull topology on $G(\widetilde{F}/F)$, then this map is an isomorphism. [As μ_m lies in F, its topology is discrete.] We leave this as an exercise. Write this as

$$1 \to (F^\times)^m \to P \to \mathrm{Hom}_{\mathrm{cont}}(G, \mu_m) \to 1 \tag{62.11}$$

is exact where $\mathrm{Hom}_{\mathrm{cont}}(G, \mu_m)$ denotes the group of continuous maps $G \to \mu_m$.

More generally, let F_{sep} denote the *separable closure of* F in $\widetilde{F}$, i.e., the maximal separable extension of F in $\widetilde{F}$. Then $\Gamma_F := G(\mathrm{F}_{sep}/F)$ is $G(\widetilde{F}/F)$ and called the *absolute Galois group of* F. If $\mu_m \subset F$ and has m elements, then we have an exact sequence

$$1 \to \mu_m \to F_{\mathrm{sep}}^\times \xrightarrow{g} F_{\mathrm{sep}}^\times \to 1.$$

where $g(x) = x^m$ By using 'group cohomology' — a subject that we shall not discuss — one obtains an exact sequence

$$1 \to \mu_m \to F^\times \xrightarrow{g} (F^\times) \xrightarrow{\delta} \mathrm{Hom}_{\mathrm{cont}}(\Gamma_F, \mu_m) \to 1.$$

called the *Kummer sequence* with δ a map coming from group cohomology theory called a connecting map. The surjectivity of δ arises from a homological version of Hilbert Theorem 90, the finite case can be found in Corollary 61.3.

Exercises 62.12.

1. Prove 62.1
2. Prove that ζ_σ in the proof of the theorem is independent of the choice of α.
3. Prove Claim 62.5.
4. Prove that (62.10) implies (62.11).

63. Normal Basis Theorem

In this section, we shall show that every finite Galois field extension K/F has an F-basis $\{\sigma(\alpha) \mid \sigma \in G(K/F)\}$. If the extension is cyclic, this is easy to prove using our knowledge of linear algebra (in particular, using facts about the minimal polynomial of a linear operator on a finite dimensional vector space). We have seen that Hilbert Theorem 90 was the crucial fact needed to classify finite cyclic extensions (with appropriate roots of unity). We used the norm map to proved it. An additive version of this result is also true, which we now state.

Lemma 63.1. *Let K/F be a finite, cyclic extension with Galois group $G(K/F) = \langle\sigma\rangle$ and $x \in K$. Then $\operatorname{Tr}_{K/F}(x) = 0$ if and only if there exists a $y \in K$ satisfying $x = y - \sigma(y)$.*

The proof that we gave for the multiplicative version works for this additive version by using the trace in exactly the same way (by taking the log). One can extend the definition of both the norm and trace to an arbitrary finite extension of fields. Unfortunately, the trace becomes useless in the nonseparable case. However, the trace is quite useful in the finite separable case because it is a nontrivial linear functional. In the multiplicative case, a fuller interpretation of the proof of Hilbert Theorem 90 by Noether was that $H^1(G(K/F), K^\times) = 1$ for all finite Galois extensions K/F. We also wish to indicate a stronger result in the additive case.

Construction 63.2. Let K/F be a finite, separable extension with L/K a normal closure of K/F. Let $\sigma_1, \dots, \sigma_n : K \to L$ be all the distinct F-homomorphisms of K into L. Since K/F is separable, L/F is Galois and $n = [K : F]$. Let $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ be an F-basis for K. As $\beta = \sum_{j=1}^n a_j\alpha_j$ for each $\beta \in K$ with $a_j \in F$ for all j, we have $\sigma(\beta) = \sum_{j=1}^n a_j\sigma(\alpha_j)$ lies in L. Recall that we defined the trace of a finite separable field extension in Remark 60.18. Define

$$T : L \times L \to F \text{ by } T(x, y) := \operatorname{Tr}_{K/F}(xy).$$

So

$$T(\alpha_i, \alpha_j) = \sum_{k=1}^{n} \sigma_k(\alpha_i)\sigma_k(\alpha_j).$$

Let

$$D = \big(\sigma_k(\alpha_i)\big) \in \mathbb{M}_n(L).$$

We have $\big(T(\alpha_i\alpha_j)\big) = DD^t$ in $\mathbb{M}_n(L)$. By Dedekind's Lemma 54.3, we know that the system of equations, $\sum_{k=1}^n x_k\sigma_k(\alpha_i) = 0$, $i = 1, \dots, n$, can only have the trivial solution. Therefore, $\big(T(\alpha_i\alpha_j)\big)$ must be invertible by Cramer's Rule, hence D is also invertible. The element $\det\Big(\big(T(\alpha_i\alpha_j)\big)\Big)$ is called a *discriminant* of K/F relative to the basis $\mathcal{B}$.

Application 63.3. Let K/F be finite of degee n and Galois with an F-basis $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ for K and $G(K/F) = \{\sigma_1, \dots, \sigma_n\}$. Then we have seen that $\alpha = \sum_{i=1}^n a_i\alpha_i$ with $a_i \in F$, then $\sigma(\alpha) = \sum_{i=1}^n a_i\sigma_j(\alpha_i)$ and the matrix $\big(\sigma_j(\alpha_i)\big)$ in $\mathbb{M}_n(K)$ is invertible.

Conversely, suppose that $\beta_, \dots, \beta_m$ in K satisfy $\big(\sigma_j(\beta_i)\big) \in \mathbb{M}_n(K)$ is invertible. If $\beta_1, \dots, \beta_m$ were linearly dependent, then the system of equations, $\sum_{i=1}^n \sigma_j(\beta_i)x_i = 0$ with $j = 1, \dots, n$ would have a nontrivial solution. Therefore, $\{\beta_1, \dots, \beta_n\}$ is linearly independent over F hence a basis.

Next, following Artin, we strengthen the notion of independence of characters over an infinite field under certain conditions.

Definition 63.4. Let K be an infinite field and $\sigma_1, \dots, \sigma_n : K \to L$ be (field) homomorphisms. We say that $\sigma_1, \dots, \sigma_n$ are *algebraically independent over* L if $f \in K[t_1, \dots, t_n]$ is a polynomial satisfying

$$f\big(\sigma_1(x), \dots, \sigma_n(x)\big) = 0 \text{ for all } x \in K, \text{ then } f = 0.$$

Since K is an infinite field, we can (and do) identify polynomials and polynomial functions over K by Remark 34.10(4).

Theorem 63.5. *Let F be an infinite field, K/F a finite separable extension, and L/K a normal closure of K/F. Let $\sigma_1, \dots, \sigma_n : K \to L$ be all the distinct F-homomorphisms of K into L. Then $\sigma_1, \dots, \sigma_n$ are algebraically independent over L.*

Proof. As K/F is separable, $[K : F] = n$. Let $\{\alpha_1, \dots, \alpha_n\}$ be an F-basis for K. Suppose that $f \in K[t_1, \dots, t_n]$ is a polynomial satisfying $f\big(\sigma_1(x), \dots, \sigma_n(x)\big) = 0$ for all $x \in K$. Then, for all $a_1, \dots, a_n$ in F^n, we have an equation

$$\begin{aligned} 0 &= f\big(\sigma_1(\sum_{i=1}^n a_i\alpha_i), \dots, \sigma_n(\sum_{i=1}^n a_i\alpha_i)\big) \\ &= f\big(\sum_{i=1}^n a_i\sigma_1(\alpha_i), \dots, \sum_{i=1}^n \sigma_n a_i(\alpha_i)\big) \end{aligned}$$

in L. Define $g \in K[t_1, \dots, t_n]$ by

$$g = \big(\sum_{i=1}^n \sigma_1(\alpha_i)t_i, \dots, \sum_{i=1}^n \sigma_n(\alpha_i)t_i\big).$$

So we have $g(a_1, \dots, a_n) = 0$ for all $a_i \in F$, i.e., $g|_{F^n} = 0$ Let $\mathcal{B} = \{\beta_1, \dots, \beta_m\}$ be an F-basis for L. We can write

$$g = \sum_{j=1}^m \beta_j g_j$$

for appropriate $g_j \in F[t_1, \dots, t_n]$ (as F-linear functionals are polynomial functions). Since $\mathcal{B}$ is a basis and $g(a_1, \dots, a_n) = 0$ for all $a_i \in F$, we conclude that $g_j(x_1, \dots, x_n) = 0$ for all $x_i \in F$. It follows that $g(x_1, \dots, x_n) = 0$ for all $x_i \in K$ i.e., $g|_{K^n} = 0$. Since F is infinite, $g = 0$ (in $F[t_1, \dots, t_n]$).

By Construction 63.2, $\det\big(\sigma_i(\alpha_j)\big) \neq 0$. It follows that $\big(\sigma_i(\alpha_j)\big)$ is invertible. Let (γ_{ij}) be the inverse of $\big(\sigma_i(\alpha_j)\big)$ in $\mathbb{M}_m(L)$. Since $g(t_1, \dots, t_n) = f\big(\sum_i \sigma_1(\alpha_i)t_i, \dots, \sum_i \sigma_n(\alpha_i)t_i\big)$, we have

$$0 = g\big(\sum_{j,k} \gamma_{1j}\sigma_j(\alpha_k)t_k, \dots, \sum_{j,k} \gamma_{nj}\sigma_j(\alpha_k)t_k\big) = f(t_1, \dots, t_n).$$

Since $g = 0$, we have $f = 0$ also. the result follows. $\square$

Definition 63.6. Let K/F be a finite Galois extension. An F-basis $\mathcal{B}$ for K is called a *normal basis* if there exists an element α in K such that $\mathcal{B} = \{\sigma(\alpha) \mid \sigma \in G(K/F)\}$.

Theorem 63.7. (Normal Basis Theorem) *Let K/F be a finite Galois extension. Then K has a normal basis.*

Proof. **Case 1**: F is an infinite field.

Let $G(K/F) = \{\sigma_1, \dots, \sigma_n\}$. By the previous theorem, $\sigma_1, \dots, \sigma_n$ are algebraically independent. By Application 63.3, if $\alpha \in K$, then $\{\alpha_i(\alpha), \dots, \sigma_n(\alpha)\}$ is an F-basis if and only if $\det\big(\sigma_i\sigma_j(\alpha)\big) \neq 0$. Write $\sigma_i\sigma_j = \sigma_{r(i,j)}$. For fixed i, the map $j \mapsto r(i,j)$ is a permutation of $1, \dots, n$. Let $A = \big(t_{r(i,j)}\big)$ in $\mathbb{M}_m(K[t_1, \dots, t_n])$. Evaluate A at $(t_1, \dots, t_n) \mapsto (1, 0, \dots, 0)$. Then this matrix is a permutation matrix, hence has determinant ± 1. Since det is a polynomial function on $t_1, \dots, t_n$, we have $\det(A) \neq 0$. By the algebraic independence of $\sigma_1, \dots, \sigma_n$, there exists an element u in K satisfying $\det\big(\sigma_i\sigma_j(u)\big) \neq 0$. Consequently, $\{\sigma_1(u), \dots, \sigma_n(u)\}$ is a normal basis for K.

Case 2: F is a finite field.

We know that K/F is a cyclic extension. We prove the result in this more general situation. Let $G(K/F) = \langle \sigma \rangle$ As K is a finite dimensional vector space over F and $\sigma : K \to K$ is an F-linear operator, it has a minimal polynomial q_σ. We know that $\deg q_\sigma \leq f_\sigma = [K : F]$ (and, in fact, $q_\sigma \mid f_\sigma$) by the Cayley-Hamilton Theorem, where f_σ is the characteristic polynomial of σ. If $n = [K : F] = |G(K/F)|$, then $1, \sigma, \dots, \sigma^{n-1}$ are independent by Dedekind's Lemma. It follows that $\deg q_\sigma = n$. Since q_σ is the monic polynomial of minimal degree satisfying $q_\sigma(\sigma)(u) = 0$ for all u in the vector space K, there must exist an $\alpha \in K$ such that $\{\alpha, \sigma(\alpha), \dots, \sigma^{n-1}(\alpha)\}$ is a basis. $\square$

We showed that the cohomological version of Hilbert Theorem 90 is that $H^1(G(K/F), K^\times) = 1$ for any finite Galois extension K/F. In general, as we shall see that $H^2(G(K/F), K^\times) \neq 1$. [For example, $H^2(G(\mathbb{C}/\mathbb{R}) = 2.$] We have also seen that the additive version of Hilbert Theorem 90 $H^1(G(K/F), K^+) = 0$ is also true. Using homological methods, one can show that the Normal Basis theorem implies that $H^n(G(K/F), K^+) = 0$ for all $n > 0$.

Exercise 63.8. Let K/F be a finite separable extension with bases $\mathcal{B}$ and $\mathcal{C}$. Show that the discriminants of K/F relative to $\mathcal{B}$ and $\mathcal{C}$ respectively differ by a square in F.

64. Addendum: Galois' Theorem

In this section, we establish Galois' characterization of the Galois group of an irreducible polynomial of prime degree over a field of characteristic zero. We begin with a lemma, previously given as Exercise 22.15 (23). If p is a prime, then a finite group is called an *elementary p-group* if it is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$ for positive integer n. Note that any elementary p-group can be viewed as a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$. A *minimal normal* subgroup of a nontrivial group G is a normal subgroup of G such that the trivial group is the only normal subgroup of G properly contained in it.

Lemma 64.1. *Let G be a nontrivial finite solvable group. Then any minimal normal subgroup of G is an elementary p-group.*

Proof. Let H be a minimal normal subgroup of G. As H is a subgroup of a solvable group, it is solvable. We also know that any characteristic subgroup of H is normal in G (cf. Exercise 11.9(22)). Since the commutator subgroup of any group is a characteristic subgroup, the series obtained by successively taking commutators gives a characteristic series for H (cf. Proposition 17.5). As H is solvable, it contains a characteristic nontrivial abelian normal subgroup, so we may assume H is abelian. Since Sylow groups of an abelian group are characteristic, we may assume that H is a p-group. Since $\{x \in H \mid x^p = 1\}$ is easily seen to be a characteristic subgroup of H, the lemma follows. □

More generally, it can be shown that a minimal normal subgroup of a nontrivial finite group is either simple or a product of simple groups all which are isomorphic.

Let p be a prime and V a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$. If v_0 is a vector in V, then the map $\tau_{v_0} : V \to V$ given by $v \mapsto v + v_0$ is called *translation* by v_0. [Note that it is not an F-linear transformation unless $v_0 = 0$.] The set of all translations of V form a subgroup of the the *group of affine transformations* of V, the group defined by

$$\operatorname{Aff}(V) := \{\tau : V \to V \mid \tau : v \mapsto av + v_0 \text{ with } a \text{ in } \mathbb{Z}/p\mathbb{Z} \text{ and } v_0 \text{ in } V\}.$$

So every element in $\operatorname{Aff}(V)$ is a composition $\tau\lambda_a$ with τ a translation and $\lambda_a : V \to V$ the F-linear map given by $x \mapsto ax$ for some a in F.

Theorem 64.2. *Let p be a prime and S a set with p elements. Suppose that G is a transitive subgroup of $\Sigma(S) \cong S_p$. Then the following are equivalent:*

(1) G is solvable.
(2) Each non-identity element in G fixes at most one element in S.
(3) There exists a transitive normal subgroup T of G satisfying $|T| = p$ and T is its own centralizer in G, i.e., $T = Z_G(T)$.
(4) G is isomorphic to a subgroup of the group of affine transformations $\operatorname{Aff}(\mathbb{Z}/p\mathbb{Z})$ and contains the subgroup of translations.

Moreover, if G is solvable and if T is as in (3), then there exists a cyclic subgroup H in G satisfying $G = HT$, $H \cap T = 1$, and $|H| \mid p - 1$. Furthermore, any non-identity normal subgroup of G is of the form $H'T$ for some subgroup H' of H.

Proof. We may assume that $G \subset S_p$, so $S = \{1, \dots, p\}$. As G is transitive, the orbit Gs for any $s \in S$, has order p, so $p \mid |G|$ and all the Sylow p-groups of G are cyclic of order p, i.e., is generated by a p-cycle.

(1) $\Rightarrow$ (3): Let N be a normal subgroup of G.

Claim. The group N acts transitively on S_p. In particular, $p \mid |N|$:

Let x and y be elements in S. As G acts transitively on S, there exists an element σ in G satisfying $y = \sigma(x)$. As the map $Nx \to Ny$ given by $zx \mapsto z\sigma(x)$ is a well-defined bijection, we see that all orbits of N have the same number of elements. If $\mathcal{O}$ is a system of representatives for the action of N on S, then $p = |S| = \sum_{\mathcal{O}} |Nx| = \sum_{\mathcal{O}} [N : N_x]$, by the Orbit Decomposition Theorem 19.9. So either S has one orbit under the action of N, i.e., N is transitive, or every orbit of S under the action of N has one point, i.e., N is the identity subgroup. This establishes the claim.

Let T be a minimal normal subgroup of G. By the lemma and claim, T is a cyclic subgroup of order p, hence is generated by a p-cycle, say $T = \langle\sigma\rangle$. The normality of T implies that it is the unique Sylow p-subgroup of G. Suppose τ centralizes T, i.e., lies in $Z_G(T)$. If

$\tau(s) = s$ for some $s \in S$, then $\tau\sigma(s) = \sigma\tau(s) = \sigma(s)$, hence $\tau\sigma^i(s) = \sigma^i\tau(s) = \sigma^i(s)$ for all i implying $\tau = 1$. Therefore τ is either 1 or is fixed point-free, so a p-cycle.

(3) $\Rightarrow$ (4): Suppose that $T = \langle\sigma\rangle$ is a cyclic subgroup of G of order p. We may assume that $S = \mathbb{Z}/p\mathbb{Z}$ and σ is the p-cycle $(\overline{0}\,\overline{1}\cdots\overline{p-1})$ where $^{-} : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ is the canonical epimorphism. Therefore, σ generates the group of translations in $\mathrm{Aff}(\mathbb{Z}/p\mathbb{Z})$. Let τ be an element in G. Then there exists an integer n with $p \not| \, n$, unique modulo p, such $\tau\sigma\tau^{-1} = \sigma^n$, i.e., we have a well-defined map $\varphi : G \to (\mathbb{Z}/p\mathbb{Z})^\times$ given by $\tau \mapsto n$. Clearly, this map is a group homomorphism with $\ker\varphi = Z_G(T)$. As $\tau\sigma = \sigma^{\varphi(\tau)}\tau$, we have

$$\tau(x + \overline{1}) = \tau\sigma(x) = \sigma^{\varphi(\tau)}\tau(x) = \tau(x) + \varphi(\tau).$$

In particular, $\tau(\overline{1}) = \tau(\overline{0}) + \varphi(\tau) = \varphi(\tau) + b$ with $b = \tau(\overline{0})$. Inductively, $\tau(x) = \varphi(\tau)x + b$. This yields (4).

(4) $\Rightarrow$ (1): The translations form a cyclic normal subgroup T in G with quotient embedding in $(\mathbb{Z}/p\mathbb{Z})^\times$.

(4) $\Rightarrow$ (2): The equation

$$ax + b \equiv x \mod p,$$

with a and b integers, has a unique solution modulo p if either $a \not\equiv 1 \mod p$ or $b \not\equiv 0 \mod p$. It follows that any non-identity element in G can fix at most one point.

(2) $\Rightarrow$ (3): We apply Exercise 21.25(17) — known as the *Burnside Counting Theorem* – to see that

$$|G| = \sum_G |F_\tau(S)| \text{ where } F_\tau(S) := F_{\langle\tau\rangle}(S).$$

As $\tau = 1$ fixes p points, we have $\sum_{G\setminus\{1\}} |F_\tau(S)| = |G| - p$. By hypothesis, if $\tau \neq 1$ then $|F_\tau(S)| \leq 1$, so we must have precisely $p-1$ non-identity elements τ in G satisfying $|F_\tau(S)| = 0$, i.e., $p-1$ elements in G have no fixed points in S. If σ is one such, then so is σ^i for $i = 1, \ldots, p-1$. It follows that σ is a p-cycle and $T = \langle\sigma\rangle$, a group of order p, contains all the p-cycles in G. As the conjugate of a p-cycle is a p-cycle, $T \triangleleft G$. Suppose that τ lies in $Z_G(T)$. If $F_\tau(S)$ is empty, then τ lies in T, so assume that $F_\tau(S)$ is not empty, say $\tau(x) = x$. It follows that $\tau\sigma(x) = \sigma\tau(x) = \sigma(x)$, i.e., both x and $\sigma(x)$ lie in $F_\tau(S)$. By assumption this means $\tau = 1$, so $T = Z_G(T)$, establishing (3).

For the last two statements, let G satisfy (4). Set

$$H = \{\tau \in G \mid \tau(x) \equiv ax \mod p \text{ with } a \text{ an integer satisfying } p \not| \, a\}.$$

Clearly, $H \cap T = 1$ and $HT = G$. Certainly, we have a monomorphism $H \to (\mathbb{Z}/p\mathbb{Z})^\times$, so H is cyclic and $|H| \mid p-1$. If $N \triangleleft G$ is not the trivial group, then by the claim, N acts transitively on S and contains T, the unique Sylow p-subgroup of G. It follows that $N = H'T$ for some subgroup H' of H by the Correspondence Principle. $\square$

Theorem 64.3. (Galois) *Suppose that F is a field of characteristic zero and f an irreducible polynomial in $F[t]$ of prime degree p. Let K be a splitting field of f over F. Then f is solvable by radicals if and only if $K = F(\alpha, \beta)$ for any two roots of f in K. Moreover, if f is solvable by radicals, $G(K/F)$ is isomorphic to a subgroup of* $\mathrm{Aff}(\mathbb{Z}/p\mathbb{Z})$ *containing the subgroup of translations, and there exists an intermediate field $K/E/F$ with E/F cyclic, $[K : E] = p$. Furthermore, if $K/L/F$ is an intermediate field with L/F normal with $F < L$, then $L \subset E$.*

PROOF. We know that $G(K/F)$ can be viewed as a transitive subgroup of S_p and f is solvable by radicals if and only if $G(K/F)$ is solvable. By the previous theorem, $G(K/F)$ is solvable if and only if no non-identity element in $G(K/F)$ fixes two roots of f. Let α and β be two roots of f in K. By the Fundamental Theorem of Galois Theory 57.3, we know that $K = F(\alpha, \beta)$ if and only if $G\big(K/F(\alpha, \beta)\big) = 1$ if and only if no non-identity element of $G(K/F)$ fixes α and β. The result now easily follows. $\square$

65. The Discriminant of a Polynomial

Let R be a commutative ring and let the symmetric group S_n act on $R[t_1, \ldots, t_n]$ by $\sigma t_i = t_{\sigma(i)}$ for all i and $\sigma \in S_n$. This induces a group monomorphism $S_n \to \mathrm{Aut}(R[t_1, \ldots, t_n])$ by viewing $\sigma \in S_n$ as a ring homomorphism fixing R. Analogous to the field case of $F(t_1, \ldots, t_n)^{S_n}$, one can show that the set of elements $R[t_1, \ldots, t_n]^{S_n}$ in $R[t_1, \ldots, t_n]$ fixed by S_n, is equal to $R[s_1(t_1, \ldots, t_n), \ldots, s_n(t_1, \ldots, t_n)]$, where

$$s_j(t_1, \ldots, t_n) = \sum_{1 \le i_1 < \cdots < i_j \le n} t_{i_1} \cdots t_{i_j}$$

with $1 \le j \le n$ are the elementary symmetric functions in $t_1, \ldots, t_n$. [Cf. the Fundamental Theorem of Symmetric Polynomials 72.4.] Let $D = \prod_{i<j}(t_i - t_j)$ in $\mathbb{Z}[t_1, \ldots, t_n]$. If σ is the transposition (ij) in S_n, then $\sigma D = -D$. It follows if σ lies in S_n, then

$$\sigma(D) = \begin{cases} -D & \text{if } \sigma \notin A_n \\ D & \text{if } \sigma \in A_n. \end{cases}$$

Let F be a field of characteristic different from two, f a polynomial in $F[t]$, and K/F a splitting field of f. Suppose that $\deg f = n$ and f has no multiple roots in K, say the roots are $\alpha_1, \ldots, \alpha_n$. Hence K/F is Galois and the Galois group $G(K/F)$ permutes $\alpha_1, \ldots, \alpha_n$, so defines a monomorphism $G(K/F) \to S_n$. If σ lies in $G(K/F)$, we have a commutative diagram:

$$\begin{tikzcd} F[t_1, \ldots, t_n] \arrow[r, "\widehat{\sigma}"] \arrow[d, "e_{\alpha_1, \ldots, \alpha_n}"'] & F[t_1, \ldots, t_n] \arrow[d, "e_{\alpha_1, \ldots, \alpha_n}"] \\ K \arrow[r, "\sigma"'] & K, \end{tikzcd}$$

where $\widehat{\sigma}$ is the automorphism of $F[t_1, \ldots, t_n]$ induced by $t_i \mapsto t_{\sigma(i)}$ for $i = 1, \ldots, n$, and $e_{\alpha_1, \ldots, \alpha_n}$ is the obvious evaluation map. It follows that

$$d := e_{\alpha_1, \ldots, \alpha_n}(D) = \prod_{i<j}(\alpha_i - \alpha_j) \neq 0$$

satisfies

$$\sigma(d) = \begin{cases} -d & \text{if } \sigma \notin A_n \\ d & \text{if } \sigma \in A_n. \end{cases}$$

Therefore, $G\big(K/F(d)\big) = G(K/F) \cap A_n$. Let

$$\Delta := d^2 = \prod_{i<j}(\alpha_i - \alpha_j)^2.$$

Then $\sigma(\Delta) = \Delta$ for every σ in $G(K/F)$, hence $d^2 = \Delta$ lies in $K^{G(K/F)} = F$ and is called the *discriminant* of f when f is monic. [In general, if $\mathrm{lead}(f) = a_n$, then $a_n^{2n-2}d^2$ is called the discriminant of f.] As

$$[G(K/F) : G(K/F(d)] \leq [S_n : A_n] = 2,$$

either $F = F(d)$ or $F(d)/F$ is of degree two, as $d \in F$ if and only if $\Delta \in F^2$ if and only if $G(K/F)$ is a subgroup of A_n (where we view $G(K/F) \to S_n$ as an inclusion).

Computation 65.1. In the setup above, let $s_i = s_i(\alpha_1, \dots, \alpha_n)$. As S_n fixes each $s_j(t_1, \dots, t_n)$, we have each s_i lies in $K^{G(K/F)} = F$. By linear algebra, the *Vandermonde determinant*

(65.2)
$$\det \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_n \\ \vdots & & \vdots \\ \alpha_1^{n-1} & \cdots & \alpha_n^{n-1} \end{pmatrix} = \prod_{i>j} (\alpha_i - \alpha_j) = d,$$

so if

(65.3)
$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_n \\ \vdots & & \vdots \\ \alpha_1^{n-1} & \cdots & \alpha_n^{n-1} \end{pmatrix},$$

is the *Vandermonde matrix* , we have $\det AA^t = d^2 = \Delta$.

Check 65.4. Let $e_i := \sum_{j=1}^n \alpha_j^i$ for $i = 1, \dots, 2n-2$. Then

$$AA^t = \begin{pmatrix} n & e_1 & e_2 & \cdots & e_{n-1} \\ e_1 & e_2 & e_3 & \cdots & e_n \\ \vdots & & & & \vdots \\ e_{n-1} & e_n & e_{n+1} & \cdots & e_{2n-2} \end{pmatrix}.$$

As $\sum_{j=1}^n t_j^i$ lies in $F[t_1, \dots, t_n]^{S_n}$, for every positive integer i, each $\sum_{j=1}^n t_j^i$ is a polynomial in the $s_i(t_1, \dots, t_n)$'s (in fact, with integer coefficients), hence each e_i lies in $F[s_1, \dots, s_n]$. We can, therefore, compute formulas for Δ for specific n.

Example 65.5. In the computation and the notation there, we let $n = 2$. Then

$$f = (t - \alpha_1)(t - \alpha_2) = t^2 - (\alpha_1 + \alpha_2)t + \alpha_1\alpha_2 = t^2 - s_1 t + s_2.$$

So $s_1 = \alpha_1 + \alpha_2$ and $s_2 = \alpha_1\alpha_2$, hence $e_1 = s_1 = \alpha_1 + \alpha_2$ and $e_2 = \alpha_1^2 + \alpha_2^2 = (\alpha_1 + \alpha_2)^2 - 2\alpha_1\alpha_2 = s_1^2 - 2s_2$. It follows that

$$\Delta = \det \begin{pmatrix} 2 & e_1 \\ e_1 & e_2 \end{pmatrix} = 2e_2 - e_1^2 = 2(s_1^2 - 2s_2) - s_1^2 = s_1^2 - 4s_2.$$

In general, if f is a polynomial of degree n and splits over K with the roots $\alpha_1, \dots, \alpha_n$ (possibly not distinct), we define the discriminant of f to be $\Delta = \prod_{i<j} (\alpha_i - \alpha_j)^2$.

If one knows a root of an irreducible polynomial, then we can express the discriminant of it in another way.

Proposition 65.6. *Let f be an irreducible and separable polynomial in $F[t]$ of degree n, L/F a splitting field of f, $\alpha = \alpha_1, \alpha_2, \dots, \alpha_n$ the (distinct) roots of f in L, $K = F(\alpha)$, and $T_\alpha : L \to L$ the F-linear isomorphism given by $x \mapsto \alpha x$. Then the following are true:*

(1) *$\alpha_1, \dots, \alpha_n$ are the eigenvalues of T_α. In particular, T_α is diagonalizable.*
(2) *$\alpha_1^k, \dots, \alpha_n^k$ are the eigenvalues of T_{α^k} for all positive integers k and T_{α^k} is diagonalizable.*
(3) *If $f = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $F[t]$, then $f_{T_\alpha} = f^{[L:K]}$, where f_{T_α} is the characteristic polynomial of T_α. In particular,*
   (a) $\mathrm{Tr}_{K/F}(\alpha) = -a_{n-1}$.
   (b) $\mathrm{N}_{K/F}(\alpha) = (-1)^n a_0$.
(4) *Let A be the Vandermonde matrix 65.3. Then the (ij)th term of AA^t satisfies*
$$(AA^t)_{ij} = \mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1}).$$
In particular, $\det\big(\mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1})\big) = \Delta(\alpha)$, *hence nonzero.*
(5) *We have*
$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}} \mathrm{N}_{K/F}\big(f'(\alpha)\big).$$

Proof. (1) and (2) are left as new exercises and (3) is Exercise 60.26(3).
(4): Using (2), we have $\mathrm{Tr}_{K/F}(\alpha^k) = \alpha_1^k + \cdots + \alpha_n^k$ for all $k \geq 1$, hence
$$(AA^t)_{ij} = \sum_{l=1}^{n} \alpha_l^{i-1}\alpha_l^{j-1} = \sum_{l=1}^{n} \alpha^{i+j-2} = \mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1}).$$
(5): We have $\det A = \prod_{i>j}(\alpha_i - \alpha_j)$ and there exist $n(n-1)/2$ pairs of integers (i,j) with $1 \leq j < i \leq n$ in the product $\Delta(\alpha) = \prod_{i>j}(\alpha_i - \alpha_j)^2$. As $(\alpha_i - \alpha_j)^2 = -(\alpha_i - \alpha_j)(\alpha_j - \alpha_i)$, it follows that
$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j).$$
Since $f = \prod_{i=1}^{n}(t - \alpha_i)$ in $L[t]$, we have
$$f' = \sum_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (t - \alpha_j), \quad \text{so} \quad f'(\alpha_i) = \prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j)$$
and
$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} f'(\alpha_i).$$
Let $\sigma_i : K \to L$ be the F-homomorphisms satisfying $\alpha \mapsto \alpha_i$ for $i = 1, \dots, n$. Then $\sigma_i\big(f'(\alpha)\big) = f'(\alpha_i)$, hence
$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}} \mathrm{N}_{K/F}\big(f'(\alpha)\big)$$
as claimed. □

We use the above in the following computation:

Example 65.7. In Proposition 65.6, set $F = \mathbb{Q}$ and $L = K = \mathbb{Q}(\zeta)$, where ζ a primitive pth root of unity in $\mathbb{C}$ with p an odd prime. The pth cyclotomic polynomial $\Phi_p = t^{p-1} + \cdots + 1 = m_{\mathbb{Q}}(\zeta)$. So by Proposition 65.6(5), we have

$$\Delta(\zeta) = (-1)^{\frac{(p-1)(p-2)}{2}} \mathrm{N}_{K/F}\left(\Phi_p'(\zeta)\right) = (-1)^{\frac{p-1}{2}} \mathrm{N}_{K/F}\left(\Phi_p'(\zeta)\right).$$

We compute the right hand side of this equation. Since $t^p - 1 = (t-1)\Phi_p$, taking the derivative yields $pt^{p-1} = \Phi_p + (t-1)\Phi_p'$. Hence

$$\Phi_p'(\zeta) = \frac{p\zeta^{p-1}}{\zeta - 1}.$$

By Proposition 65.6(5),

$$\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}\left(\Phi_p'(\zeta)\right) = \frac{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(p\zeta^{-1})}{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\zeta - 1)} = \frac{p^{p-1} \cdot 1}{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\zeta - 1)}.$$

As $\prod_{i=1}^{p-1}(t - \zeta^i) = t^{p-1} + \cdots + t + 1$, we have

$$\prod_{i=1}^{p-1}(\zeta^i - 1) = (-1)^{p-1}\Phi_p(1) = (-1)^{p-1}p = p.$$

Consequently, $\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}\left(\Phi_p'(\zeta)\right) = p^{p-2}$, hence

$$\Delta(\zeta) = (-1)^{\frac{p-1}{2}} p^{p-2}.$$

In particular, $\mathbb{Q}\left(\sqrt{\Delta(\zeta)}\right)$ is the unique quadratic extension of $\mathbb{Q}$ in $\mathbb{Q}(\zeta)$ by Proposition 59.14.

Note, since

$$\prod_{p-1 \geq i > j \geq 1} (\zeta^i - \zeta^j)^2 = \prod_{p-1 \geq i > j \geq 1} \left((\zeta^i - 1) - (\zeta^j - 1)\right)^2,$$

we also have $\Delta(\zeta) = \Delta(1 - \zeta)$.

The technique above can be generalized in order to compute $\Delta(\zeta)$ with ζ a primitive p^mth root of unity (and from it ζ a primitive nth root of unity). To show this, one uses:

Lemma 65.8. *Let ζ be a primitive nth root of unity in $\mathbb{C}$ and $K = \mathbb{Q}(\zeta)$. Then $\Delta(\zeta) \mid n^{\varphi(n)}$.*

PROOF. Let $\Phi_n = m_{K/F}(\zeta)$, then $t^n - 1 = \Phi_n g$ for some $g \in \mathbb{Z}[t]$. Taking derivatives, we have $nt^{n-1} = \Phi_n g' + \Phi_n g$. Evaluating at $t = \zeta$ and taking norms give

$$n^{\varphi(n)} = \mathrm{N}_{K/\mathbb{Q}}(n\zeta^{n-1}) = \mathrm{N}_{K/\mathbb{Q}}(\Phi_n') \, \mathrm{N}_{K/\mathbb{Q}}(g) = (-1)^{\frac{n(n-1)}{2}} \Delta(\zeta) \, \mathrm{N}_{K/\mathbb{Q}}(g)$$

in $\mathbb{Z}$. The result follows. □

Exercises 65.9.

1. Show that equation (65.2) is valid.
2. Prove Check 65.4.
3. Let $f = t^3 - a_1t^2 - a_2t - a_3 \in \mathbb{R}[t]$. Show
   (a) The discriminant $\Delta = -4a_1^3a_3 + a_1^2a_2^2 - 18a_1a_2a_3 + 4a_2^3 - 27a_3^2$.

(b) f has multiple roots if and only if $\Delta = 0$.
(c) f has three distinct real roots if and only if $\Delta > 0$.
(d) f has one real root and two non-real roots if and only if $\Delta < 0$.

4. Let F be a field and f be an irreducible separable cubic in $F[t]$. Show the Galois group of f is either A_3 or S_3 and if the characteristic of FE is not two, it is A_3 if and only if the discriminant of f is a square in F.
5. Let $t^3 + pt + q$ be irreducible over a finite field F of characteristic not 2 or 3. Show that $-4p^3 - 27q^2$ is a square in F.
6. Prove (1) and (2) of Proposition 65.6
7. Let ζ be a primitive p^mth root of unity in $\mathbb{C}$ and $K = \mathbb{Q}(\zeta)$. Show that $\Delta(\zeta) = \pm p^{m-1(mp-m-1)}$ with the plus sign occurring if and only if either $p \equiv 1 \mod 4$ or $a = 2^a$ with $a > 2$.

66. Purely Transcendental Extensions

Let K/F be a finitely generated extension of fields in which K is *purely transcendental* over F, i.e., $K \cong F(t_1, \dots, t_n)$. If $K/E/F$ is an intermediate field, the question arises whether E/F is a purely transcendental extension. In general, this has been shown to be false, but if $n = 1$ this is true and is the subject of this section.

Lemma 66.1. *Let F be a field and x an element transcendental over F. Suppose that u is an element in $F(x)$ not lying in F, say $u = f/g$ with f and g nonzero non-constant relatively prime polynomials in $F[t]$, and $r = \max\{\deg f, \deg g\}$. Then x is algebraic over $F(u)$ with $m_{F(u)}(X) = f - ug$ in $F(u)[t]$.*

Proof. Let $h = f - ug$ in $F[u][t] \subset F(u)[t]$ with f and g as in the lemma. Suppose that $f = \sum a_i t^i$ and $g = \sum b_i t^i$. If b_j is nonzero then so is $a_j - ub_j$, lest u lies in F. Thus $\deg h = \max\{\deg f, \deg g\}$ and x is algebraic over $F(u)$. It follows that u is transcendental over F as x is. To finish, we must show that h is irreducible in $F(u)[t]$. Suppose this is false, then h is also reducible in the UFD $F[u, t]$ by a consequence of Gauss' Lemma (Lemma 35.7). As h is primitive, we have $h = h_1 h_2$ in $F[u, t]$ with h_1 and h_2 lying in $F[u, t] \setminus F$. As u occurs linearly in h, we may assume that it occurs in h_1, say $h_1 = uh_3 + h_4$ with h_3, h_4 in $F[t]$ and h_2 lies in $F[t]$. So we have

$$g = f - ug = uh_2 h_3 - h_2 h_4.$$

As u is transcendental over $F(t)$, w must have $f = -h_2 h_4$ and $g = -h_2 h_3$. Therefore, $h_2 \mid f$ and $h_2 \mid g$ in $F[t]$, a contradiction. □

Corollary 66.2. *Let F be a field, x a transcendental element over F, and u an element of $F(x)$. Then $F(u) = F(x)$ if and only if there exist elements a, b, c, d in F with $ad - bc$ nonzero and $u = \dfrac{ax + b}{cx + d}$.*

[Note the condition on a, b, c, d insures that u does not lie in F.]

Proposition 66.3. *Let F be a field and x a transcendental element over F. Then*

$$\begin{aligned} G(F(x)/F) = \\ &\{f : F(x) \to F(x) \mid x \mapsto \frac{ax+b}{cx+d} \text{ an } F\text{-homomorphism with} \\ &\qquad\qquad a, b, c, d \text{ in } F \text{ satisfying } ad - bc \neq 0\} \\ &\cong \mathrm{GL}_2(F)/Z\big(\mathrm{GL}_2(F)\big). \end{aligned}$$

The group $\mathrm{GL}_2(F)/Z\big(\mathrm{GL}_2(F)\big)$ is called the (second) *projective linear group* and denoted by $\mathrm{PGL}_2(F)$. If $F = \mathbb{C}$, then $\mathbb{C} = \mathbb{C}^2$, so $\mathrm{PGL}_2(\mathbb{C}) = \mathrm{PSL}_2(\mathbb{C}) := \mathrm{SL}_2(F)/Z\big(\mathrm{SL}_2(F)\big)$. (If $A \in \mathrm{GL}_2(\mathbb{C})$ multiply it by $\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}$ where $d = (\sqrt{\det A})^{-1}$.)

Theorem 66.4. (Lüroth's Theorem) *Let F be a field and x a transcendental element over F. If $F(x)/E/F$ is an intermediate field, then there exists an element y in $F(x)$ such that $E = F(y)$.*

PROOF. We may assume that $F < E \subset F(x)$. Let z be an element in E not in F. Then $F(x)/F(z)$ is algebraic by the lemma, hence $F(x)/E$ is algebraic. Let $f = am_E(x)$ with $a \in F[x]$ and $f \in F[x][t]$ primitive. Say

$$f = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i \in F[x] \text{ for all } i$$

where n is the degree $\deg_t$ of f in t, so $a_n(x)$ is nonzero and $n = \deg m_E(x) = [F(x) : E]$. We also have all $a_i(x)/a_n(x)$ lie in E but as x is transcendental over F, there exists an i such that $u = a_i(x)/a_n(x)$ does not lie in F. We can also write $u = g(x)/h(x)$ with g and h nonzero relatively prime polynomials in $F[t]$. If $r = \max\{\deg g, \deg h\}$ then $r = [F(x) : F(u)]$ by the lemma, so

$$n = [F(x) : E] \leq [F(x) : F(u)] = r.$$

Let $d = g(x)h(t) - h(x)g(t)$ in $F[x, t]$. As g and h are relatively prime in the UFD $F[x, t]$, the polynomial d is nonzero. In addition, the polynomial

$$uh(t) - g(t) = h(x)^{-1}d \text{ in } E[t] \text{ has } x \text{ as a root,}$$

so $m_E(x) \mid h(x)^{-1}d$ in $E[t]$, hence $f = a(x)m_E(x) \mid h(x)^{-1}d$ in $F(x)[t]$. Since f is primitive in $F[x, t]$ and d lies in $F[x, t]$, we must have $f \mid d$ in $F[x, t]$, as a consequence of Gauss' Lemma (e.g., by Exercise 3ii(35.12)). Write $d = \alpha f$ with α in $F[x, t]$ and let $\deg_x$ denote the degree in x. We have $\deg_x d \leq r$ and $\deg_x f \geq \max\{\deg a_j(x) \mid a_j(x) \neq 0\}$. By choice, $u = g(x)/h(x) = a_i(x)/a_n(x)$, so $\deg_x f \geq \max\{\deg g, \deg h\} = r$. Consequently, $d = \alpha f$ in $F[x, t]$ and $f \mid d$ in $F[x, t]$ implies that $\deg_x d = r = \deg_x f$ and $\deg_x \alpha = 0$. In particular, α lies in $F[t]$, so is primitive when viewed as a polynomial over $F[x]$, i.e., an element in $F[x][t]$. By Gauss' Lemma, $d = f\alpha$ is also primitive in $F[x][t]$. Since d is skew symmetric in x and t, it must also be primitive in $F[t][x]$. As α lies in $F[t]$ and $\alpha \mid d$ in $F[x, t]$, we must have α lies in $F[t]^\times = F^\times$, hence

$$n = \deg_t f = \deg_t d = \deg_x d = \deg_x f = r,$$

so $E = F(u)$. $\square$

It is known if F is an algebraically closed field of characteristic zero, e.g., $\mathbb{C}$, that any subfield of $F[t_1, t_2]$ is purely transcendental. [A similar result holds if the characteristic is positive with a mild additional assumption.] However, the result is false for $F[t_1, t_2, t_3]$.

Suppose that F is a field and x is transcendental over F. Let $H \subset G(F(x)/F) \cong \mathrm{PGL}_2(F)$ be a finite subgroup. Then $F(x)/F(x)^H$ is a finite Galois extension. By Lüroth's Theorem, there exists an element u in $F(x)$ such that $F(u) = F(x)^H$. Suppose that $F = \mathbb{C}$, Then using a bit of complex analysis (stereographic projections, etc.), one can show that if $H \subset \mathrm{PGL}_2(\mathbb{C})$ is a finite subgroup, then H is isomorphic to a finite rotation group in $\mathbb{R}^3$. In Section 20 we classified such rotation groups in $\mathbb{R}^3$. Recall that they are the following:

(1) cyclic: $\cong \mu_n$, all n (planar).
(2) dihedral: $\cong D_n$, all n (planar).
(3) tetrahedral: $\cong A_4$ (the rotations of a regular tetrahedron).
(4) octahedral: $\cong S_4$ (the rotations of a cube or a regular octahedron).
(5) icosohedral: $\cong A_5$ (the rotations of a regular dodecahedron or a regular icosahedron).

Exercise 66.5. Prove that the rotations of a regular tetrahedron is isomorphic to A_4, the rotations of a cube or octahedron is isomorphic to S_4, and the rotations of a dodecahedron or icosahedron is isomorphic to A_5 using Exercise 24.24(9).

67. Finite Fields

A major property of a finite field F is that the multiplicative group of F is cyclic. If you have done the appropriate exercises, you see that we have all of the following:

Theorem 67.1. *Let F be a finite field with $|F| = q$. Then all the following are true:*

(1) $F^\times$ is a cyclic group.
(2) The characteristic of F is p for some prime p and $q = p^n$.
(3) If K/F is a finite extension of fields of degree m, then $|K| = q^m$.
(4) If p is a prime and m a positive integer, then there exists a field with p^m elements.
(5) If $|F| = p^n$, with p a prime, then F is a splitting field of $t^{p^n} - t$ over $\mathbb{Z}/p\mathbb{Z}$.
(6) Any two finite fields with the same number of elements are isomorphic. In particular, in a fixed algebraic closure of $\mathbb{Z}/p\mathbb{Z}$, with p a prime, there exists a unique field of order p^m for each positive integer m.
(7) If $|F| = q = p^n$, then $\varphi_F : F \to F$ by $x \mapsto x^q$ is an F-automorphism, the Frobenius automorphism, and $G(F/\mathbb{Z}/p\mathbb{Z}) = \langle \varphi_F \rangle$.
(8) If $|F| = p^n$ and K a field with $|K| = p^m$ in an algebraic closure of F, then K/F if and only if $n \mid m$.
(9) If K/F is a finite extension of degree m and $\varphi_K : K \to K$ the Frobenius automorphism $x \mapsto x^{q^m}$, then K/F is a cyclic extension with $G(K/F) = \langle \varphi_K^q \rangle$.
(10) Any finite field is perfect.
(11) There exists an x in F such that $F = (\mathbb{Z}/p\mathbb{Z})(x)$.

Given a finite field F with $q = p^n$ elements, one would like to find irreducible polynomials in $F[t]$. If f is an irreducible polynomial of degree d in $F[t]$, then by the theorem,

$f \mid t^{p^n} - t$ if and only if $d \mid n$. In particular, we must have

$$t^{p^n} - t = \prod_{d|n} f_{i,d}$$

where the $f_{i,d}$ in $F[t]$ run over all the monic irreducible polynomials of degree d. So we have $p^n = \sum_{d|n} dN_{p,d}$, where $N_{p,d}$ is the number of monic irreducible polynomials of degree d in $F[t]$, by comparing degrees. By the general form of Möbius Inversion (Exercise 59.22(6)), we conclude that

Proposition 67.2. *The number $N_{p,n}$ of monic irreducible polynomials of degree n in $(\mathbb{Z}/p\mathbb{Z})[t]$, with p a prime, satisfies*

$$N_{p,n} = \frac{1}{n}\sum_{d|n} \mu(\frac{n}{d})p^d.$$

Using the general form of the Möbius Inversion given by Exercise 59.22(6), we can determine irreducible polynomials, just as in the case of cyclotomic extensions of $\mathbb{Q}$. Let ζ_n be a primitive root of unity over $\mathbb{Z}/p\mathbb{Z}$, i.e., a generator for $F^\times$ if F is a field with p^n elements satisfying $p \nmid n$. Let Φ_n denote the product of $t - \zeta$ where ζ is a primitive nth root of unity in an algebraic closure of $\mathbb{Z}/p\mathbb{Z}$, so $\deg \Phi_n = \varphi(n)$. Unlike the case of $\mathbb{Q}$, the polynomial Φ_n is not irreducible. We have $t^n - 1 = \prod_{d|n} \Phi_d$. Arguing as the case over $\mathbb{Q}$, we conclude that

$$\Phi_n = \prod_{d|n}(t^d - 1)^{\mu(\frac{n}{d})} = \prod_{d|n}(t^{\frac{n}{d}} - 1)^{\mu(d)}.$$

in $(\mathbb{Z}/p\mathbb{Z})[t]$ by Möbius Inversion.

Example 67.3. Let p be a prime not dividing 12. Then

$$\begin{aligned}\Phi_{12} &= \prod_{d|12}(t^{\frac{12}{d}} - 1)^{\mu(d)}\\ &= (t^{12} - 1)^{\mu(1)}(t^6 - 1)^{\mu(2)}(t^4 - 1)^{\mu(3)}(t^3 - 1)^{\mu(4)}\\ &\qquad (t^2 - 1)^{\mu(6)}(t - 1)^{\mu(12)}\\ &= \frac{(t^{12} - 1)(t^2 - 1)}{(t^6 - 1)(t^4 - 1)} = t^4 - t^2 + 1.\end{aligned}$$

Lemma 67.4. *Let F be a finite field with q elements and i a positive integer. Set*

$$S(\widehat{t^i}) := \sum_{x \in F} x^i \ \text{ in } F,$$

then

$$S(\widehat{t^i}) = \begin{cases} -1 & \text{if } q - 1 \mid i \\ 0 & \text{otherwise.}\end{cases}$$

For convenience, as $q = 0$ in F, we shall set $S(\widehat{t^0}) = 0$ also, i.e., let $0^0 = 1$.

PROOF. Suppose that $q-1 \mid i$. Then $x^i = 1$ for all x in $F^\times$ and $S(\widehat{t^i}) = \sum_{F^\times} x^i = |F^{q-1}|1_F = -1$ in F. So we may assume that $q-1 \not| \, i$. Then there exists an element y in $F^\times$ such that $y^i \neq 1$ as $F^\times$ is cyclic of order $q-1$. Since $yF^\times = F^\times$, we have

$$S(\widehat{t^i}) = \sum_{F^\times} x^i = \sum_{F^\times} (yx)^i = y^i \sum_{F^\times} x^i,$$

so $(1-y^i)\sum_{F^\times} x^i = 0$. It follows that $S(\widehat{t^i}) = 0$ □

If $f = \sum a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n}$ is a polynomial in $R[t_1,\ldots,t_n]$ (R a commutative ring), then define the *total degree* deg of f is defined to be $\max\{i_1 + \cdots + i_n \mid a_{i_1,\ldots,i_n} \text{ nonzero}\}$. We are interested in solutions of $f = 0$ in R^n for such an f when R is a finite field.

Theorem 67.5. (Chevalley-Warning Theorem) *Let F be a finite field of characteristic p and f a non-constant polynomial in $F[t_1,\ldots,t_n]$ having total degree $\deg f < n$. Set*

$$V = \{\underline{x} = (x_1,\ldots,x_n) \in F^n \mid f(\underline{x}) = 0 \text{ in } F\}.$$

Then $|V| \equiv 0 \mod p$.

PROOF. Suppose that F has q elements. Set $P = 1 - f^{q-1}$ in $F[t_1,\ldots,t_n]$. As $|F^\times| = q-1$, if $\underline{x}$ lies in V, then $f(\underline{x}) = 0$ in F, so $P(\underline{x}) = 1$ in F; and if $\underline{x}$ does not lie in V, then $P(\underline{x}) = 0$ in F, i.e.,

$$P(\underline{x}) = \begin{cases} 0 & \text{if } \underline{x} \notin V \\ 1 & \text{if } \underline{x} \in V. \end{cases}$$

(So P is the characteristic function for V.) For any g in $F[t_1,\ldots,t_n]$, let

$$S(\widehat{g}) := \sum_{F^n} g(\underline{x}) \text{ in } F,$$

so

$$S(\widehat{P}) = \sum_{F^n} P(\underline{x}) = \sum_{V} 1 = |V|1_F \text{ in } F.$$

In particular, if we show $S(\widehat{P}) = 0$ in F, then $|V| \equiv 0 \mod p$ as needed.

We know that for all $(i_1,\ldots,i_n)$,

$$\sum_{F^n} x_1^{i_1} \cdots x_n^{i_n} = \big(\sum_F x_1^{i_1}\big) \cdots \big(\sum_F x_n^{i_n}\big)$$

and $\deg P < n(q-1)$. If $P = \sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n}$, we have $\sum_{j=1}^n i_j < n(q-1)$ if $a_{i_1,\ldots,i_n} \neq 0$. In particular, if $a_{i_1,\ldots,i_n} \neq 0$ is a coefficient of P, then, there exists a j, $1 \leq j \leq n$, with $i_j < q-1$. Therefore, we have

$$\begin{aligned} S(\widehat{P}) = \sum_{F^n} P(\underline{x}) &= \sum_{F^n} \sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n} \\ &= \sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} \sum_{F^n} x_1^{i_1} \cdots x_n^{i_n} \\ &= \sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} \big(\sum_F x_1^{i_1}\big) \cdots \big(\sum_F x_n^{i_n}\big) = 0, \end{aligned}$$

as $\sum_F y^j = S(\widehat{t^j}) = 0$ for $j < q - 1$ by the lemma. $\square$

Of course, one is interested in solutions of such a non-constant polynomial f in $F[t_1, \ldots, t_n]$ if every variable t_i occurs in f nontrivially, since the result is always true otherwise. A polynomial f in $F[t_1, \ldots, t_n]$ is called *homogeneous of degree* d if every nonzero monomial in f has the same total degree d.

Corollary 67.6. *Let F be a finite field and f a non-constant polynomial in $F[t_1, \ldots, t_n]$ having total degree $\deg f < n$. Suppose that $f(\underline{0}) = 0$. Then there exists a point $\underline{x}$ in F^n besides the origin such that $f(\underline{x}) = 0$. In particular, if f is homogeneous of degree d satisfying $n > d$, then f has a nontrivial solution in F^n.*

A homogeneous polynomial of degree two is called a *quadratic form*. The corollary says that any quadratic form in at least three variables over a finite field has a nontrivial solution. A much deeper theorem (due to Weil) says that if f is any homogeneous polynomial in $\mathbb{Z}[t_1, \ldots, t_n]$ that remains irreducible in $\mathbb{C}[t_1, \ldots, t_n]$, then $f \equiv 0 \mod p$ has a nontrivial zero for all primes $p >> 0$.

A theorem of Jacobson says that if R is a ring (even a rng) in which for each element x there exists an integer $n = n(x) > 1$ satisfying $x^n = x$, then R is commutative. For a proof of this more general result, see Section 68. The following result is the starting point.

Theorem 67.7. (Wedderburn) *Every finite division ring is a field.*

Proof. (Witt) Let D be a division ring. Define the *center* $Z(D)$ of D to be $Z(D) := \{x \in D \mid xy = yx \text{ for all } y \text{ in } D\}$. So $Z(D^\times)$ is the center of the group $D^\times$. Clearly, $Z(D)$ is a field. Now suppose that D is finite. Let $F = Z(D)$. Then F is also finite, so $\operatorname{char} F = p$ for some prime p and $|F| = q = p^n$, for some n. We know that D is an F-vector space, so $|D| = q^m$ where $m = \dim_F D$. Assume that D is not commutative, i.e., $m > 1$. If $x \in D$ does not lie in F, then the conjugacy class of x in $D^\times$ is $\{y \in D^\times \mid y = axa^{-1} \text{ for some } a \text{ in } D^\times\}$ and the centralizer of x in $D^\times$ is $Z_{D^\times}(x) = \{y \in D^\times \mid xy = yx\}$. Set $Z_D(x) = Z_{D^\times}(x) \cup \{0\}$. Then $F \subset \mathbb{Z}_D(x) \subset D$. Clearly, $Z_D(x)$ is also a division ring and a finite dimensional vector space over F. Set $\delta(x) := \dim_F Z_D(x)$. We have $|Z_D(x)| = q^{\delta(x)}$. The notation and basic theorems about vector spaces over fields hold over division rings (i.e., as (left) modules over division rings) with the same proofs. [The one exception is the Representation Theorem of linear operators where matrix multiplication of matrices A and B must be modified so that the $ij$the entry of AB is $\sum B_{kj}A_{ik}$ – unless linear operators and scalars are written on opposite sides.] In particular, D is a (left) $Z_D(x)$-vector space. It is easy to see that $q^m = |D| = |Z_D(x)|^r = q^{\delta(x)r}$, some r. In particular, $m = \delta(x)r$, so $\delta(x) \mid m$ (cf. with the analogue of a tower of fields). By assumption $\delta(x) < m$ if x does not lie in F. Let $\mathcal{C}^*$ be a system of representatives of the conjugacy classes of $D^\times$ not in the center of $D^\times = F^\times$. By the Class Equation (21.3), we have

$$q^m - 1 = |F^\times| + \sum_{\mathbb{C}^*}[D^\times : Z_{D^\times}(x)] = (q - 1) + \sum_{C(x) \in \mathcal{C}^*} \frac{q^m - 1}{q^{\delta(x)} - 1}.$$

Let Φ_m be the mth cyclotomic polynomial $\prod_{\zeta\in\mu_m \text{ primitive}} \zeta$ in $\mathbb{Z}[t]$. Then we know that $\Phi_m \mid \dfrac{t^m-1}{t^\delta - l}$ in $\mathbb{Z}[t]$ if $\delta \mid m$ and $\delta \neq m$, as $t^m - 1 = \prod_{d|m}\Phi_d$. Therefore,

$$\Phi_m(q) \mid q^m - 1 \text{ and } \Phi_m(q) \mid \sum_{\mathcal{C}^*} \frac{q^m-1}{q^{\delta(x)}-1}$$

in $\mathbb{Z}$. Consequently, $\Phi_m(q) \mid q-1$ in $\mathbb{Z}$.

Claim. If ζ is a primitive mth roof of unity, then $|q-\zeta| > |q-1|$:

This is clear, if you draw a picture or if $\zeta = e^{2\pi\sqrt{-1}r/m}$ with r and m relatively prime, then

$$\begin{aligned}|q-\zeta|^2 &= (q - e^{2\pi\sqrt{-1}r/m})(q - e^{-2\pi\sqrt{-1}r/m})\\ &= q^2 + 1 - 2q\cos\frac{2\pi r}{m} > q^2 + 1 - 2q = (q-1)^2.\end{aligned}$$

As $-1 < \cos\frac{2\pi r}{m} < 1$, the claim is established. The claim implies that

$$|\Phi_m(q)| = \prod_{\zeta\in\mu_n \text{ primitive}} |q-\zeta| > |q-1|.$$

Hence $\Phi_m(q) \mid q-1$ in $\mathbb{Z}$ is impossible. This contradiction shows that $D = F$. $\square$

Exercises 67.8.

1. Prove Theorem 67.1.
2. Show that the nth cyclotomic polynomial $\Phi_n \in \mathbb{Z}/p\mathbb{Z}[t]$ with p a prime not dividing n is a product of $\varphi(n)/d$ of irreducible polynomials of the same degree d.
3. Show that over over finite field F, every element of F is a sum of two squares.
4. Let K/F be a finite extension of finite fields. Prove that the norm map $N_{K/F} : K \to F$ is surjective.
5. Let F be a finite field and f a homogeneous form of degree 2 (i.e., a quadratic form) in two variables with both variables occurring nontrivially. Then every nonzero element x in F is a value of f, i.e., there exists a y in F satisfying $f(y) = x$.

68. Addendum: Jacobson's Theorem

In this section, we generalize Wedderburn's Theorem that finite division rings are fields.

Let R be any ring. Then its *center*, $Z(R) := \{x \in R \mid xy = yx \text{ for all } y \text{ in } R\}$, is a commutative subring. If D is a division ring, then $Z(D)$ is a field. In addition, if $x \in D$ and F any subfield of $Z(D)$, then $F[x]$ is a commutative subring of D and its quotient field $F(x)$ is a field and also lies in D. Moreover, if x is algebraic over F, i.e., $F[x]$ is a finite F-vector space, then $F(x) = F[x]$ and $F(x)/F$ is finite. We have looked at two cases for D, the case that D is finite and the case that D is the Hamiltonian quaternions (where it is a finite dimensional $\mathbb{R}$-vector space). In the first case, D was commutative, but in the second case this might not be true. When D is a finite division ring, we have $x^{|D|} = 1$ for all x in D. Jacobson's Theorem says that a ring R that satisfies $x^n = x$ for each x in R for some integer $n = n(x) > 1$ is commutative. We prove this in this section. We also show that any such ring is a subring of a product of fields, each algebraic over a finite field.

If D is a division ring, then two nonzero elements x and y in D commute if and only if $yxy^{-1} = x$. We want to exploit this in the special case of Jacobson's Theorem when R is a division ring (and reduce to the finite division ring case that we have answered).

Proposition 68.1. *Let D be a division ring that satisfies $x^n = x$ for each x in D and some integer $n > 1$ depending on x. Then D is a field.*

PROOF. We know that the center of D contains a prime field, so $n1_D$ lies in the center for all integers n. Let x be a nonzero element of D. Then there exist positive integers n and n' such that $x^n = x$ and $(2x)^{n'} = 2x$. It follows with $N = (n-1)(n'-1)+1$ that $(2^N - 2)x = 0$ in D, so $2^N - 2 = 0$. We conclude that the center of D has characteristic p for some prime p. Let F be the prime field in the center of D, so $F \cong \mathbb{Z}/p\mathbb{Z}$. Suppose that D is not commutative. Let x in D not lie in the center. As x is algebraic over F, since it is a root of $t^n - t$ in $F[t]$ for some positive integer n, the field extension $F(x)/F$ is finite by the discussion prior to the proposition. Therefore, $F(x)$ is a finite field. Indeed the same argument shows that $F(y)$ is a finite field for all y in D. Let $|F(x)| = p^m$. We also know that any element z in $F(x)$ satisfies $z^{p^m} = z$.

Claim: There exists an element y in D and an integer k such that $x^k \neq x$ and $yxy^{-1} = x^k$: Let $T_x : D \to D$ be the F-linear transformation defined by $z \mapsto zx - xz$. As x is not in the center of D, this map is not the trivial map, i.e., $T_x(z) \neq 0$ for some z. Since

$$T_x^2(z) = T_x(zx - xz) = zx^2 - 2xzx + x^2z,$$

inductively, we see that

$$T_x^l(z) = \sum_{i=0}^{l} (-1)^i \binom{l}{i} x^i z x^{l-i} \text{ for all positive integers } l \text{ and all } z \text{ in } D.$$

Since $\operatorname{char} F = p$, the Children's Binomial Theorem implies that $T_x^p(z) = zx^p - x^pz$, hence

$$T_x^{p^m}(z) = zx^{p^m} - x^{p^m}x = zx - xz = T_x(z) \text{ for all } z \text{ in } D,$$

i.e., $T_x^{p^m} = T_x$. For each w in D, left multiplication by w, $\lambda_w : D \to D$, i.e., $z \mapsto wz$ is also an F-linear transformation. We have for each w in $F(x)$ (so $wx = xw$)

$$T_x\lambda_w(z) - T_x(wz) = wzx - x(wz) = w(zx - xz) = \lambda_w T_x(z),$$

i.e., $T_x\lambda_w = \lambda_w T_x$ for all w in $F(x)$ and z in D. As $\lambda_w\lambda_{w'} = \lambda_{w'}\lambda_w$ if w, w' lie in $F(x)$, we have

$$(*) \qquad 0 = T_x^{p^m} - T_x = \prod_{F(x)} (T_x - \lambda_w).$$

By assumption, there exists an element y in D satisfying $xy - yx \neq 0$. As D is a division ring, $D^\times$ satisfies the cancellation law and y is invertible in D. It follows that there exists a nonzero element w in $F(x)$ such that $0 = (T_x - \lambda_w)(y) = yx - xy - wy$, i.e., $yxy^{-1} = x + w$ in $F(x)$ by (*). Let $s = p^m - 1$, the order of $F(x)^\times = \langle x \rangle$. As $1, x, \dots, x^{s-1}$ are all the roots of $t^s - 1$ in $F(x)[t]$, we must have $yxy^{-1} = x^k$ for some $k > 1$. This proves the claim.

Let y in D be as in the claim. We know by the argument above that $F(y)$ is also a finite field with say p^n elements. So we have $y^{p^n} = y$ and $xy \neq yx$. Set

$$W := \{z \in D \mid z = \sum_{i=0}^{p^m} \sum_{j=0}^{p^n} a_{ij} x^i y^j \text{ with } a_{ij} \text{ in } F\}.$$

We have W a finite set closed under addition. By the claim, W is closed under multiplication, so W is a finite subring of D. If z lies in W, $F(z) = F[z] \subset W$, so W is a finite division ring. Consequently, W is a field by Wedderburn's Theorem, contradicting x and y do not commute. □

Let F be a field and D a division ring containing F. If $x \in D$, then $F(x)$ is a field in D. We say that D is *algebraic* over F if every element in D is algebraic over F.

Corollary 68.2. *Let D be a division algebra that is algebraic over a finite field. Then D is commutative. In particular, D/F is an algebraic extension of fields.*

PROOF. If F is the finite field in D and of characteristic p, then $F(x)$ is a finite field for each $x \in D$, hence $x^{p^n} = x$ for some n. It follows that D is commutative by the proposition. □

We next generalize the proposition.

Theorem 68.3. (Jacobson) *Let R be a rng (i.e., a ring possibly without a one). Suppose for each element x in R, there exists an integer $n > 1$ depending on x such that $x^n = x$. Then R is commutative.*

PROOF. We may assume that R is not the trivial rng. We want to reduce to the proposition. Let x be an element in R and $n > 1$ an integer such that $x^n = x$. Since for every element z in R, $z^m = z$ for some integer $m > 1$, R cannot have any nonzero nilpotent elements, i.e., we cannot have $z^m = 0$ unless $z = 0$ – why?

Step 1. The element x^{n-1} is a *central idempotent* in R, i.e., x^{n-1} lies in the center of R and $(x^{n-1})^2 = x^{n-1}$:

Set $e = x^{n-1}$. We have $e^2 = x^{2n-2} = x^n x^{n-2} = xx^{n-2} = x^{n-1} = e$, so e is an idempotent in R. If y lies in R, then $(ye - eye)^2 = 0 = (ey - eye)^2$. As R has no nonzero nilpotent elements, we have $ye = eye = ye$, so e is central.

Step 2. Every right ideal $\mathfrak{A}$ in R is a (two-sided) ideal:

Let r be an element in R and x an element in $\mathfrak{A}$. Then, as $\mathfrak{A}$ is a right ideal, with x^{n-1} central and lying in $\mathfrak{A}$, we have $rx = rx^{n-1}x = \big((x^{n-1})r\big)x$ lies in $\mathfrak{A}$.

Step 3. Suppose that R is a ring (i.e., has a one). Define the *right radical* $\mathfrak{R}$ of R to be the intersection of all maximal right ideals in R. [It can be shown that the *left radical* of R and the right radical of R are the same, but we do not need this.] Then for all y in R, the element $xy - yx$ lies in $\mathfrak{R}$:

Since 1 lies in R, the Zorn's Lemma argument showing that nonzero commutative rings have maximal ideals, shows that maximal right ideals exist in R. Let $\mathfrak{m}$ be any such maximal right ideal. By Step 2, $\mathfrak{m}$ is an ideal in R. By the Correspondence Principle, $R/\mathfrak{m}$ is a ring with no nontrivial right ideals. It follows that every nonzero element in

$R/\mathfrak{m}$ has a right inverse and therefore an inverse, i.e., $R/\mathfrak{m}$ is a division ring – why? If z is an element in R with $z^m = z$ and $m > 1$, we have $(z+\mathfrak{m})^m = z^m + \mathfrak{m} = z + \mathfrak{m}$, so $R/\mathfrak{m}$ is a division ring satisfying the hypothesis of the proposition, hence commutative. Therefore, $(z+\mathfrak{m})(y+\mathfrak{m}) = (y+\mathfrak{m})(z+\mathfrak{m})$ for all y, z in R, i.e., $zy - yz$ lies in $\mathfrak{m}$ for any right maximal ideal in R as needed.

Step 4. If R is a ring, then R is commutative:

It suffices to show the (right) radical $\mathfrak{R}$ in R is zero. Suppose not and x is a nonzero element in $\mathfrak{R}$. Let $x^n = x$ with $n > 1$ and $e = x^{n-1}$. By Step 1, e is a central idempotent and lies in $\mathfrak{R}$ as x does. Suppose that $1 - e$ is not a unit. Then $1 - e$ lies in a maximal right ideal $\mathfrak{m}$ in R. This implies that 1 lies in $\mathfrak{m}$ as $e \in \mathfrak{R} \subset \mathfrak{m}$, which is impossible. Therefore, the right ideal $(1-e)R = R$, so $(1-e)r = e$ for some r in R. We conclude that $0 = e(1-e)r = e^2 = e = x^{n-1}$, contradicting x cannot be nilpotent.

Step 5. Finish:

By Step 4, we may assume that R does not have a one. Let e be a central idempotent in R as in Step 1 and set $T = eR = Re$. Check that T is a ring with $e = 1$ and for each y in T, there exists an integer $n > 1$ such that $y^n = y$. By Step 4, we know that T is a commutative ring. Let x and y be elements of R. Then using the commutativity of T, we have $xye = (xe)(ye) = (ye)(xe) = yxe$. It follows that $(xy - yx)e = 0$. Since e was a central idempotent in R $(xy - yx)e = 0$ for every central idempotent e in R. In particular, if $m > 1$ an integer such that $(xy - yx)^m = xy - yx$ in R, then $(xy - yx)^{m-1}$ is a central idempotent by Step 1, so $xy - yx = (xy - yx)(xy - yx)^{m-1} = 0$. This proves that R is commutative. □

Corollary 68.4. *Let R be a ring satisfying for each element x in R, there exists an integer $n > 1$ depending on x such that $x^n = x$. Then R is isomorphic to a subring that is a product of fields each algebraic over a finite field of positive characteristic.*

Proof. We know that R is commutative. Let $R \to \bigtimes_{\mathrm{Min}(R)} R/\mathfrak{p}$ be the natural map where $\mathrm{Min}(R)$ is the set of minimal prime ideals in R. Then each $R/\mathfrak{p}$ is a commutative domain. Since $x^n = x$ for some x, either $x = 0$ or $x^{n-1} = 1$. It follows that $R/\mathfrak{p}$ is a field and R has no nonzero nilpotent elements. Therefore, $\mathrm{nil}(R) = 0$ and the map is a ring monomorphism. Finally since each element in $R/\mathfrak{p}$, $\mathfrak{p} \in \mathrm{Min}(R)$, is a root of a polynomial of the form $t^m - 1$ over its prime field, we must have $R/\mathfrak{p}$ must be of characteristic p and be an algebraic extension over its prime field. The result follows □

Exercises 68.5.

1. Show that if R is a rng in which every element x satisfies $x^n = x$ for some integer n (depending on x), that R has no nontrivial nilpotent elements.
2. If R is a nontrivial ring in which every nonzero element has an inverse, show that R is a division ring.

69. Addendum: Hilbert Irreducibility Theorem

Throughout this section, $x, y, t, t_1, \dots, t_n$ will be independent variables. Hilbert's Irreducibility Theorem says if $f(t_1, \dots, t_n, t)$ is an irreducible polynomial in $\mathbb{Q}[t_1, \dots, t_n, t]$, then $f(\alpha_1, \dots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$ for infinitely many rational numbers $\alpha_1, \dots, \alpha_n$.

Hilbert used this theorem to prove that there exists a finite Galois extension of $\mathbb{Q}$ with Galois group S_n and A_n for every n. We shall apply the theorem to show S_n occurs as a Galois group over $\mathbb{Q}$ for all n. This will show that there exist irreducible polynomials of any degree over $\mathbb{Q}$ having no root solvable by radicals.

We begin by looking at a polynomial $f(x,t)$ in $\mathbb{Q}[x,t]$. Write it as

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i(x) \in \mathbb{Q}[x] \text{ and } a_n(x) \text{ nonzero,}$$

so the degree of $f(x,t)$ in t, $\deg_t f$, is n. We can view $f(x,t)$ in $\mathbb{C}[x,t]$. Since $\mathbb{Q}$ is infinite, we can and do identify polynomial functions and polynomials over $\mathbb{C}$. We know by Proposition 42.8 if, in addition, that f is irreducible in $\mathbb{C}[x,y]$ that all but finitely many complex numbers x_0 are regular values, i.e., $a_n(x_0)$ is not zero and $f(x_0,t)$ has n distinct roots in $\mathbb{C}$. We say that a real or complex valued function $\rho(x)$ is *analytic* at x_0 if ρ converges to its Taylor series in a neighborhood of x_0, i.e., has a positive radius of convergence R centered at x_0. If x_0 is a regular value of $f(x,t)$, we shall call an analytic function $\rho = \rho(x)$ a *root function* of $f(x,t)$ in a neighborhood of x_0 if $f\big(x,\rho(x)\big) = 0$ in a neighborhood of x_0, i.e., there exists a real number $R > 0$ such that $\rho(x)$ are power series in $x - x_0$ for all x satisfying $|x_{-}x_O| < R$. In particular, $\rho(x_0)$ is a root of $f(x_0,t)$. We begin by constructing root functions in a neighborhood of regular value x_0 of any given polynomial in $\mathbb{C}[x,t]$.

Proposition 69.1. *Let*

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i(x) \in \mathbb{Q}[x]$$

be a polynomial in $\mathbb{Q}[x,t]$ of degree n in t (so ($a_n(x)$ is nonzero) and x_0 a regular value of $f(x,t)$. Then there exist precisely n distinct root functions $\rho_1,\dots,\rho_n$ of $f(x,t)$ at x_0 and they satisfy

$$(*) \qquad f(x,t) = a_n(x)\prod_{i=1}^{n}\big(t-\rho_i(x)\big) \text{ in a neighborhood of } x_0$$

with the $\rho_i(x)$ power series expansions in $x-x_0$ with coefficients in $\widetilde{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, in this neighborhood of x_0.

Proof. Let $\alpha_1,\dots,\alpha_n$ be the (distinct) roots of $f(x_0,t)$ in $\mathbb{C}$, so lie in a finite extension of $\mathbb{Q}$, e.g., any field containing $K = \mathbb{Q}(\alpha_1,\dots,\alpha_n)$ in $\widetilde{\mathbb{Q}}$. For each $i = 1,\dots,n$, we shall construct a unique root function ρ_i satisfying $\rho_i(x_0) = \alpha_i$ with coefficients K in its power series expansion about x_0. We make the following reductions:

(*i*) As $a_n(x_0) \neq 0$, we may assume that $a_n(x_0) = 1$.
(*ii*) We may assume that $x_0 = 0$ by replacing $f(x,t)$ by the polynomial $g(x,t) = f(x+x_0,t)$.
(*iii*) We may assume that $\alpha_i = 0$ by replacing $f(x,t)$ by the polynomial $g(x,t) = f(x,t-\alpha_i)$.

So we can assume that $x_0 = 0$ is a regular value of $f(x,t)$ having root $\alpha_i = 0$ and we must construct a unique root function in a neighborhood of $x = 0$ with the given conditions.

Therefore, we must find a unique power series

$$\rho(x) = \sum_{i=1}^{\infty} b_i x^i \text{ with } b_i \in K \text{ converging for all } |x| < R \tag{69.2}$$

for some positive real number R that satisfies $f\big(x, \rho(x)\big) = 0$ in a neighborhood of $x = 0$. As $f(0,0) = 0$, we can write the polynomial

$$f(x,t) = a_{10}x + a_{01}t + \sum_{i+j\geq 2} a_{ij}x^i t^j \text{ in } \mathbb{Q}[x,t]$$

with almost all a_{ij} zero and the $a_{ij} \in K$. Since $x_0 = 0$ is a regular value of $f(x,t)$, we also have $\dfrac{\partial f}{\partial t}(0,0) = a_{01}$ is nonzero, so we can further assume that $a_{01}(0,0) = -1$, i.e.,

$$f(x,t) = a_{10}x - t + \sum_{i+j\geq 2} a_{ij}x^i t^j \text{ in } \mathbb{Q}[x,t]. \tag{69.3}$$

Assume that $\rho(x)$ has a positive radius of convergence about 0. We show that the b_i in (69.2) are uniquely determined by the equation

$$f(x, \sum_{i=1}^{\infty} b_i x^i) = 0 \tag{69.4}$$

in a neighborhood of $x = 0$, i.e., $\rho(x)$ is the unique root function in a neighborhood of 0 and with the b_i lying in K.

The coefficient of x^i in the expansion of the left hand side of equation (69.4) must be zero for each i. In particular, evaluating the x term shows that $0 = a_{10} - b_1$, so we must have $b_1 = a_{10}$. Inductively, we may suppose that we have constructed unique $b_1, \ldots, b_{k-1}$ in K. It is convenient to view y as a new variable to replace t. Let

$$h(x,y) = a_{10}x - y + \sum_{i+j\geq 2} a_{ij}x^i y^j \text{ and } y_0 = \sum_{i=1}^{k-1} b_i x^i.$$

Then the (formal) Taylor expansion for $h(x,y)$ at y_0 exists by Exercise 69.19(2) and is

$$\begin{aligned} h(x,y) = f(x,y_0) &+ \frac{\partial f}{\partial y}(x,y_0)(y - y_0) \\ &+ \text{ terms in higher powers of } y - y_0. \end{aligned}$$

By (69.3), we must have

$$\frac{\partial f}{\partial y}(x,y) = -1 + g(x,y)$$

with every term in $g(x,y)$ having total degree at least one. Evaluating y at $\rho(x)$ yields $\rho(x) - y_0 = \sum_{i=k}^{\infty} b_i x^i$, so

$$\begin{aligned} 0 = f(x, \sum_{i=1}^{k-1} b_i x^i) &+ \big(-1 + g(x, \sum_{i=1}^{k-1} b_i x^i)\big)(\sum_{i=k}^{\infty} b_i x^i) \\ &+ \text{ terms with power of } x \text{ at least } 2k. \end{aligned}$$

As the coefficient of x^k in this equation is zero, the coefficient $-b_k$ of x^k of $f(x, y_0)$ in this equation is uniquely determined by f and $b_1, \ldots, b_{k-1}$ as desired.

We can say more. If $\rho(x)$ is an analytic function, i.e., has positive radius of convergence, then in some neighborhood of $x_0 = 0$, we would have $b_1 = a_{10}$ and $f(x, \sum_{i=1}^{\infty} b_i x^i) = 0$, so

$$\sum_{i=2}^{\infty} b_i x^i = \sum_{i+j\geq 2} a_{ij} x^i (\sum_{l=1}^{\infty} b_l x^l)^j.$$

Expanding this and looking at the coefficient of x^k for $k > 1$ leads to an equation $b_k = \sum_{i+lj=k} c_{ijl} a_{ij} b_l$ for some nonnegative integers c_{ijl}. In particular, $l < k$. For each of the finite nonzero a_{ij} occurring in $f(x, t)$, let t_{ij} be a new variable. By induction, it follows that there exist nonzero polynomials $p_k \in \mathbb{Z}[t_{ij}]_{i,j}$ having positive integer coefficients (in the nonzero terms) and satisfying $b_k = p_k(a_{ij})$.

To finish we must still show that $\rho(x) = \sum b_i x^i$ has a positive radius of convergence in a neighborhood of $x_0 = 0$. Choose a positive integer A satisfying $A > |a_{ij}|$ for all of the finitely many nonzero a_{ij} occurring in $f(x, t)$ and set

$$f_A(x, t) = Ax - t + A(\sum_{i+j\geq 2} x^i t^j).$$

By the argument above, we know if f_A has a positive radius of convergence, then there exist unique $\widetilde{b}_k \in \mathbb{C}$ and $\widetilde{p}_k \in \mathbb{Z}[t_{ij}]_{i,j}$ satisfying $f_A(x, \sum_{i=1}^{\infty} b_i x^i) = 0$ and $\widetilde{b}_k = \widetilde{p}_k(A, \ldots, A)$ are positive integers for all k. Since $\widetilde{b}_k \geq p_k(|a_{ij}|) \geq |b_k|$ for all k, it suffices to show that $y = \sum_{i=1}^{\infty} \widetilde{b}_i x^i$ converges for $|x| < R$ for some positive real number R. Summing the geometric series, shows

$$\begin{aligned}
0 &= Ax - y + A \sum_{i+j\geq 2} x^i y^j \\
&= Ax - y + Ax^0 \sum_{j=2}^{\infty} y^j + Ax^1 \sum_{j=1}^{\infty} y^j + A \sum_{i=2}^{\infty} x^i (\sum_{j=0}^{\infty} y^j) \\
&= Ax - y + A\frac{y^2}{1-y} + Ax\frac{y}{1-y} + A\frac{x^2}{1-x}\frac{1}{1-y}.
\end{aligned}$$

Multiplying by $1 - y$ yields

$$\begin{aligned}
0 &= Ax - Ayx + -y + y^2 + Ay^2 + Axy + A(\frac{x^2}{1-x}) \\
&= (A+1)y^2 - y + A\frac{x}{1-x}.
\end{aligned}$$

One of the two solutions for $y(0) = 0$, i.e., when $x = 0$ is

$$y = \frac{1 - \sqrt{1 - 4(A+1)Ax/(1-x)}}{2(A+1)}.$$

Since $\sqrt{1 - \big(bx/(1-x)\big)} = (\sqrt{1-cx})/(\sqrt{1-x})$ has a positive radius of convergence about $x = 0$ for $b = 4A(A+1)$ and $c = b+1$, it follows that f_A is analytic in some neighborhood of 0, hence $\rho(x)$ is a root function in a neighborhood of 0.

We therefore conclude that for each $i = 1, \ldots, n$, there exist a unique root function $\rho_i(x)$ in a neighborhood of x_0 for the root α_i of $f(x_0, y)$ with coefficients in its power series expansion in K. Since the the constant term of $\rho_i(x)$ is α_i for $i = 1, \ldots, n$, the ρ_i are all distinct in some common neighborhood of x_0 and satisfy $f(x,t) = \prod_{i=1}^n \big(t - \rho_i(x)\big)$ in this neighborhood. $\square$

We shall need the following well-known lemma whose proof we leave as an exercise.

Lemma 69.5. (Lagrange Interpolation) *Suppose that F is an infinite field, $a_0, \ldots, a_n$ distinct elements in F and $c_0, \ldots, c_n$ elements in F. Then there exists a unique polynomial f in $F[t]$ of degree at most n satisfying $f(a_i) = c_i$ for each $i = 0, \ldots, n$.*

Let $\alpha_0 < \cdots < \alpha_m$ be real numbers. Recall that the $(m+1)$st Vandermonde determinant is given by

$$\begin{aligned} V_m = V_m(\alpha_0, \alpha_1, \ldots, \alpha_m) &= \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^m \end{pmatrix} \\ &= \prod_{0 \le i < j \le m} (\alpha_j - \alpha_i) \neq 0. \end{aligned}$$

If z is a real-valued function of x, let

$$W_m = W_m(\alpha_0, \alpha_1, \ldots, \alpha_m, z) = \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{m-1} & z(\alpha_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^{m-1} & z(\alpha_m) \end{pmatrix}.$$

With this notation, we have:

Lemma 69.6. *Let $\alpha_0 < \cdots < \alpha_m$ be real numbers and $z(x)$ an m times differential real-valued function on the closed interval $[\alpha_0, \alpha_m]$. Then there exist a real number α satisfying*

$$\frac{z^{(m)}(\alpha)}{m!} = \frac{W_m}{V_m}$$

satisfying $\alpha_0 < \alpha < \alpha_m$ with W_m and V_m as above. In particular, if $z^{(m)}(x)$ is nonzero on $[\alpha_0, \alpha_m]$, then W_m is also nonzero.

Proof. Let $y(t) = a_0 + a_1 t + \cdots + a_m t^m$ in $\mathbb{R}[t]$ be the unique polynomial satisfying $y(\alpha_i) = z(\alpha_i)$ for $i = 0, \ldots, m$ given by Lagrange Interpolation. By Cramer's Rule, the system of equations

$$a_0 + a_1\alpha_i + \cdots + a_m\alpha_i^m = z(\alpha_i) \ \text{ for } \ i = 0, \ldots, m$$

satisfies $a_m = W_m/V_m$.

By Rolle's Theorem there exists m distinct real numbers $\overline{\alpha}_i$ with each $\overline{\alpha}_i$ lying in the open interval (α_i, α_{i+1}) and satisfying $y^{(1)}(\overline{\alpha}_i) = z^{(1)}(\overline{\alpha}_i)$ for $i = 0, \ldots, m-1$. Similarly, we see that there exist $m-1$ distinct points on which $y^{(2)}$ and $z^{(2)}$ agree. Inductively, we find that there exists a point α satisfying $\alpha_0 < \alpha < \alpha_m$ on which y and z agree. Since $y^{(m)}(x) = m!\, a_m$, we have $z^{(m)}(\alpha) = y^{(m)}(\alpha) = m!\, a_m$ as needed. $\square$

We now prove the Hilbert Irreducibility Theorem in the special case of two variables. We shall use the following observation.

Observation 69.7. As $\operatorname{char}(\mathbb{C}) = 0$, the domain of polynomials in $\mathbb{C}[t]$, can be identified with the domain of polynomial functions on an interval I (of nonzero length) of $\mathbb{C}$, hence as a subring of the domain R of convergent functions on I. It follows that we can identify the quotient field of $\mathbb{C}[t]$ with a subfield of the quotient field of R, the field of finite Laurent series on I. [Note we can also work with polynomials in $1/t$ with the appropriate modifications.] We shall utilize this as an identification below.

Proposition 69.8. *Let $f(x,t)$ be an irreducible polynomial in $\mathbb{Q}[x,t]$. Then there exist infinitely many rational numbers α such that $f(\alpha,t)$ is irreducible in $\mathbb{Q}[t]$.*

Proof. Let

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x)$$

with $a_i(x)$ in $\mathbb{Q}[x]$ for $i = 0,\ldots,n$ and $a_n \neq 0$. Let $d = \max\{\deg_x a_i(x) \mid 0 \leq i \leq n\}$. Almost all elements of $\mathbb{C}$ are regular values of $f(x,t)$. Let x_0 be one such. Replacing $f(x,t)$ by the irreducible polynomial $h(x,t) = f(x+x_0,t)$, we may assume $x_0 = 0$. By Proposition 69.1, we can write

$$(i) \qquad f(x,t) = a_n(x)\prod_{i=1}^{n}\bigl(t-\rho_i(x)\bigr)$$

for each x in some neighborhood of $x_0 = 0$ with ρ_i the n distinct root functions of f at $x = 0$ (so $\rho_i(x_0)$ are the roots of $f(x_0,t)$) and these root functions have coefficients lying in $\widetilde{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, in their power series expansion in this neighborhood. It is convenient to work at 'infinity' rather than at 0, where a real or complex valued function $h(x)$ is said to be *analytic at infinity* if it can be represented by a power series $\sum_{i=1}^{\infty} c_i\dfrac{1}{x^i}$ converging for all $|x| > R$ for some positive real number R. Set

$$g(x,t) = x^d f(\frac{1}{x},t),$$

then $g(x,t)$ also lies in $\mathbb{Q}[x,t]$. Moreover, $g(x,t)$ is irreducible. Indeed, suppose that

$$g(x,t) = r(x,t)s(x,t)$$

with $r(x,t)$ and $s(x,t)$ polynomials in $Q[x,t]$ of degrees m_1 and m_2 in x, respectively. Then $d = m_1 + m_2$ and

$$f(x,t) = x^d g(\frac{1}{x},t) = x^{m_1}r(\frac{1}{x},t)x^{m_2}s(\frac{1}{x},t)$$

is a factorization of $f(x,t)$ into polynomials. As $f(x,t)$ is irreducible, this must be a trivial factorization, i.e., either r or s lies in $\mathbb{Q}$ as required. Now note if $g(\alpha,t)$ is irreducible for $\alpha \in \mathbb{Q}$, so is $f(\frac{1}{\alpha},t) = \frac{1}{\alpha^d}g(\alpha,t)$. Therefore, we can replace $f(x,t)$ by $g(x,t)$ and work with power series analytic at infinity, i.e., we may assume all the root functions ρ_i of $f(x,t)$ have power series that converge for all $|x| > R$ and (i) holds for all $|x| > R$ for some real number R.

Let S be a nonempty proper subset of $\{1,\dots,n\}$ and consider the equation

$$(ii)\qquad \prod_{i=1}^{n}\bigl(t-\rho_i(x)\bigr)=\prod_{i\in S}\bigl(t-\rho_i(x)\bigr)\prod_{i\notin S}\bigl(t-\rho_i(x)\bigr).$$

We know that for every value of x with $|x|>R$, we have $f(x,t)=a_n(x)\prod_{i=1}^{n}\bigl(t-\rho_i(x)\bigr)$ is a factorization into linear terms in $\widetilde{\mathbb{Q}}[t]$, Since $f(x,t)$ is irreducible in $\mathbb{Q}[x,t]$ and $a_n(x)\in\mathbb{Q}[x]$, neither factor on the right hand side of (ii) can lie in $\mathbb{Q}(x)[t]$ by Lemma 35.7, a consequence of Gauss' Lemma, and Observation 69.7. Consequently,

$$(iii)\qquad \prod_{i\in S}\bigl(t-\rho_i(x)\bigr)=t^{|S|}+b_1t^{|S|-1}+\cdots+b_{|S|}\ \text{ in } \widetilde{\mathbb{Q}}[t],$$

with each $b_j=b_j(x)$, $1\le j\le |S|$, analytic at infinity, but at least one, say $b_i=b_i(x)$ does not lie in $\mathbb{Q}(x)$. Let N be the number of such factorizations (ii), so $N=2^{n-1}-1$. For each such S, let $y_S(x)$ be a coefficient $b_i(x)$ on the right hand side of (iii) analytic at infinity and not in $\mathbb{Q}(x)$.

Claim. Let $R'\ge R$ be arbitrary. Then there exists a rational number x_0 satisfying $x_0>R'$ (so a regular value of f) satisfying $y_S(x_0)$ is not rational for all nonempty proper subsets S of $\{1,\dots,n\}$. In particular, $f(x_0,t)$ is irreducible and the proposition is true.

Suppose that x_0 satisfies the property of the claim, but that $f(x_0,t)=r(t)s(t)$ in $\mathbb{Q}[t]$ with r and s non-constant polynomial in $\mathbb{Q}[t]$. Then there exists a nonempty proper subset S of $\{1,\dots,n\}$ satisfying

$$r(t)=a_n(x_0)\prod_{i\in S}\bigl(t-\rho_i(x_0)\bigr)\ \text{ and }\ s(t)=\prod_{i\notin S}\bigl(t-\rho_i(x_0)\bigr)$$

in $\mathbb{Q}[t]$. This means that $y_S(x_0)$ lies in $\mathbb{Q}$, a contradiction. So we need only construct x_0 as claimed, since $R'>R$ is arbitrary.

We construct an x_0 in $\mathbb{Q}$ satisfying the claim. In fact, we shall produce an integer x_0 satisfying the claim. Fix a nonempty proper S and let $y=y_S(x)$. Then y lies in $\mathbb{Q}[\rho_1(x),\dots,\rho_n(x)]\subset\widetilde{\mathbb{Q}}$, so $y(x)$ satisfies an equation

$$(iv)\qquad d_ly^l+d_{l-1}y^{l-1}+\cdots d_0=0$$

with $d_i\in\mathbb{Q}(x)$ for $i=0,\dots,l$ and d_l nonzero for some integer l. Clearing denominators, we may assume all the d_i lie in $\mathbb{Q}[x]$ and then clearing denominators again that all d_i lie in $\mathbb{Z}[x]$.

Multiplying equation (iv) by d_l^{l-1} shows that $z=z(x)=d_ly(x)$ satisfies an equation

$$z^l+b_{l-1}z^{l-1}+\cdots+b_0=0$$

with $b_0,\dots,b_{l-1}$ all lying in $\mathbb{Z}[x]$. If $\alpha>R$ satisfies $y(\alpha)$ is rational, then $z(\alpha)$ is an integer as each $b_i(\alpha)$ is an integer and any rational root of a monic equation with integer coefficients must be an integer. Therefore, we conclude that $z(x)$ has an integral value at some integer $\alpha>R$ if and only if $y(x)$ has a rational value at that α. Consequently, it suffices to work with $z(x)$. As $z = d_l y$ with $d_l \in \mathbb{Z}[x]$ and $y(x)$ analytic at infinity, we have

$$(v) \qquad z(x) = c_k x^k + \cdots + c_0 + c_{-1}\frac{1}{x} + \text{ higher powers of } \frac{1}{x}$$

with $c_i \in \mathbb{C}$ and z converges for all $|x| > R$.

We must show that $z(\alpha)$ is not an integer for some integer $\alpha > R'$ for infinitely many integers $\alpha > R$, hence for any $R' \geq R$. There are three cases to consider.

Case 1. $z(x)$ is a polynomial in $\mathbb{C}[x]$:

The polynomial $z(x)$ can have at most k integer values by Lagrange Interpolation, lest $z(x)$ hence $y(x)$ be a polynomial in $\mathbb{Q}[x]$, contrary to choice. So $z(x)$ has no integer values at integer points for sufficiently large x.

Case 2. There exists a coefficient c_j of $z(x)$ in (v) that is not real:

We may assume that j has been chosen maximal with c_j not real. Then we have, where Im is the imaginary part,

$$\lim_{x\to\infty} \operatorname{Im}\big(\frac{z(x)}{x^j}\big) = \operatorname{Im}(c_j) \neq 0,$$

hence $z(x)$ is not real for all large real values of x.

Case 3. All the c_j in (v) are real and there exist a $j > 0$ with c_{-j} not zero:

Choose $m = m(S) > 0$ such that the mth derivative of $z(x)$ has no non-negative powers of x, say

$$z^{(m)}(x) = \frac{d}{x^k} + \text{terms with larger powers of } \frac{1}{x}.$$

So d is a nonzero real number and k a positive integer. In particular, we have

$$\lim_{x\to\infty} x^k z^{(m)}(x) = d \neq 0.$$

Choose $R_1 > R'$ such that if $\alpha > R_1$, then

$$0 < |z^{(m)}(\alpha)| < \frac{2|d|}{\alpha^k}.$$

Suppose that there exist $m+1$ integers α_i satisfying

$$R_1 < \alpha_0 < \cdots < \alpha_m \text{ with } z(\alpha_i) \in \mathbb{Z} \text{ for } i = 0, \ldots, m.$$

Let

$$V_m = \det\begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^m \end{pmatrix}$$

and

$$W_m = \det\begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{m-1} & z(\alpha_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^{m-1} & z(\alpha_m) \end{pmatrix}.$$

Applying Lemma 69.6, we see that both V_m and W_m are nonzero integers and

$$\frac{2|d|}{m!\,\alpha_0^k} \geq \frac{2|d|}{m!\,\alpha^k} \geq \frac{|z^{(m)}(\alpha)|}{m!} = \frac{|W_m|}{|V_m|} \geq \frac{1}{|V_m|}$$

for some α satisfying $\alpha_0 < \alpha < \alpha_m$. Therefore,

$$\frac{m!}{2|d|}\alpha_0^k \leq |V_m| = \prod_{i<j}(\alpha_j - \alpha_i) < (\alpha_m - \alpha_0)^{\frac{m(m+1)}{2}}.$$

Consequently, there exist positive real numbers γ and δ depending on m satisfying

$$(vi) \qquad \alpha_m - \alpha_0 > \gamma\alpha_0^\delta.$$

Note that the right hand side of (vi) goes to infinity as $\alpha_0 \to \infty$, i.e., the α_m's get further and further from α_0 as it gets larger. Recall that $N = 2^{n-1} - 1$ is the number of the $y(x)$'s. (We can replace N by any larger integer.) Choose $R_2 > R_1 > R' \geq R$ satisfying

$$\gamma R_2^\delta > Nm.$$

In the above, we may also assume that $\alpha_0 > R_2$. Therefore, we have $\alpha_0 < \alpha_1 < \cdots < \alpha_m$ with each α_i and each $z(\alpha_i)$ integers, $i = 1, \dots, m$, and

$$\alpha_m - \alpha_0 > \gamma\alpha_0^\delta > \gamma R_2^\delta > Nm.$$

We conclude that among any $Nm+1$ sufficiently large consecutive integers $l, l+1, \dots, l+Nm$, the function $z(x)$ takes on integer values on at most m of these, equivalently $y(x)$ takes on rational values on at most m of these.

If $z(x)$ satisfies Case 1 or 2, it has only finitely many integer values at integer points. Choose $R_0 > R_2$ such that all such $z(x)$'s have no integer values, hence $y(x)$ no rational values, at integers greater than R_0.

For each of the finite number of S's, we have constructed an integer $m = m(S)$ and a real number $R_0 = R_0(S) > R'$ such that for any $x > R_0$ and string of $Nm+1$ consecutive integers greater than R_0, the analytic function $y_S(x)$ has at most m rational values at such a string of $Nm + 1$ consecutive integers. Let

$$M \geq \max_S \big(m(S)\big) \text{ and } R_3 > \max_S \big(R_0(S)\big).$$

For any string of $NM + 1$ consecutive integers greater than R_3, each $y_S(x)$ can have at most M rational values. In particular, in each string of $NM + 1$ consecutive integers greater than R_3, there exists an integer α such that $y_S(\alpha)$ is not rational for every one of the N functions $y_S(x)$. This establishes the claim and completes the proof. □

Corollary 69.9. *Let $f_1(x,t), \dots, f_m(x,t)$ be irreducible polynomials in $\mathbb{Q}[x,t]$. Then there exist infinitely many rational numbers α such that $f_1(\alpha,t), \dots, f_m(\alpha,t)$ are all irreducible in $\mathbb{Q}[t]$.*

Proof. By the proof of the proposition, we can work with the union of all the $y(x)$'s that arise from the $f_i(x,t)$'s with N the number of all of these. □

To prove the full version of the Hilbert Irreducibility Theorem, we use a trick developed by Kronecker.

Definition 69.10. Let F be a field, d and n positive integers. Set

$$P_d(n,F) := \{f \in F[t_1, \dots, t_n] \mid \deg_{t_i} f < d \text{ for all } i = 1, \dots, n\}$$

and

$$S_d : P_d(n,F) \to P_{d^n}(1,F)$$

defined by

$$S_d\big(f(t_1,\dots,t_n)\big) = f(t,t^d,t^{d^2},\dots,t^{d^{n-1}}).$$

For example, $S_d(t_1^{i_1}\cdots t_n^{i_n}) = t^{i_1+i_2d+i_3d^2+\cdots i_nd^{n-1}}$. Let m be a positive integer, then m has a unique d-adic expansion. In particular, applying this to all positive integers $m < d^n$, we deduce that S_d is a bijection. Although $P_d(n,F)$ is not closed under products, we do have:

Remark 69.11. If $f,g,fg \in P_d(n,F)$, then

$$S_d(fg) = S_d(f)S_d(g).$$

In particular, since S_d is a bijection, if $S_d(f)$ is irreducible in $F[t]$, then f is irreducible in $F[t_1,\dots,t_n]$.

The converse of this remark is false. For example, if $F = \mathbb{Q}$, then $f = t_1^2 + t_2^2$ in $\mathbb{Q}[t_1,t_2]$ is irreducible, but $S_3(f) = t^2 + t^6$ is not. Suppose that we have an irreducible polynomial $f \in P_d(n,F)$ but

$$S_d(f) = G(t)H(t) \text{ in } F[t] \text{ with } 0 < \deg G, \deg H < d^n,$$

i.e., $S_d(f)$ is not irreducible. Then G and H lie in P_{d^n}, so there exist unique polynomials g and h in $P_d(n,F)$ satisfying $G = S_d(g)$ and $H = S_d(h)$, i.e.,

$$S_d(f) = S_d(g)S_d(h) = S_d(gh)$$

Theorem 69.12. (Kronecker's Criterion) *Let F be a field and f an element in $P_d(n,F)$. Then f is irreducible in $F[t_1,\dots,t_n]$ if and only if for all nontrivial factorizations (i.e., polynomials of positive degree)*

$$S_d(f) = S_d(g)S_d(h)$$

with g and h in $P_d(n,F)$, the polynomial gh contains a nonzero monomial of the form $bt_1^{i_1}\cdots t_n^{i_n}$ for some $i_j \geq d$.

Proof. ($\Leftarrow$): If $f = f_1f_2$ with $f_1,f_2 \in F[t_1,\dots,t_n]$, then we have $f_1,f_2 \in P_d(n,f)$, $S_d(f) = S_d(f_1f_2) = S_d(f_1)S_d(f_2)$, and $f = f_1f_2$ has no nonzero monomial $bt_1^{i_1}\cdots t_n^{i_n}$, some $i_j > 0$.

($\Rightarrow$): Suppose that f is irreducible and $S_d(f) = S_d(f_1)S_d(f_2)$ for some $f_1,f_2 \in P_d(n,F)$. If $f_1f_2 \in P_d(n,F)$, then $S_d(f) = S_d(f_1)S_d(f_2) = S_d(f_1f_2)$. As $S_d : P_d(n,F) \to P_{d^n}(1,F)$ is one-to-one, $f = f_1f_2$, and one of f_1, f_2 lies in F. $\square$

Example 69.13. If $f = t_1^2 + t_2^2$ in $\mathbb{Q}[t_1,t_2]$, we have $S_3(f) = t^2(1+t^4)$, $S_3(g) = t^2$ with $g = t_1^2$, $S_3(h) = 1 + t^4$ with $h = 1 + t_1t_2$, and $gh = t_1^2 + t_1^3t_2$.

Corollary 69.14. *Let F be a field and $f \in F[t_0,\dots,t_n]$ irreducible with $\deg_{t_i} < d$ for $i = 1,\dots,n$. Let $S_d : P_d\big(n,F(t_0)\big) \to P_{d^n}\big(1,F(t_0)\big)$ be defined by $S_d\big(f(t_0,t_1,\dots t_n)\big) = f(t_0,x,x^d,\dots,x^{d^n-1})$. Suppose that*

$$(*) \qquad S_d(f) = \prod_{i=1}^{m} p_i(t_0,x)$$

with $p_i(t_0, x) \in F[t_0][x]$ irreducible for $i = 1, \ldots, m$. Then there exists an element $\varphi \in F(t_0)$ with the following property: If $\alpha \in F$ satisfies $\varphi(\alpha)$ is defined and nonzero and $p_i(\alpha, x)$ is irreducible in $F[x]$ for $i = 1, \ldots, n$, then $f(\alpha, t_1, \ldots, t_n)$ is irreducible in $F[t_1, \ldots, t_n]$.

PROOF. As $F[t_0]$ is a UFD, all factorizations of $S_d(f)$ into two polynomials in $F[t_0, x]$ arise from (*) (up to constants in F). View f as an irreducible polynomial in $F(t_0)[t_1, \ldots, t_n]$. By Kronecker's Criterion, any such factorization can also be written $S_d(f) = S_d(g)S_d(h)$ with $g, h \in P_d\big(n, F(t_0)\big)$ and satisfying some nonzero monomial $\varphi_{gh} t_1^{i_1} \cdots t_n^{i_n}$ occurs in gh with $\varphi_{gh} \in F(t_0)$ (and, in fact, in $F[t_0]$), for some $i_j > d$. Let φ be the product of all the φ_{gh} arising from such factorizations. If $\alpha \in F$ satisfies $\varphi(\alpha)$ is defined and nonzero and $p_i(\alpha, x)$ is irreducible in $F[x]$ for $1 \leq i \leq m$, then $S_d(f(\alpha, t_1, \ldots, t_n) = \prod_{i=1}^m p_i(\alpha, x)$ is a factorization into irreducibles in $F[x]$. It follows that $f(\alpha, t_1, \ldots, t_n)$ is irreducible in $F[t_1, \ldots, t_n]$ by Kronecker's Criterion. □

Theorem 69.15. (Hilbert Irreducibility Theorem) *Let $f(x_1, \ldots, x_n, t)$ be an irreducible polynomial in $\mathbb{Q}[x_1, \ldots, x_n, t]$. Then there exist infinitely many $(\alpha_1, \ldots, \alpha_n)$ in $\mathbb{Q}^n$ such that $f(\alpha_1, \ldots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$.*

PROOF. We induct on n. The case of $n = 1$ is Proposition 69.8. We shall apply the previous corollary and its notation to complete the proof. If $f \in P_d\big(n, \mathbb{Q}(x_1)\big)$ and $S_d : P_d\big(n, \mathbb{Q}(x_1)\big) \to P_1\big(d^n, \mathbb{Q}(x_1)\big)$, we are done if $S_d(f)$ is irreducible by induction by choosing α in $\mathbb{Q}(x_1)$ such that $\varphi(\alpha)$ is defined and nonzero. [For almost all x_0, we have $\varphi(x_0)$ is defined and nonzero.] So we may assume that we can factor $S_d(f) = \prod_{i=1}^m p_i(x_1, t)$ as in (*). By Corollary 69.9, there exist infinitely many α in $\mathbb{Q}$ satisfying $p_i(\alpha, t)$ is irreducible in $\mathbb{Q}[t]$ for $i = 1, \ldots, m$. By the previous corollary and induction, the theorem follows. □

It can be shown that the Hilbert Irreducibility Theorem holds with F replacing $\mathbb{Q}$ for any number field, i.e., any finite field extension of $\mathbb{Q}$ or more generally any non-algebraic finitely generated field extension of any field. However, it does not hold for every field replacing $\mathbb{Q}$, e.g., it does not hold for $\mathbb{C}$ (although it holds for $\mathbb{C}(t)$).

We now apply the Hilbert Irreducibility Theorem to show that S_n is a Galois group over $\mathbb{Q}$.

Theorem 69.16. *Let n be any positive integer. Then there exists a finite Galois extension $L/\mathbb{Q}$ with $G(L/\mathbb{Q}) \cong S_n$.*

PROOF. Let $K = \mathbb{Q}(t_1, \ldots, t_n)$ and S_n act on K by $\sigma(t_i) = t_{\sigma(i)}$ for $i = 1, \ldots, n$. Set $F = K^{S_n}$. By Example 57.6(3), we know that $F = \mathbb{Q}(s_1, \ldots, s_n)$ with each s_i is the ith elementary symmetric function of $t_1, \ldots, t_n$, the extension K/F is a finite Galois extension with Galois group $G(K/F) \cong S_n$, and K is a splitting field of

$$f = \prod_{i=1}^n (t - t_i) = t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n.$$

By Remark 57.10 to the Primitive Element Theorem, we can find distinct integers $m_1, \ldots, m_n$ satisfying $K = F(m_1 t_1 + \cdots + m_n t_n)$.

Let $c = m_1t_1 + \cdots + m_nt_n$. Then the $\sigma(c)$, $\sigma \in G(K/F)$ ($\cong S_n$), give $n!$ distinct elements in K. Let $g = \prod_{G(K/F)} \big(t - \sigma(c)\big) \in K[t]$. Since $\sigma(g) = g$ for all $\sigma \in G(K/F)$, we know that $g \in F[t]$ and $K = F(c)$ is a splitting field of g over F. In particular, as $\deg g = [K : F]$ and $m_F(c) \mid g$ in $F[t]$, we must have $g = m_F(c)$ is irreducible in $F[t]$. Since $s_1, \ldots, s_n$ are algebraically independent, $g(s_1, \ldots, s_n, t)$ is irreducible in $\mathbb{Q}[s_1, \ldots, s_n, t]$, so we can choose $\alpha_1, \ldots, \alpha_n$ in $\mathbb{Q}$, such that $g(\alpha_1, \ldots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$ by the Hilbert Irreducibility Theorem 69.15, i.e, we apply the evaluation map $e_{\alpha_1,\ldots,\alpha_n} : \mathbb{Q}[s_1, \ldots, s_n] \to \mathbb{Q}$ with $s_i \mapsto \alpha_i$ for each i. Let $\beta_1, \ldots, \beta_n$ be the roots of the polynomial $f(\alpha_1, \ldots, \alpha_n, t)$ in $Q[t]$ in its splitting field L in $\mathbb{C}$. Then

$$f(\alpha_1, \ldots, \alpha_n, t) = \prod_{i=1}^{n}(t - \beta_i) = t^n - \alpha_1 t^{n-1} + \cdots + (-1)^n \alpha_n$$

in $\mathbb{Q}[t]$. Let $e = m_1\beta_1 + \cdots + m_n\beta_n$ in $L = \mathbb{Q}(\beta_1, \ldots, \beta_n)$. As $c = m_1t_1 + \cdots + m_nt_n$ is a root of g with $t_1, \ldots, t_n$ the roots of f, we must have e is a root of the irreducible polynomial $g(\alpha_1, \ldots, \alpha_n, t)$. Therefore,

$$|G(L/\mathbb{Q})| = [L : \mathbb{Q}] \geq \deg m_{\mathbb{Q}}(e) = \deg g(\alpha_1, \ldots, \alpha_n, t) = n!.$$

It follows that $G(L/\mathbb{Q}) \cong S_n$, since $\deg f = n$ and we can view $G(L/F) \subset S_n$. □

Corollary 69.17. *There exists an irreducible polynomial of every degree $n \geq 5$ not solvable by radicals.*

Corollary 69.18. *Let $m \geq 2$. Then there exists an element algebraic of degree 2^m over $\mathbb{Q}$ that is not constructible.*

Proof. Let f be an irreducible polynomial of degree n in $\mathbb{Q}[t]$ with Galois group S_n with $n = 2^m$ and $K = F(\alpha)$ a splitting field of f over $\mathbb{Q}$. If α was constructible (from $z = 0, z = 1$), then $n! = [K : \mathbb{Q}] = 2^e$ for some e by the refined form of the Constructibility Criterion 60.3, which is impossible □

Exercises 69.19.

1. Let F be a field of characteristic zero and g and h two polynomials in $F[t]$. Show if the (formal) derivative $(g - h)' = 0$, then $h = g + \big(h(0) - g(0)\big)$.
2. Let F be a field of characteristic zero, h be a polynomial of degree n in $F[t]$ and t_0 an element of F. Show that the (formal) Taylor expansion of h exists at t_0, i.e.,
$$h(t) = \sum_{i=0}^{n} \frac{h^{(i)}}{i!}(t - t_0)^i$$
where $h^{(i)}$ is the ith derivative of h.
3. Let $f(x, t)$ be an irreducible polynomial in $\mathbb{Q}[x, t]$, Show that the set $\{\alpha \in \mathbb{Q} \mid f(\alpha, t) \in \mathbb{Q}[t]$ is irreducible$\}$ is dense in $\mathbb{R}$.

CHAPTER XIII

Transcendental Numbers

70. Liouville Numbers

Although we have seen that there are uncountably many real numbers transcendental over $\mathbb{Q}$, it is usually quite difficult to prove a particular real number that is not easily seen to be algebraic over $\mathbb{Q}$ is transcendental. In this section, we demonstrate that certain numbers are in fact transcendental over $\mathbb{Q}$. The first numbers shown to be transcendental over $\mathbb{Q}$ were constructed by Liouville. His approach depends on the following theorem:

Theorem 70.1. (Liouville) *Let α be a complex number algebraic over $\mathbb{Q}$ of degree $n > 1$. Then there exists a positive constant $c = c(\alpha)$ satisfying*

$$|\alpha - \frac{p}{q}| > \frac{c}{|q|^n} \quad \textit{for all integers } p \textit{ and } q \textit{ with } q \textit{ nonzero.}$$

Proof. By definition, $\deg m_{\mathbb{Q}}(\alpha) = n > 1$, so α is not a rational number. Clearing denominators, we see that there exists a nonzero integer a such that $am_{\mathbb{Q}}(\alpha)$ lies in $\mathbb{Z}[t]$. Dividing $am_{\mathbb{Q}}(\alpha)$ by its content, we see that there exists an irreducible polynomial f in $\mathbb{Z}[t]$ of degree n with $f(\alpha) = 0$. As $m_{\mathbb{Q}}(\alpha)$ has no rational roots, neither does f. (Of course, the ideals $\big(m_{\mathbb{Q}}(\alpha)\big)$ and (f) are the same in $\mathbb{Q}[t]$.) As the characteristic of $\mathbb{Q}$ is zero,

$$f'(\alpha) = \lim_{z\to\alpha}\Big|\frac{f(z) - f(\alpha)}{z-\alpha}\Big| = \lim_{z\to\alpha}\Big|\frac{f(z)}{z-\alpha}\Big|$$

is nonzero, equivalently

$$\lim_{z\to\alpha}\frac{\big|\frac{f(z)}{z-\alpha}\big|}{|f'(\alpha)|} = 1.$$

Choose $\delta' > 0$ such that if $0 < |z-\alpha| < \delta'$, then

$$\Big|\frac{f(z)}{z-\alpha}\Big| < 2|f'(\alpha)|, \quad \text{so} \quad |f(z)| < 2|f'(\alpha)||z-\alpha|.$$

Let p and q be integers with q nonzero. Let $0 < \delta < \delta'$. Suppose that $0 < |\frac{p}{q} - \alpha| \le \delta < \delta'$ (as α is not rational), then

$$|f(\frac{p}{q})| < 2|f'(\alpha)||\frac{p}{q} - \alpha|.$$

Since $\deg f = n$ and $f(\frac{p}{q})$ is nonzero, $f(\frac{p}{q}) = \frac{b}{q^n}$ for some nonzero integer b. It follows, if $|\frac{p}{q} - \alpha| \le \delta$, then

$$\frac{1}{|q^n|} \le |f(\frac{p}{q})| < 2|f'(\alpha)||\frac{p}{q} - \alpha|,$$

i.e.,

$$|\frac{p}{q} - \alpha| > \frac{1}{2|f'(\alpha)|}\frac{1}{|q^n|}.$$

Now suppose that $|\frac{p}{q} - \alpha| > \delta$. Then $\delta > \frac{\delta}{|q^n|}$, so $|\frac{p}{q} - \alpha| > \frac{\delta}{|q^n|}$. Set $c := \min\{\delta, \frac{1}{2|f'(\alpha)|}\}$. Then $|\frac{p}{q} - \alpha| > \frac{c}{|q^n|}$ for all nonzero integers q. □

Let $\alpha \in \mathbb{C}$ be a nonrational number. We say that α is a *Liouville number* if there exist no pair of real numbers $c > 0$ and $n \geq 2$ satisfying

$$|\frac{p}{q} - \alpha| > \frac{c}{|q^n|} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

Note that if α is a Liouville number, it must be real, since a nonreal complex number has positive distance from any rational number.

Corollary 70.2. *Liouville numbers are transcendental over* $\mathbb{Q}$.

Remark 70.3. Let α be a Liouville number, N a positive integer. Then there exist infinitely many rational numbers $\frac{p}{q}$, with p, q with $q \neq 0$ satisfying

$$0 < |\frac{p}{q} - \alpha| < \frac{1}{|q^N|} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

Indeed suppose this is false, then we could choose a real number $c > 0$ smaller than the distance of α to each of these finitely many rational numbers $\frac{p}{q}$, $p, q \in \mathbb{Z}$, $q \neq 0$, satisfies

$$0 < |\frac{p}{q} - \alpha| < \frac{c}{|q^N|} \text{ holds for no integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

We use this to give an alternative characterization of Liouville numbers.

Proposition 70.4. *Let α be a real number. Then α is a Liouville number if and only if for every positive integer N there exist integers p and q, with $q \geq 2$ satisfying*

$$0 < |\frac{p}{q} - \alpha| < \frac{1}{q^N}. \tag{*}$$

Proof. ($\Rightarrow$) is the remark above.
($\Leftarrow$): Let $c > 0$ and $n \geq 2$ be given. Choose an integer m so that $\frac{1}{2^m} < c$ and set $N = n + m$. By assumption, there exist integers p and q, with $q \geq 2$ satisfying

$$|\frac{p}{q} - \alpha| < \frac{1}{q^N} \leq \frac{1}{2^m q^n} < \frac{c}{q^n}.$$

By the theorem, it follows that α is a Liouville number if we can show that α is not a rational number. Suppose to the contrary that $\alpha = r/s$, with r and s integers, s positive. We may assume that we have chosen N so that it satisfies $2^{N-1} > s$. So by assumption, there exist integers p and q, with $q \geq 2$ with satisfying (*). However, we also have

$$|\frac{p}{q} - \alpha| = |\frac{p}{q} - \frac{r}{s}| = |\frac{ps - qr}{qs}| > \frac{1}{sq} > \frac{1}{2^{N-1}q} \geq \frac{1}{q^N},$$

a contradiction. □

Example 70.5. Let $\alpha = \sum_{k=1}^{\infty} \frac{1}{10^{k!}}$. Then α is a Liouville number:

Let $\sum_{k=0}^{N} \frac{1}{10^{k!}} = \frac{P_N}{10^{N!}}$, so P_N and $10^{N!}$ are positive integers satisfying

$$\begin{aligned} 0 < |\alpha - \frac{P_N}{10^{N!}}| &= \sum_{k=N+1}^{\infty} \frac{1}{10^{k!}} \leq \frac{1}{10^{(N+1)!}} \sum_{k=0}^{\infty} \frac{1}{10^{k!}} \\ &\leq \frac{1}{9}\Big(\frac{1}{(10^{N!})^{N+1}}\Big) < \frac{1}{(10^{N!})^{N}}. \end{aligned}$$

Hence α is a Liouville number.

The basic reason that the example is transcendental over $\mathbb{Q}$ is that the series converges very rapidly. Liouville numbers have this property. Of course, most transcendental numbers do not, e.g., π. Mahler showed that $|\pi - \frac{p}{q}| > q^{-42}$ if $q \geq 2$.

Much work was done to strengthen Liouville's theorem. The best theorem was established by Roth who proved the following

Theorem 70.6. (Roth's Theorem) *Let α be a complex number algebraic over the rational numbers but not rational. Suppose that there exists a non-negative real number μ satisfying*

$$|\alpha - \frac{p}{q}| < \frac{1}{q^{\mu}}$$

for infinitely many p/q with p, q relatively prime integers $q > 0$. Then $\mu \leq 2$. Moreover, this is best possible, i.e., if α is algebraic over $\mathbb{Q}$ and $\epsilon > 0$, then

$$|\alpha - \frac{p}{q}| > \frac{1}{q^{2+\epsilon}}$$

holds for all but finitely many p/q with p, q relatively prime integers $q > 0$, and there exists a positive constant $c = c(\alpha, \epsilon)$ such that

$$|\alpha - \frac{p}{q}| > \frac{c}{|q|^{2+\epsilon}} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero,}$$

i.e., is independent of $n > 0$.

Roth won the Fields Medal for this work. This subject is part of what is called diophantine approximation. The proof is delicate, but does not use a lot of theory. A proof can be found in Leveque's book *Topics in Number Theory, Volume II*, Chapter 4.

Exercises 70.7.

1. Let $b_0 < b_1 < b_2 < \cdots$ an increasing sequence of positive integers satisfying $\limsup_{k\to\infty} b_{k+1}/b_k = \infty$ and let $\{e_k\}$ be a bounded sequence of positive integers. If $a \geq 2$ is an integer, show

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{a^{b_k}} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{e_k}{a^{b_k}}$$

are Liouville numbers.

2. For any choice of signs, show that
$$1 \pm \frac{1}{2^{1!}} \pm \frac{1}{2^{2!}} \pm \frac{1}{2^{3!}g} \pm \cdots$$
is a Liouville number.

71. Transcendence of e

In this section we prove that the real number e is transcendental over $\mathbb{Q}$. The ideas give the basis for showing that π is transcendental, which we shall do in Section 73. The key to the proof is the two basic properties of the exponential function, viz., the derivative $(e^x)'$ of e^x is e^x and e^x has the Taylor expansion $\sum_{n=o}^{\infty} \frac{x^n}{n!}$, so converges rapidly because of the factorial in each term.

Theorem 71.1. (Hermite) *The real number e is transcendental over $\mathbb{Q}$.*

Proof. Suppose that there exist integers $a_0, \ldots, a_m$ not all zero satisfying
$$a_m e^m + a_{m-1}e^{m-1} + \cdots + a_1 e + a_0 = 0,$$
i.e., e is a root of the nonzero polynomial $\sum_{i=0}^{m} a_i t^i$ in $\mathbb{Z}[t]$, equivalently, e is algebraic over $\mathbb{Q}$, by clearing denominators. We may assume that a_m is nonzero and without loss of generality that a_0 is nonzero as well. Define a *polynomial function*
$$f(x) := \frac{x^{p-1}(x-1)^p(x-2)^p \cdots (x-m)^p}{(p-1)!}.$$
where $p > 2$ is a prime number to be specified later.

Note that $f(x) = \dfrac{x^{p-1}}{p-1!}g(x)$ with $g(x) = (x-1)^p(x-2)^p \cdots (x-m)^p$ satisfying $g(x) = 0$ for $x = 1, \ldots, m$.

Now set
$$F(x) := f(x) + f'(x) + f^{(2)}(x) + \cdots + f^{(mp+p-1)}(x)$$
where $f^{(j)}$ is the jth derivative of f. Note that $\deg f = mp + p - 1$, so the $(mp+p)$th derivative of f, $f^{(mp+p)}(x) = 0$. In particular, $F(x) = \sum_{i=0}^{\infty} f^{(i)}(x)$.

We shall need the following observation. For $0 < x < m$, we have the upper bound

$$|f(x)| < \frac{m^{p-1}m^p \cdots m^p}{(p-1)!} = \frac{m^{mp+p-1}}{(p-1)!}. \tag{71.2}$$

Check 71.3. We have the following:

(1)
$$\frac{d}{dx}\big(e^{-x}F(x)\big) = e^{-x}\big(F'(x) - F(x)\big) = -e^{-x}f(x)$$

(2)
$$a_j \int_0^j e^{-x} f(x)dx = a_j\big[-e^{-x}F(x)\big]_{x=0}^{x=j} = a_j F(0) - a_j e^{-j}F(j).$$

Multiplying equation (2) by e^j and adding yields

$$\begin{aligned} A := \sum_{j=0}^{m} a_j e^j \int_0^j e^{-x} f(x)dx &= \sum_{j=0}^{m} a_j e^j F(0) - \sum_{j=0}^{m} a_j F(j) \\ &= -\sum_{j=0}^{m} a_j F(j) = -\sum_{j=0}^{m} \sum_{i=0}^{mp+p-1} a_j f^{(i)}(j). \end{aligned} \tag{71.4}$$

Claim: $f^{(i)}(j)$ is an integer for all i and j and $p \mid f^{(i)}(j)$ in $\mathbb{Z}$ if $(i,j) \neq (p-1, 0)$. The claim will follow from the following:

Lemma 71.5. *Let* $h(x) = \dfrac{x^n g(x)}{n!}$ *with* g *a polynomial in* $\mathbb{Z}[t]$. *Then*

(1) The number $h^{(j)}(0)$ *is an integer for all* j.
(2) The integer $n+1$ *satisfies* $n+1 \mid h^{(j)}(0)$ *in* $\mathbb{Z}$ *if* $j \neq n$.
(3) If $g(0) = 0$, *then* $n+1 \mid h^{(n)}(0)$ *in* $\mathbb{Z}$.

Proof. Let $t^n g(t) = \sum_{j=0}^{\infty} c_j t^j$ in $\mathbb{Z}[t]$ (almost all $c_j = 0$). So we have c_j are integers and zero if $j < n$. Since $h(x) = \dfrac{x^n g(x)}{n!}$, we see that $h^{(j)}(0) = c_j j!/n!$. If $j < n$, then $c_j = 0$. Hence $h^{(j)}(0) = 0$ and if $j > n$, then $h^{(j)}(0) = c_j(j!/n!)$ in $\mathbb{Z}$. If $j = n$, then $h^{(n)}(0) = c_n$ in $\mathbb{Z}$. Finally, if g has a zero constant term, i.e., $g(0) = 0$, then $c_n = 0$ and the lemma follows. □

We return to the proof of the theorem. We apply the lemma to $f(x), f(x+1), \ldots, f(x+m)$, respectively to see that $f^{(i)}(j)$ is an integer for all i and j and $p \mid f^{(i)}(j)$ with the possible exception when $i = p-1$ and $j = 0$. Therefore, as a_i are integers, we see that A is an integer by 71.4 satisfying

$$A = -\sum_{j=0}^{m} \sum_{i=0}^{mp+p-1} a_j f^{(i)}(j) \equiv -a_0 f^{(p-1)}(0) \mod p.$$

We also have $f^{(p-1)}(0) = (-1)^p(-2)^p \cdots (-m)^p$. (This is just $g(0)$, if in the lemma, $g(x) = (x-1)^p \cdots (x-m)^p$.) In particular, if $p > m$, then $p \nmid f^{(p-1)}(0)$. Consequently, as $a_0 \neq 0$, we must have: if p is a prime satisfying $p > |a_0|$ and $p > m$, then

$$A \equiv -a_0 f^{(p-1)}(0) \not\equiv 0 \mod p.$$

In particular, $|A| \geq 1$. By (71.2) and (71.4), we have

$$\begin{aligned} |A| = \Big|\sum_{j=0}^{m} a_j e^j \int_0^j e^{-x} f(x)dx\Big| &\leq \sum_{j=0}^{m} \Big|a_j e^j \int_0^j f(x)dx\Big| \\ &\leq \sum_{j=0}^{m} |a_j| e^j j \frac{m^{mp+p-1}}{(p-1)!} \leq \sum_{j=0}^{m} |a_j| e^m m \frac{m^{mp+p-1}}{(p-1)!} \\ &= \Big(\sum_{j=0}^{m} |a_j|\Big) e^m \frac{m^{mp+p}}{(p-1)!} = \Big(\sum_{j=0}^{m} |a_j|\Big) e^m \frac{(m^{m+1})^p}{(p-1)!}. \end{aligned}$$

Choosing p sufficiently large leads to $|A| < 1$, a contradiction. $\square$

Exercise 71.6. Verify the two statements in Check 71.3.

72. Symmetric Polynomials

We discussed symmetric polynomials in the study of Galois Theory. In order to prove that π is transcendental, we shall need a finer discussion. This material is very important in its own right and is fundamental in the study of polynomials as well as its various generalizations.

We first establish some nomenclature for polynomials in finitely many variables over a commutative ring. Let R be a commutative ring and f a nonzero polynomial in $R[t_1, \ldots, t_n]$. Write

$$f = \sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n}.$$

Each nonzero term $a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n}$ in f is called a *monomial* of *total degree* $i_1 + \cdots + i_n$ of f in $t_1, \ldots, t_n$ and the *total degree* of f in $t_1, \ldots, t_n$ is defined to be the maximum of the total degree of the nonzero monomials of f in $t_1, \ldots, t_n$. For each positive integer k, order the subset

$$\{a t_1^{i_1} \cdots t_n^{i_n} \mid a \text{ nonzero in } R \text{ with } k = i_1 + \cdots + i_n\}$$

of $R[t_1, \ldots, t_n]$ *lexicographically*, i.e., if $a, b \in R$ are nonzero and $k = i_1 + \cdots + i_n = j_1 + \cdots + j_n$, then

$$a t_1^{i_1} \cdots t_n^{i_n} \leq b t_1^{j_1} \cdots t_n^{j_n}$$

if

$$\begin{aligned} &i_l = j_l \text{ for all } l \text{ or} \\ &\text{there exists an } l \text{ satisfying } i_s = j_s \text{ for all } s \leq l \text{ and } i_{l+1} < j_{l+1}. \end{aligned}$$

For example,

$$t_1 t_2 t_3^2 < t_1^3 t_2 t_3 \text{ and } t_2^{27} < t_1 t_2^{27}$$

in $R[t_1, t_2, t_3]$ where $<$ of course means $\leq$ but not $=$. We define the *leading term* of a nonzero f in $R[t_1, \ldots, t_n]$ to be the nonzero monomial of f that has maximal total degree relative to the lexicographic ordering on $R[t_1, \ldots, t_n]$. In particular, the total degree of the leading term is the same as the total degree of f, although there can be many nonzero monomials of f of the same total degree. The coefficient of the leading term is called the *leading coefficient*.

Definition 72.1. Let R be a commutative ring. A polynomial f in $R[t_1, \ldots, t_n]$ is called *symmetric* in $t_1, \ldots, t_n$ if for all permutations $\sigma \in S_n$, we have $f(t_1, \ldots, t_n) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)})$.

Example 72.2. The following are symmetric in $R[t_1, \ldots, t_n]$:

1. Any element of R.
2. $t_1^r + \cdots + t_n^r$ for every positive integer r.
3. $t_1 t_2 + t_1 t_3 + \cdots + t_1 t_n + t_2 t_3 + \cdots + t_{n-1} t_n = \sum_{i<j} t_i t_j$

4. $\displaystyle\sum_{1\le i_1<i_2<\cdots i_r\le n} t_{i_1}^s\cdots t_{i_r}^s$ for each $1 \le r \le n$ and each positive integer s.

The rth *elementary symmetric polynomials* in $t_1, \dots, t_n$ is defined to be

$$s_r(t_1,\dots,t_n) = \sum_{1\le i_1<i_2<\cdots i_r\le n} t_{i_1}\cdots t_{i_r}.$$

For example,

$$\begin{aligned}
&s_3(t_1,t_2,t_3,t_4) = t_1t_2t_3 + t_1t_2t_4 + t_1t_3t_4 + t_2t_3t_4 \text{ and}\\
&s_4(t_1,t_2,t_3,t_4) = t_1t_2t_3t_4.
\end{aligned}$$

We shall write s_r for $s_r(t_1,\dots,t_n)$ if n is clear and we set $s_0 = 1$.

Observation 72.3. Let R be a commutative ring, $S = R[t_1,\dots,t_n]$, and $s_i(t_1,\dots,t_n)$ in S. Then in $S[t]$, we have

$$\begin{aligned}
f &:= \prod(t+t_i) := t^n + s_1t^{n-1} + \cdots + s_n \ \text{ and}\\
g &:= \prod(t-t_i) := t^n - s_1t^{n-1} + \cdots + (-1)^n s_n
\end{aligned}$$

with $-t_1,\dots,-t_n$ (respectively, $t_1,\dots,t_n$) the roots of f (respectively, of g) in S. Set $S_0 = R[s_1,\dots,s_n] \subset S = R[t_1,\dots,t_n]$. Then the subset

$$S^{S_n} := \{f \in S \mid f \text{ symmetric in } t_1,\dots,t_n\}$$

satisfies $S_0 \subset S^{S_n} \subset S$. We shall show that $S_0 = S^{S_n}$.

Theorem 72.4. (Fundamental Theorem of Symmetric Polynomials) *Let R be a commutative ring, $s_i = s_i(t_1,\dots,t_n)$ in $R[t_1,\dots,t_n]$. Then*

$$R[s_1,\dots,s_n] = R[t_1,\dots,t_n]^{S_n} := \{f \in S \mid F \text{ symmetric in } t_1,\dots,t_n\}.$$

More specifically, let $f \in R[t_1,\dots,t_n]$ be symmetric in $t_1,\dots,t_n$ of total degree k in $t_1,\dots,t_n$. Then there exists a unique polynomial g in $R[s_1,\dots,s_n]$ satisfying $f = g$ and the total degree of g in $s_1,\dots s_n$ is at most k.

PROOF. **Existence**: Let $f \in R[t_1,\dots,t_n]$ be symmetric in $t_1,\dots,t_n$ of total degree k and leading term $at_1^{i_1}\dots t_n^{t_n}$ in the lexicographical ordering. In particular, $k = i_1+\cdots+i_n$. As f is symmetric, $at_{\sigma(1)}^{i_1}\dots t_{\sigma(n)}^{i_n}$ is also a nonzero monomial in f for every σ in S_n. It follows that $i_1 \ge i_2 \ge \cdots \ge i_n$. Set

$$j_l := i_l - i_{l-1} \le i_l \ \text{ for } \ l = 1,\dots,n$$

where $i_{n+1} = 0$. Therefore, $j_l \ge 0$ and $i_l = j_l + j_{l+1} + \cdots j_n$ for every l. The monomial $as_1^{j_1}\cdots s_n^{j_n}$ has total degree in $t_1,\dots,t_n$ equal to

$$j_1 + 2j_2 + \cdots + nj_n = i_1 + i_2 + \cdots + i_n = k$$

and total degree in $s_1,\dots,s_n$ equal to

$$j_1 + j_2 + \cdots + j_n \le i_1 + i_2 + \cdots + i_n = k.$$

The leading term of $as_1^{j_1}\cdots s_n^{j_n}$ as a polynomial in $t_1,\ldots,t_n$ is

$$at_1^{j_1}(t_1t_2)^{j_2}(t_1t_2t_3)^{j_3}\cdots(t_1\cdots t_n)^{j_n} = at_1^{i_1}\ldots t_n^{i_n},$$

so $f_1 := f - as_1^{j_1}\cdots s_n^{j_n}$ has leading term in $t_1,\ldots,t_n$ less than $as_1^{j_1}\cdots s_n^{j_n}$ in the lexicographic ordering. Iterating the process rids us of all monomials of total degree k in $t_1,\ldots,t_n$ except for monomials in $s_1,\ldots,s_n$ of total degree at most k in $s_1,\ldots,s_n$ that when viewed as polynomials in $t_1,\ldots,t_n$ have total degree k. Induction on the total degree of f produces the existence of the required g.

Uniqueness: Suppose that f has two such expressions. Then we would have an equation

$$\sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n}s_1^{i_1}\cdots s_n^{i_n} = 0$$

not all $a_{i_1,\ldots,i_n}$ zero. Let $as_1^{j_1}\cdots s_n^{j_n}$ be the leading term of this expression in $s_1,\ldots,s_n$. Writing this term in $t_1,\ldots,t_n$ would produce a leading term in $t_1,\ldots,t_n$

$$at_1^{i_1}\cdots t_n^{i_n} \text{ with } j_l = i_l - i_{l+1}, \ \textit{i.e.}, \ i_l = j_l + \cdots + j_n$$

just as before. As the t_i are indeterminants, we would have $a = 0$, a contradiction. $\square$

We deduce the immediate specialization of this result.

Corollary 72.5. *Let R be a commutative ring and $f \in R[t_1,\ldots,t_n]$ a symmetric polynomial. Suppose that $\alpha_1,\ldots,\alpha_n$ are elements in R and $s_i(t_1,\ldots,t_n)$ the elementary symmetric polynomials in $t_1,\ldots,t_n$. Then the evaluation $e_{\alpha_1,\ldots,\alpha_n}$ of f, $f(\alpha_1,\ldots,\alpha_n)$, lies in $R[s_1(\alpha_1,\ldots,\alpha_n),\ldots,s_n(\alpha_1,\ldots,\alpha_n)]$.*

We shall need the following consequence of this.

Corollary 72.6. *Let F be a field, K/F a field extension, and f in $F[t]$ a polynomial of degree n with roots $\alpha_1,\ldots,\alpha_n$ in K. Suppose that $P \in F[t_1,\ldots,t_n]$ is symmetric in $t_1,\ldots,t_n$. Then $P(\alpha_1,\ldots,\alpha_n)$ lies in F.*

Proof. Using Observation 72.3, we know if $f = a_nt^n + \cdots + a_0$ in $F[t]$, then

$$\pm s_{n-i}(\alpha_1,\ldots,\alpha_n) = a_i/a_n$$

lies in F. The result follows. $\square$

73. Transcendence of π

The proof of the transcendence of π over the rationals takes the approach of the analogous statement for e, but it is significantly harder. In particular, we shall need the Fundamental Theorem of Symmetric Polynomials.

We begin with the observation that a function $f : \mathbb{Z} \to \mathbb{Z}$ satisfying $f(p) \to 0$ as $p \to \infty$ must satisfy $f(p) = 0$ for large enough primes, since $|f(p)| < 1$ for all primes large enough.

We give Hilbert's proof of the following famous result.

Theorem 73.1. (Lindemann) *The real number π is transcendental over $\mathbb{Q}$.*

PROOF. Suppose that π is algebraic over $\mathbb{Q}$. Then $\alpha_1 := \sqrt{-1}\pi$ is also algebraic over $\mathbb{Q}$. We have

(i) (Euler) $$e^{\alpha_1} + 1 = 0.$$

Let $g_1 = m_{\mathbb{Q}}(\alpha_1)$ in $\mathbb{Q}[t]$. Supposes that $\deg g_1 = n$ and

$$\alpha_1, \ldots, \alpha_n \text{ are all the (distinct) roots of } g_1 \text{ in } \mathbb{C}.$$

Then by (i), we have

(ii) $$\prod_{i=1}^{n}(e^{\alpha_i} + 1) = 0.$$

Expanding (ii) yields

(iii) $$\Big(\sum_{1\le r\le n} \sum_{1\le i_1<\cdots<i_r\le n} e^{\alpha_{i_1}+\cdots+\alpha_{i_r}}\Big) + 1 = 0.$$

Let $s_j(t_1, \ldots, t_n) := \sum_{1\le i_1<\cdots<i_j\le n} t_{i_1}\cdots t_{i_j}$ be the jth elementary polynomial in $\mathbb{C}[t_1, \ldots, t_n]$. Then the monic polynomial

$$\prod_{i=1}^{n}(t - \alpha_i) = t^n - s_1(\alpha_1, \ldots, \alpha_n)t^{n-1} + \cdots + (-1)^n s_n(\alpha_1, \ldots, \alpha_n)$$

in $\mathbb{C}[t]$ has precisely $\alpha_1, \ldots, \alpha_n$ as roots. In particular, this polynomial is just g_1. It follows that each $s_j(\alpha_1, \ldots, \alpha_n)$ lies in $\mathbb{Q}$.

Fix r with $1 \le r \le n$ and let

$$T_{i_1,\ldots,i_r}, \quad 1 \le i_1 < \cdots < i_r \le n,$$

be $k := \binom{n}{r}$ independent variables, ordered lexicographically on the $i_1, \ldots, i_r$. Let s'_j, $1 \le j \le k$, be the elementary symmetric polynomials in the $T_{i_1,\ldots,i_r}$ evaluated at the corresponding $\alpha_{i_1} + \cdots + \alpha_{i_r}$. Set

$$g_r := t^k - s'_1 t^{k-1} + \cdots + (-1)^k s'_k, \text{ a polynomial in } \mathbb{C}[t].$$

Therefore, g_r has roots

$$\alpha_{i_1} + \cdots + \alpha_{i_r} \text{ for } 1 \le i_1 < \cdots < i_r \le n.$$

We can view the evaluation of the elementary symmetric polynomials in the $T_{i_1,\ldots,i_r}$ in two steps:

(a) Evaluate these elementary symmetric polynomials in the $T_{i_1,\ldots,i_r}$ at $t_{i_1} + \cdots + t_{i_r}$ for each $1 \le i_1 < \cdots < i_r \le n$.
(b) Evaluate the resulting polynomials in (a) at the corresponding $\alpha_{i_1} + \cdots + \alpha_{i_r}$.

Check 73.2. The polynomials in (a) above are symmetric in $t_1, \ldots, t_n$

It follows by Corollary 72.6, the corollary to the Fundamental Theorem of Symmetric Polynomials 72.4, that the polynomials in (a) lie in $\mathbb{Q}[s_1(t_1,\dots,t_n),\dots,s_n(t_1,\dots,t_n)]$ and

$$s'_j \in \mathbb{Q}[s_1(\alpha_1,\dots,\alpha_n),\dots,s_n(\alpha_1,\dots,\alpha_n)] \subset \mathbb{Q}.$$

Consequently, $g_r \in \mathbb{Q}[t]$ for $r = 1,\dots,n$.

Claim: There exists a polynomial g in $\mathbb{Z}[t]$ having as its roots in $\mathbb{C}$ precisely all the nonzero $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$:

Let $\widetilde{g} = g_1 \dots g_n$, a polynomial in $\mathbb{Q}[t]$. We know that $\widetilde{g}$ has as its roots all $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$. Let

$$\begin{aligned} M &:= \text{precisely the number of } (i_1,\dots,i_r),\ 1 \le i_1 < \cdots < i_r \le n \\ &\quad 1 \le r \le n,\ \text{ satisfying } \alpha_{i_1}+\cdots+\alpha_{i_r} = 0. \end{aligned}$$

Therefore, the non-negative integer M satisfies $t^M \mid \widetilde{g}$, but $t^{M+1} \nmid \widetilde{g}$ in $\mathbb{Q}[t]$. Let $\widehat{g} := \widetilde{g}/t^M$ in $\mathbb{Q}[t]$. Consequently, $\widehat{g}$ has as roots, precisely those $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$, that are nonzero. Let m be the least common multiple of all the denominators of the (reduced) coefficients of $\widehat{g}$ (i.e., if $\widehat{g} = \sum \frac{a_i}{b_i} t^i$ with a_i, b_i relatively prime integers $(b_i \ne 0)$ if $a_i \ne 0$ (and $b_i = 1$ otherwise), then m is the least common multiple of the b_i's). Set $g = m\widehat{g}$, a polynomial in $\mathbb{Z}[t]$ that satisfies the claim.

Let $r = \deg g$ and rename the roots of g by

$$\beta_1,\dots,\beta_r \ \text{ not necessarily distinct.}$$

By choice, all the β_i are nonzero, so (iii) becomes

(iv) $$e^{\beta_1} + e^{\beta_2} + \cdots + e^{\beta_r} + k = 0, \text{ for some positive integer } k.$$

Write

$$g = ct^r + c_1 t^{r-1} + \cdots + c_r, \ \text{ a polynomial in } \mathbb{Z}[t].$$

As $\deg g = r$ and $g(0)$ is not zero, we have

$$c \ \text{ and } \ c_r \text{ are nonzero integers.}$$

Now fix a prime p and set

$$s = rp - 1.$$

Define

$$\begin{aligned} f &= \frac{c^s t^{p-1} g^p(t)}{(p-1)!} \\ F &= f + f' + f^{(2)} + \cdots + f^{(s+p)} \end{aligned}$$

where $f^{(j)}$ is the jth derivative of f.

Note that $\deg f = p - 1 + pr = p + s$, so $f^{(s+p+1)} = 0$.

It follows that

$$\frac{d}{dt}\big(e^{-t}F(t)\big) = -e^{-t}f(t) \ \text{ so } \ e^{-t}F(t) - F(0) = -\int_0^t e^{-x} f(x)dx.$$

Multiply this equation by e^t and let $x = \lambda t$ (λ is a new variable), to get

(v) $$F(t) - e^t F(0) = t\int_0^1 e^{(1-\lambda)t} f(\lambda t) d\lambda.$$

Setting $t = \beta_1, \dots, \beta_r$ (one at a time) into (v) and summing, we get, using (iv),

$$\text{(vi)} \qquad \sum_{i=1}^{r} F(\beta_i) + kF(0) = \sum_{i=1}^{r} \beta_i \int_0^1 e^{(1-\lambda)\beta_i} f(\beta_i \lambda) d\lambda.$$

We shall show that this leads to a contradiction by analyzing

$$A = \sum_{i=1}^{r} F(\beta_i) + kF(0), \text{ the left hand side of (vi), and}$$

$$B = \sum_{i=1}^{r} \beta_i \int_0^1 e^{(1-\lambda)\beta_i} f(\beta_i \lambda) d\lambda, \text{ the right hand side of (vi).}$$

Claim: If $p >> 0$, then A is a non-zero integer.
To prove this, we have to first evaluate $\sum_{i=1}^{r} f^{(j)}(\beta_i)$. We show

$$\text{(a)} \qquad \sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ is an integer.}$$

$$\text{(b)} \qquad p \mid \sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ in } \mathbb{Z} :$$

As $g(\beta_i) = 0$ for all i, all terms in

$$(\star) \qquad \frac{d^j}{dt^j}\Big|_{t=\beta_i} \left(\frac{c^s t^{p-1} g^p(t)}{(p-1)!} \right)$$

having a g^l term with $l > 0$ are zero. In particular, (a) and (b) hold if $j < p$.
So we may assume that $j \geq p$.

Check 73.3. $\dfrac{1}{pc^s} f^{(j)}(t)$ lies in $\mathbb{Z}[t]$ for $j \geq p$. (Cf. Example 2.11.)

Therefore, $\dfrac{1}{pc^s} \displaystyle\sum_{i=1}^{r} f^{(j)}(t_i) \in \mathbb{Z}[t_1, \dots, t_r]$ is symmetric in $t_1, \dots, t_r$ of total degree $(p-1) + rp - j = s + p - j \leq s$ in $t_1, \dots, t_r$ as $j \geq p$. By the Fundamental Theorem of Symmetric Polynomials 72.4,

$$(\dagger) \qquad \frac{1}{pc^s} \sum_{i=1}^{r} f^{(j)}(t_i) \text{ lies in } \mathbb{Z}[s_1(t_1, \dots, t_r), \dots, s_r(t_1, \dots, t_r)]$$

and is of total degree in $s_1(t_1, \dots, t_r), \dots, s_r(t_1, \dots, t_r)$ at most $(p-1+pr)-j = s+p-j \leq s$.
We conclude by $(\dagger)$ that the sum

$$\frac{1}{pc^s} \sum_{i} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[s_1(\beta_1, \dots, \beta_r), \dots, s_r(\beta_1, \dots, \beta_r)]$$

by evaluating the t_i at the β_i. Since

$$\frac{g}{c} = t^r + \frac{c_1}{c} t^{r-1} + \cdots + \frac{c_r}{c}$$

has roots $\beta_1, \ldots, \beta_r$, we see that

$$s_{n-j}(\beta_1, \ldots, \beta_r) = (-1)^j \frac{c_j}{c} \quad \text{for } j = 1, \ldots, r,$$

so

$$\frac{1}{pc^s} \sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[\frac{c_1}{c}, \ldots, \frac{c_r}{c}].$$

Since (†) has total degree in $s_1(t_1, \ldots, t_r), \ldots, s_r(t_1, \ldots, t_r)$ at most s, it follows that

$$\frac{1}{p} \sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[c_1, \ldots, c_r] \subset \mathbb{Z},$$

i.e., $\sum_{i=1}^{r} f^{(j)}(\beta_i)$ is an element of $p\mathbb{Z}$. This shows (a) and (b).

Check 73.4. We have

$$f^{(j)}(0) = \begin{cases} 0 \quad \text{if } j \leq p-2 \\ c^s c_r^p \quad \text{if } j = p-1 \\ l_j p \quad \text{if } j \geq p \text{ some integer } l_j. \end{cases}$$

We conclude that there exists an integer N satisfying

$$\big(\sum_{i=1}^{r} F(\beta_i)\big) + kF(0) = Np + kc^s c_r^p \quad \text{in } \mathbb{Z}.$$

As $k,\ c$, and c_r are not zero, if p is chosen such that $p > \max\{k, |c|, |c_r|\}$, then

$$p \not| \big(\sum_{i=1}^{r} F(\beta_i)\big) + kF(0).$$

This shows that A is a nonzero integer for all $p >> 0$.

We turn to computing B.

Claim: $B = 0$ for all $p >> 0$.

Of course, if we show this, then we have finished the proof. For each j with $1 \leq j \leq r$, let

$$m(j) := \sup_{0 \leq \lambda \leq 1} |g(\beta_j \lambda)|.$$

Hence if $0 \leq \lambda \leq 1$, we have

$$|f(\beta_j \lambda)| \leq \frac{|c|^s |\beta_j|^{p-1} m(j)^p}{(p-1)!}.$$

Let

$$N = \max_j \Big(\sup_{0 \leq \lambda \leq 1} \big|e^{(1-\lambda)\beta_j}\big|\Big),$$

then

$$\Big|\sum_{j=1}^{r} \beta_j \int_0^1 e^{(1-\lambda)\beta_j} f(\beta_j \lambda) d\lambda\Big| \leq \sum_{j=1}^{r} \frac{|c|^s |\beta_j|^p m(j)^p}{(p-1)!} N,$$

so the integer

$$\sum_{j=1}^{r} \beta_j \int_0^1 e^{(1-\lambda)\beta_j} f(\beta_j\lambda)d\lambda \to 0 \text{ as } p \to \infty.$$

It follows (by the remark at the beginning of the section) that this sum must be zero for all $p >> 0$ □

Corollary 73.5. *It is impossible to construct a square from a given circle with the same area using only straight-edge and compass.*

Lindemann's proof of the transcendence of π was generalized by Lindemann and proved completely by Weierstraß. The proof is based on our proofs of the transcendence of e and π. To carry this out, we introduce some notions previously mentioned that we shall study more deeply later. Let Ω be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. We use this notation below. An element α of Ω is called an *algebraic number*. If, in addition, $\alpha \in \Omega$ is the root of a monic polynomial in $\mathbb{Z}[t]$, it is called an *algebraic integer*. If $\Omega/K/F$, set $\mathbb{Z}_K = \{\alpha \in K \mid \alpha \text{ is an algebraic integer}\}$. We shall see in Corollary 79.7 (where $\mathbb{Z}_K$ is the *integral closure* of $\mathbb{Z}$ in K) that $\mathbb{Z}_K$ is a domain. This is the analogue of the fact that the set of elements in Ω forms a field. For example, $\mathbb{Z} = \mathbb{Q}_{\mathbb{Z}}$ by the rational root test (Exercise 35.12(5)).

Remarks 73.6. Let α be an algebraic number.

1. Clearing denominators, we may assume that α satisfies a nonzero polynomial $a_nt^n + \cdots + a_1t + a_0 \in \mathbb{Z}[t]$. Multiplying this polynomial by a_n^{n-1} shows that $a_n\alpha$ is an algebraic integer. In particular, given a finite set S of algebraic numbers, there exist a nonzero integer N such that $N\beta$ is an algebraic integer for all $\beta \in S$.
2. If $\alpha \in \mathbb{Z}_K$ and $\sigma : K \to \Omega$ is a $\mathbb{Q}$-automorphism, then $\sigma(\alpha) \in \mathbb{Z}_{\sigma(K)}$. In particular, if $K/\mathbb{Q}$ is normal then $\sigma(\alpha) \in \mathbb{Z}_K$. This implies if $K/\mathbb{Q}$ is finite, then $N_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}_K \cap \mathbb{Q} = \mathbb{Z}$, since $\mathbb{Z}_K$ is a domain.
3. if L/K is a field extension in Ω, then $\mathbb{Z}_K = \mathbb{Z}_L \cap K$ as algebraic integers are defined by monic polynomials over $\mathbb{Z}$.

Theorem 73.7. (Lindemann-Weirstraß Theorem) *Let $\alpha_1, \ldots, \alpha_n$ be distinct algebraic numbers with $n \geq 1$ and $d_1, \ldots, d_n$ be nonzero algebraic numbers. Then*

$$\sum_{i=0}^{n} d_i e^{\alpha_i} \neq 0.$$

In particular, $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are linearly independent over Ω.

Much of the proof is analogous to the proofs of the transcendence of e and π over the rationals except for the generalization to a set of arbitrary nonzero algebraic numbers. So we leave many details to the reader.

Proof. Suppose the result is false. Then there exist algebraic numbers $d_1, \ldots, d_m$ not all zero satisfying

$$\text{(i)} \qquad \sum_{i=0}^{n} d_i e^{\alpha_i} = 0.$$

Let $K = \mathbb{Q}(d_1, \dots, d_k)$, a finite extension of $\mathbb{Q}$ say of degree m. Multiplying by an appropriate nonzero integer, we may assume that all the d_i are algebraic integers, i.e., lie in $\mathbb{Z}_K$. Let E/K be the normal closure of K/Q in Ω and $\sigma_j : K \to L$, $j = 1, \dots, m$, be the m distinct $\mathbb{Q}$-homomorphisms (as $L/\mathbb{Q}$ is separable). Then we have an equation

$$\text{(ii)} \qquad \prod_{j=1}^{m} \Big(\sum_{i=0}^{n} \sigma_j(d_i) e^{\alpha_i} \Big) = 0.$$

As $d_i \in \mathbb{Z}_L$ for all i, so is $\sigma_j(d_i)$ for all i and all j. Multiply equation (i) out. Let γ_i, $i = 1, \dots, m$, be all the distinct $\alpha_{i_1} + \dots + \alpha_{i_k}$, $1 \leq i_1 < \dots i_k \leq n$ for $k = 1, \dots, n$, that occur when multiplied out. So equation (ii) becomes

$$\text{(iii)} \qquad a_1 e^{\gamma_1} + \dots + a_m e^{\gamma_m} = 0$$

for some $a_i \in \mathbb{Z}_L$. Since the a_i are symmetric in the $\sigma_j(d_i)$, they must be rational numbers. Multiplying by a suitable positive integer (or by the Fundamental Theorem of Symmetric Polynomials 72.4), we may assume that $a_1, \dots, a_m$ are all integers not all zero in (iii).

Now let $L/E(\gamma_1, \dots, \gamma_m)$ be the normal closure of $E/\mathbb{Q}$ in Ω. Add all the unlisted conjugates of γ_i to $\gamma_1, \dots, \gamma_m$ if any. To each added γ_i, let $a_i = 0$. Changing notation again, we may assume that (iii) contains all the conjugates of all the γ_i.

Let $\gamma_j^{(l)}$ denote the lth conjugate of γ_j under the action of $G(L/Q)$. Set

$$C_l(x) := \sum_{k=1}^{m} a_k e^{\gamma_k^{(l)} x}.$$

As $\gamma_1^{(l)}, \dots, \gamma_m^{(l)}$ are distinct for each l, by linear algebra, $e^{\gamma_1 x}, \dots, e^{\gamma_m x}$ are linearly independent over $\mathbb{C}$ (proven by induction or by using the Wronkian). In particular, $C_l(x)$ is not the zero function for any l. Viewing each $C_l(x)$ in the domain $\mathbb{C}[[x]]$ of (convergent) powers at zero, we also have $C(t) := \prod_l C_l(x)$ is not the zero function. So

$$C(x) = \prod_l C_l(x) = \sum_{k=1}^{N} b_k e^{\beta_k x}.$$

with $b_i \in \mathbb{Z}$ not all zero. By hypothesis, $C(1) = 0$. Finally, choose an integer $c > 0$ such that $c\beta_i$ is an algebraic integer in $\mathbb{Z}_L$ for $i = 1, \dots, N$

We have now completed the analogue to the proofs of the transcendence of e and π before the introduction of the associated polynomials that led to a contradiction, that we now do.

Let $p > 0$ be a prime. For each r, $1 \leq r \leq N$, let

$$f_r(x) := \frac{c^{Np}}{t - \beta_r} \prod_{k=1}^{N} (t - \beta_k)^p.$$

So $s := \deg f_r = Np - 1$ for all r. Set

$$F_r(x) := \sum_{i=0}^{\infty} f^{(i)}(x) = \sum_{i=0}^{s} f^{(i)}(x)$$

where $f^{(i)}(x)$ is the ith derivative of f. Arguing as before, we see that

$$F_r(t) - e^t F_r(0) = -\int_0^t e^{t-x} f_r(x)dx.$$

This implies that

$$\begin{aligned} A_r : &= \sum_{i=1}^{N} b_k \big(e^{\beta_k} \sum_{j=0}^{s} f_r^{(j)}(0) - \sum_{j=0}^{s} f_r^{(j)}(\beta_k) \big) \\ &= -\sum_{k=1}^{N} b_k \sum_{j=0}^{s} f_r^{(j)}(\beta_k), \end{aligned}$$

as $C(1) = 0$. We now can apply our arguments before to conclude that $A := \prod_{i=1}^{N} A_i$ is an algebraic integer fixed by the Galois group $G(L/\mathbb{Q})$ so must be an integer. Moreover, for each prime p, we have $(p-1)! \mid A$ and for all sufficiently large p that $p \nmid A$. It follows that the integer A is not zero. Approximating $|A|$, as in the proof of the transcendence of e and π, shows that $|A| < 1$ for all sufficiently large p. This gives the desired contradiction. $\square$

Corollary 73.8. *Let α be a nonzero algebraic number. Then e^{α} is transcendental over $\mathbb{Q}$.*

PROOF. If $\gamma = e^{\alpha}$ is algebraic, then $e^{\alpha} - \gamma e^0 = 0$, a contradiction. $\square$

Corollary 73.9. *Let α be a nonzero algebraic number. Then the numbers e, π, $\sin\alpha$, $\cos\alpha$ are transcendental over $\mathbb{Q}$ as is $\log\alpha$ if $\alpha \neq 1$.*

PROOF. Suppose that $\log\alpha$ is algebraic. Then $\gamma = e^{\alpha}$ is transcendental, a contradiction. If $\sin\alpha$ or $\cos\alpha$ is transcendental over $\mathbb{Q}$, then the equations

$$e^0 \sin\alpha - \frac{1}{2\sqrt{-1}} e^{\sqrt{-1}\alpha} + \frac{1}{2\sqrt{-1}} e^{-\sqrt{-1}\alpha} = 0$$

$$e^0 \cos\alpha - \frac{1}{2} e^{\sqrt{-1}\alpha} - \frac{1}{2} e^{-\sqrt{-1}\alpha} = 0$$

would lead to a contradiction. $\square$

Corollary 73.10. *Let $\alpha_1, \dots, \alpha_n$ be algebraic integers that are linearly independent over $\mathbb{Q}$. Then $e^{\alpha_1}, \dots, e^{\alpha_n}$ are algebraically independent over $\mathbb{Q}$.*

PROOF. If this is false, then there exists a nonzero polynomial $f(t_1, \dots, t_n) \in \mathbb{Z}[t_1, \dots, t_n]$ such that

$$0 = f(e^{\alpha_1}, \dots, e^{\alpha_n}) = \sum_{i_1, \dots, i_n} a_{i_1, \dots, i_n} e^{i_1\alpha_1 + \cdots + i_n\alpha_n}$$

with the $a_{i_1, \dots, i_n}$ rational numbers but not all zero. By the Lindemann-Weierstraß Theorem 73.7, the integers $i_1\alpha_1 + \cdots i_n\alpha_n$ cannot be all distinct. It follows that $\alpha_1, \dots, \alpha_n$ are linearly dependent over $\mathbb{Q}$. $\square$

One of the famous Hilbert Problems is:

Problem 73.11. (Hilbert Seventh Problem) Let α and β be algebraic numbers with α not zero or one and β not rational, is it true that α^β is transcendental over $\mathbb{Q}$, e.g., $\sqrt{2}^{\sqrt{2}}$ is transcendental over $\mathbb{Q}$?

This was solved in the affirmative independently by Gelfond and Schneider. We will prove that the Hilbert Problem is true in the next section. It uses some number theory and complex analysis.

One formulation of Gelfond's and Schneider's work solving this problem is the following:

Theorem 73.12. *Suppose that $\alpha_1, \alpha_2, \beta_1, \beta_2$ are nonzero algebraic numbers. If $\log \alpha_1$ and $\log \alpha_2$ are linearly independent over $\mathbb{Q}$, then $\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2$ is not zero.*

This was generalized by Baker in the 1966, who showed

Theorem 73.13. *Let $\alpha_1, \dots, \alpha_n$ be nonzero algebraic numbers satisfying $\log \alpha_1, \dots, \log \alpha_n$ are linearly independent over $\mathbb{Q}$. Then $1, \log \alpha_1, \dots, \log \alpha_n$ are linearly independent over the quotient field of Ω.*

This theorem implies the following:

Remark 73.14. Let $\alpha_1, \dots, \alpha_n$ be nonzero algebraic numbers and $\beta_0, \dots, \beta_n$ be algebraic numbers with $\beta_0 \neq 0$. Then

(1) $\beta_0 + \beta_1 \log \alpha_1 + \cdots \beta_n \log \alpha_n \neq 0$.
(2) If, in addition, all of the β_i are nonzero, then $e^{\beta_0} \alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n}$ is transcendental over $\mathbb{Q}$.
(3) If, in addition, none of the α_i are one and $1, \beta_1, \dots, \beta_n$ are linearly independent over $\mathbb{Q}$, then $\alpha_1^{\beta_1} \cdots \alpha_n^{\beta_n}$ is transcendental over $\mathbb{Q}$.

Baker won the Fields Medal for this work.

Exercises 73.15.

1. Prove Check 73.2.
2. Prove Check 73.4.
3. Prove Check 73.3.
4. Fill in the details of the proof of Theorem 73.7.
5. Prove that the Lindemann-Weierstraß Theorem 73.7 is equivalent Corollary 73.10.
6. Let α be a nonzero algebraic number. Prove that $\tan \alpha, \sinh \alpha$, and $\tanh \alpha$ are all transcendental over $\mathbb{Q}$.
7. Let $\alpha \neq 1$ be an algebraic number. Show that the inverse trigonometric and inverse hyperbolic trigonometric functions defined at α are transcendental over $\mathbb{Q}$.
8. Let $P(t_0, \dots, t_n) \in \Omega[t_0, \dots, t_n]$ be nonzero with the property that no nonzero polynomial in $\Omega[t_0]$ divides $P(t_0, \dots, t_n)$ in $\Omega[t_0, \dots, t_n]$. Show if $\alpha_0, \dots, \alpha_n$ are algebraic numbers that are linearly independent over $\mathbb{Q}$ and β is a nonzero root of $P(t_0, e^{\alpha_1 t}, \dots, e^{\alpha_n t})$, then β is transcendental over $\mathbb{Q}$.

74. Gelfond-Schneider Theorem

In this section we shall prove the Gelfond-Schneider Theorem that shows that Hilbert's seventh problem has a positive answer, i.e., if α and β are algebraic over the rationals with $\alpha \neq 0$ or 1 and β not a rational, then α^β is transcendental over $\mathbb{Q}$. A geometric formulation of Hilbert's problem is: Suppose that the ratio of the base angle to the angle at the vertex in an isosceles triangle is algebraic but not rational. Is the ratio between base and side always transcendental? Hilbert seventh problem was proved independently by Gelfond and Schneider in 1934.

Let Ω denote the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. If $\alpha \in \mathbb{C}\setminus\Omega$. we just say α is transcendental (i.e., it is transcendental over $\mathbb{Q}$ in $\mathbb{C}$). We shall prove the Gelfond-Schneider Theorem using Schneider's approach as modified by Lang. Like the proof of e and π, this calls for delicate bounds on the size of functions. In this case, our functions will be meromorphic functions on the complex plane. We shall assume facts needed from complex analysis, but these are rather minimal. We shall also need some facts about integral elements. These will be proved later in the book but stated and are not deep.

We begin with material that we shall need in this section.

We use the notion about algebraic numbers and algebraic integers set up in the last section. So if $\Omega/K/\mathbb{Q}$ is a field extension, $\mathbb{Z}_K$ is the ring of algebraic integers in K, i.e., the set of elements in K satisfying a monic polynomial in $\mathbb{Z}[t]$. (As remarked in the previous section it is a domain by Corollary 79.7 below.) It also satisfies Remarks 73.6. If $K\mathbb{Q}$ is finite, K is called an *algebraic number field of degree* $[K : \mathbb{Q}]$.*algebraic number field* In addition, we shall need the following whose proof can be found in Corollary 80.8 below:

Theorem 74.1. *Let $K/\mathbb{Q}$ be a number field of degree s. Then $\mathbb{Z}_K$ is a free $\mathbb{Z}$-module (i.e., free abelian group) of rank s.*

A basis for Z_k is called an *integral basis* for $\mathbb{Z}_K$. For example, $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}_{\mathbb{Q}(\sqrt{-1})}$ has $\{1, \sqrt{-1}\}$ as an integral basis.
[It is not true in general if $K = \mathbb{Q}(\alpha)$ with α integral over $\mathbb{Z}$ that $Z_K = Z[\alpha]$.]

We shall use the following notation:

Notation 74.2. Let $K/\mathbb{Q}$ be a number field and $\alpha \in K$. Let $\alpha_1, \dots, \alpha_n$ be all the conjugates of α in a normal extension of $Q(\alpha)/\mathbb{Q}$, e.g., in Ω. We define

$$\boxed{\alpha} := \max\{|\alpha_i| \mid i = 1, \dots n\}.$$

Let S be a non-empty connected open subset (called a *region*) of the complex plane. A complex-valued function $f(z)$ on S is called *analytic* (or *holomorphic*) if it converges to its Taylor series on an open neighborhood for each of the points in S. In particular, $f(z)$ is infinitely differentiable. A complex-valued function is called an *entire* function if it is analytic at each point in $\mathbb{C}$. A complex-valued function $f(z)$ is called *meromorphic* if at each point in $z_0 \in S$, there exists an open neighborhood U of z_0 such that $f(z)$ is analytic at each point in $U \setminus \{z_0\}$. If f is meromorphic at z_0, then $f(z) = \sum_{n=d}^{\infty} a_n(z - z_0)^n$ with $a_d \neq 0$ for some $d \in \mathbb{Z}$ on some open neighborhood of z_0. If $d \geq 0$, then $f(z)$ is analytic at z_0 and we call z_0 a *zero* of order d of $f(z)$. If $d < 0$, we call z_0 a pole of order $|d|$ of

$f(z)$. A meromorphic function f is a quotient of two analytic functions, the denominator not the zero function, with any pole z_0 of f coinciding with the zero of the denominator at z_0 of the same order.

We shall also need a special case of the Maximum Modulus Principle that we state without proof. (It is usually proved using Cauchy's Theorem.)

Theorem 74.3. (Maximum Modulus Principle) *Let $D_R = \{z \in \mathbb{C} \mid |z)| \leq R\}$ be the closed disc of radius R in $\mathbb{C}$. If $f(z)$ is a continuous function on D_R that is analytic in the interior of D_R, i.e., in $\{z \in \mathbb{C} \mid |z| < R\}$, then $f(z)$ assumes its maximal value on D_R on its boundary $\{z \in \mathbb{C} \mid |z| = R\}$.*

We shall also need the following notation:

Let $f(z)$ be a complex-valued function on an unbounded region S in $\mathbb{C}$ and $g(x)$ a positive real-valued function defined for all $x \geq 0$. We write

$$|f(z)| = O(f(|z|) \text{ if } |f(z)| \leq Cg(|z|) \text{ for all } z \in S$$

for some constant $C > 0$.

If f is an entire function, we say f is of *order* $\leq \rho$ for $\rho \geq 0$ if $f(z) = O(e^{|z|^\rho})$. As usual, we let $f^{(i)}(z)$ denote $\dfrac{d^i}{dz^i} f(z)$ for all $i \geq 0$ (with $f(z) = f^{(0)}(z)$). If $f(z)$ is a meromorphic function on $\mathbb{C}$, and $f = g/h$ with g and h entire, $h(z)$ not the zero function, with h having zeros only at the poles of f, then we say the order f has $\leq \rho$ if g has order $\leq \rho$.

The set of analytic functions on a region S is a domain with the set of meromorphic functions on S its quotient field. As with fields, a meromorphic function $f(z)$ on $\mathbb{C}$ is called *transcendental* if it is transcendental over $\mathbb{C}(z)$. A collection of meromorphic functions $f_1, \ldots, f_n$ on $\mathbb{C}$ is called *algebraically independent over* $\mathbb{C}$, if for any nonzero polynomial $P \in \mathbb{C}[t_1, ..., t_n]$, the function $P(f_1, ..., f_n)$ is not the zero function. Otherwise, the functions are called *algebraically dependent*.

With these preliminaries, we start the proof of the Gelfond-Schneider Theorem with estimates of upper bounds of nontrivial solutions of certain systems of linear equations.

Lemma 74.4. *Let*

$$(*) \qquad \begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & 0 \\ \vdots & & & & \vdots & & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

be a system of linear equations in n unknowns $x_1, \ldots, x_n$ with $a_{ij} \in \mathbb{Z}$ not all zero, $i = 1, \ldots, m$ and $j = 1, \ldots n$. Suppose that $n > m \geq 1$. Then there exists a nontrivial solution $x_1, \ldots, x_n$ to the linear system (*) in $\mathbb{Z}$ satisfying

$$|x_j| \leq 1 + (nA)^{\frac{m}{n-m}}$$

for $j = 1, \ldots, n$ with $A = \max_{i,j} |a_{ij}|$.

PROOF. We may assume that $A > 0$, otherwise we are done. Let $M > 0$ in $\mathbb{Z}$. To each of the $(2M+1)^n$ n-tuples $\underline{w} = (w_1, \ldots, w_n) \in \mathbb{Z}^n$ with $|w_i| \leq M$ for $i = 1, \ldots, n$, there exists an m-tuple $\underline{v} = (v_1, \dots, v_m) \in \mathbb{Z}^m$ defined by $v_i = \sum_{j=1}^n a_{ij} w_j$ for $i = 1, \dots, m$. Since $|v_i| \leq nAM$, at most $(2nAM+1)^m$ of these can be distinct. Thus if

$$(i) \qquad (2nAM+1)^m < (2M+1)^n$$

two of these m-tuples say $\underline{v}$ and $\underline{v}'$ must be equal. Let $\underline{w}'$ and $\underline{w}''$ correspond to $\underline{v}$ and $\underline{v}'$ in $\mathbb{Z}^n$, respectively. Therefore, the components x_j of $\underline{w}' - \underline{w}''$ with $|x_j| \leq 2M$, $j = 1, \dots, n$, give a nontrivial solution. So we must find an M satisfying (i) and

$$(ii) \qquad 2M \leq 1 + (nA)^{\frac{m}{n-m}}.$$

Let M be the greatest integer in $\frac{1}{2} + \frac{1}{2}(nA)^{\frac{m}{n-m}}$. Then

$$2\Big(\frac{1}{2} + \frac{1}{2}(nA)^{\frac{m}{n-m}}\Big) \leq 2M \leq 2\Big(\frac{1}{2} + \frac{1}{2}(nA)^{\frac{m}{n-m}}\Big) + 2.$$

Therefore, M satisfies (ii) and

$$2M > \Big(\frac{1}{2} + \frac{1}{2}(nA)^{\frac{m}{n-m}}\Big) - 2.$$

Hence $(2M+1) > (nA)^{\frac{m}{n-m}}$ and

$$(2M+1)^n = (2M+1)^{n-m}(2M+1)^m > (nA)^m(2M+1)^m > (2nAM+1)^m$$

satisfies (i) as needed. $\square$

Lemma 74.5. *Let $K/\mathbb{Q}$ be a finite extension of degree s and*

$$(*) \qquad \begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & 0 \\ \vdots & & & & \vdots & & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & 0. \end{array}$$

be a system of linear equations in n unknowns $x_1, \dots, x_n$ with $a_{ij} \in \mathbb{Z}_K$ not all zero, $i = 1, \dots, m$ and $j = 1, \dots, n$. Suppose that $n > sm$. Then there exists a nontrivial solution $(x_1, \dots, x_n)$ to the linear system (*) in $\mathbb{Z}_K^n$ satisfying

$$|x_i| \leq 2(cnA)^{\frac{sm}{n-sm}}$$

with $A = \max_{i,j} \overline{|\, a_{ij} \,|}$ and $c > 0$ a constant depending only on K.

Proof. Let $\{v_1, \dots, v_s\}$ be an integral basis for $\mathbb{Z}_K$ in K. Then there exist $b_{ijk} \in \mathbb{Z}$ such that

$$(\dagger) \qquad a_{ij} = \sum_{k=1}^s b_{ijk} v_k.$$

Hence, the system of linear equations (*) is equivalent to the system of linear equations

$$\sum_{j=1}^n b_{ijk} x_j = 0, \quad i = 1, \dots, m \text{ and } k = 1, \dots, s.$$

Therefore, to prove the lemma, it suffices to find a constant c such that

$$cA \geq B := \max_{i,j,k} |b_{ijk}| \geq 1.$$

Let $T : K \times K \to \mathbb{Q}$ be the map given by $(x, y) \mapsto \mathrm{Tr}_{K/\mathbb{Q}}(xy)$. As $K/\mathbb{Q}$ is separable, this map is non-degenerate, i.e., $T(x, y) = 0$ if and only either $x = 0$ or $y = 0$ by Dedekind's Lemma 54.3. By Lemma 80.1, there exist $w_1, \ldots, w_n$ in K satisfying $T(v_i, w_j) = \delta_{ij}$. Hence

$$b_{ijk} = \mathrm{Tr}_{K/\mathbb{Q}}(a_{ij}) = \sum_{\sigma} \sigma(a_{ij})\sigma(w_k)$$

by (†), where σ run over all the distinct $\mathbb{Q}$-homomorphisms $\sigma : K \to L$ and L/K is the normal closure of $K/\mathbb{Q}$. Therefore, $B \leq \sum_{k=1}^{s} \overline{|w_k|} A$, with $c = \overline{|w_k|}$. □

The key to Schneider's proof of the Gelfond-Schneider Theorem is his theorem that bounds m points in $\mathbb{C}$ that are not poles of two algebraically independent meromorphic functions both of order at most ρ if certain algebraic and analytic conditions are met. Explicitly,

Theorem 74.6. (Schneider) *Let $f_1(z)$ and $f_2(z)$ be meromorphic functions on the whole complex plane each of order $\leq \rho$ and $z_1, \ldots, z_m$ be distinct points in $\mathbb{C}$ that are not poles of f_1 or f_2. Suppose that f_i, $i = 1, 2$, satisfy all of the following conditions:*

1. *$f_i^{(k)}(z_l)$, $i = 1, 2$, lies in K for all $k \geq 0$ and $l = 1, \ldots, m$ with $K/\mathbb{Q}$ finite of degree s.*
2. *There exists $b \in \mathbb{Z}^+$ satisfying $b^{k+1} f_i^{(k)}(z_l) \in \mathbb{Z}_K$ for all i, k, l.*
3. *There exists a constant $\eta \in \mathbb{Z}^+$ such that $\overline{|f_i^{(k)}(z_l)|} = O(k^{\eta k})$.*
4. *f_1 and f_2 are algebraically independent.*

Then $m \leq \rho(4s\eta - 2\eta + 2s + 1)$.

We will prove the theorem in a number of steps. The idea is to define a function $F(z)$ that satisfies $F(z) = P\big(f_1(z), f_2(z)\big)$ where P is a polynomial in $\mathbb{Z}[t_1, t_2]$ having conditions bounding its coefficients in such a way that $F(z)$ has zeros of a high order depending on $\deg P$ (the total degree) at all the points $z_1, \ldots, z_m$ with $f_1(z)$ and $f_2(z)$ satisfying all the conditions of Schneider's Theorem 74.6.

We begin with a lemma extending the previous two lemmas when conditions of Schneider's Theorem 74.6 holds.

Lemma 74.7. *Let $f_1(z)$ and $f_2(z)$ be meromorphic functions on the whole complex plane each of order $\leq \rho$ and $z_1, \ldots, z_m$ be distinct points in $\mathbb{C}$ that are not poles of f_1 or f_2. Suppose that the conditions* (1) *(with $K/\mathbb{Q}$ finite of degree s),* (2)*, and* (3) *of Schneider's Theorem 74.6 holds with the same notation as in that theorem. Let M be a positive integer satisfying*

$$M \mid 2sm \ \textit{and} \ N := \frac{M^2}{2sm}.$$

and

$$F(z) = \sum_{i,j=1}^{M} c_{ij} f_1^i(z) f_2^j(z)$$

with integer coefficients c_{ij}, $i, j = 1, \ldots, M$. Suppose that $F(z)$ has a zero of order at least N at each z_i, $i = 1, \ldots, m$, i.e, $F^{(k)}(z_l) = 0$ for $k = 0, \ldots, N-1$ and $l = 1, \ldots, m$. Then there exist $c_{ij} \in \mathbb{Z}$ not all zero satisfying

$$|c_{ij}| < \delta^N N^{(\frac{1}{2}+\eta)N}$$

for some constant δ independent of M and N.

Proof. The requirement that the $F^{(k)}(z_l) = 0$ for $k = 0, \ldots, M-1$ and $l = 1, \ldots, m$ is equivalent to the system of N linear equations

$$(i) \qquad L_{kj} := \sum_{i,j=1}^{M} c_{ij}(f_1^i f_2^j)^{(k)}(z_l) = 0, \quad k = 0, \ldots, N-1, \quad l = 1, \ldots, m$$

in the M^2 variables c_{ij}, has a nontrivial solution in integers.

By Leibnitz Formula for differentiation, we see that the term $(f_1^i f_2^j)^{(k)}(z_l)$ can be expressed as a sum of $(i+j)^k$ terms, each having the form of a constant times

$$(ii) \qquad f_1^{(k_1)} \cdots f_1^{(k_i)} f_2^{(k_{i+1})} \cdots f_2^{(k_{i+j})}(z_l) \;\text{ with }\; k_1 + \cdots + k_{i+j} = k.$$

With η as in condition (3) of Schneider's Theorem 74.6 and the definition that $N = \dfrac{M^2}{2sm}$, we see that there exists constants C_1 and C_2 satisfying:

$$(74.8) \qquad \begin{aligned} \left| (f_l^i f_2^j)^{(k)}(z_l) \right| &\leq (i+j)^k C_1^{i+j} k_1^{\eta k_1} \cdots k_{i+j}^{\eta k_{i+j}} \\ &\leq (2M)^k C_1^{2M} k^{\eta(k_1+\cdots+k_{i+j})} = (2M)^k C_1^{2M} k^{\eta k} \\ &< C_2^N N^{\frac{N}{2}} N^{\eta N}. \end{aligned}$$

By condition (2) of Schneider's Theorem 74.6 that there exists a nonzero integer b such that $b^{k_n+1} f_i^{(k_n)}(z_l) \in \mathbb{Z}_K$ for all i, k_n and l. It follows by equation (ii) that

$$(iii) \qquad a_{ijkl} = b^{i+j+k}(f_1^i f_2^j)^{(k)}(z_l) \in \mathbb{Z}_K$$

for all i, j, k, l. So the system of linear equations (i) is equivalent to the linear system

$$(iv) \qquad b^{i+j+k} L_{kj} := \sum_{i,j=1}^{M} c_{ij} a_{ijkl} = 0, \quad k = 0, \ldots, N-1, \quad l = 1, \ldots, m,$$

with

$$(v) \qquad \left| a_{ijkl} \right| \leq b^{i+j+k} C_2^N N^{(\frac{1}{2}+\eta)N} < C_3^N N^{(\frac{1}{2}+\eta)N}$$

for all i, j, k, l for some constant C_3.

We can now apply Lemma 74.5 with the integer n in that lemma with our M^2 and the m in that lemma with our our mN and the same constant $c > 0$ arising in that lemma. Hence $N > sM$ and $\frac{sM}{N-sM} = 1$ assures a nontrivial solution to the linear system (iv), hence to the linear system (i) with

$$(74.9) \qquad |c_{ij}| < 2cM^2 C_3^N N^{(\frac{1}{2}+\eta)N} < \delta^N N^{(\frac{1}{2}+\eta)N}$$

for some constant δ as needed. $\square$

We now want to estimate lower and upper bounds for $|F^q(z_l)|$ for suitable $q \geq N$ where $F(z)$ is as in Lemma 74.7 under further conditions that will lead to a proof of Schneider's Theorem 74.6 if we also assume condition (4), i.e., $f_1(z)$ and$f_2(z)$ are algebraically independent.

In the proof of these bounds, we shall use the following notation:

Notation 74.10. The function $\tau_i = \tau_i(q)$ will denote a function of $q \in \mathbb{Z}^+$ satisfying $\lim_{q\to\infty} \tau_i(q)/q^{\epsilon q} = 0$ for all $\epsilon > 0$. [The τ_i will come up sequentially in the proofs below.]

Lemma 74.7 allows us to obtain a lower bound given by the following proposition.

Proposition 74.11. *Let all conditions of Lemma 74.7 hold. In addition, suppose for some l that $F^{(q)}(z_l)$ is nonzero at q. Hence by assumption $q \geq N$. Then there exists $\tau_2 = \tau_2(q)$ such that*

$$|F^{(q)}(z_l)| \geq \tau_2 q^{(1+2\eta)(s-1)q}.$$

Proof. By equations (74.8) and (74.9) in the proof of Lemma 74.7, we see that (as there are M^2 terms)

$$\overline{\left| F^{(q)}(z_l) \right|} < M^2 \delta^N N^{(\frac{1}{2}+\eta)N} (2M)^q C_1^{2M} q^{\eta q}.$$

Since $M = \sqrt{2sm}$ and $N \leq q$, we deduce that

$$\overline{\left| F^{(q)}(z_l) \right|} \leq \tau_1 q^{(1+2\eta)q}$$

for some $\tau_1 = \tau_1(q)$. As $F^{(q)}(z_l)$ has s conjugates whose product is its nonzero norm, it follows that

$$|F^{(q)}(z_l)| \geq \frac{|N_{K/\mathbb{Q}}\big(F^{(q)}(z_l)\big)|}{\tau_1^{s-1} q^{(1+\eta)q(s-1)}}.$$

By condition (2) of Schneider's Theorem 74.6 and Lemma 74.7 (and equations (*ii*) and (*iii*) in its proof), there exists a positive integer b such that $0 \neq b^{2M+q} F^{(q)}(z_1)$ lies in $\mathbb{Z}_K$. Therefore,

$$|N_{K/Q}\big(F^{(q)}(z_l)\big)| \geq \frac{1}{b^{(2M+q)s}}.$$

Hence

$$|F^{(q)}(z_1)| > \frac{1}{\tau_2 q^{(1+2\eta)(s-1)q}}$$

for some $\tau_2 = \tau_2(q)$ as needed. □

Next we compute an upper bound.

Proposition 74.12. *Let all conditions of Lemma 74.7 hold. In addition, suppose that $F^{(q)}(z_l)$, $l = 1, \dots, m$, has zeros of order $\geq q$ at each z_l, $l = 1, \dots, m$. Then there exists an l, $1 \leq l \leq m$, satisfying*

$$|F^{(q)}(z_l)| < \tau_6 q^{(\frac{3}{2}+\eta-\frac{m}{2\rho})q}$$

for some $\tau_6 = \tau_6(q)$.

Proof. By condition (4) of Schneider's Theorem 74.6, f_1 and f_2 are algebraically independent (which we did not use to obtain the lower bound). In particular, $F(z)$ cannot be the zero function. So in Lemma 74.7, we may assume at least one of the integers c_{ij} is not zero. Choose $q \geq N$ such that $F^{(q)}(z_l)$ has a zero of order $\geq q$ and exactly q for some l. Changing notation, we may assume that $l = 1$. The Taylor series of F at z_1 is

$$F(z) = \frac{F^{(q)}(z_1)}{q!}(z - z_1)^q + \cdots \quad \text{with} \quad F^{(q)}(z_1) = \frac{q!\, F(z)}{(z - z_1)^q}\Big|_{z=z_1}. \tag{1}$$

The hypotheses of Schneider's Theorem 74.6 insures that there exist entire functions h_1, h_2 of order $\leq \rho$ with $h_i(z_1) \neq 0$ for $i = 1, 2$ and $h_i f_i$ entire functions for $i = 1, 2$.

Set $H = h_1^M h_2^M$. Then HF is an entire function with zeros of order $\geq q$ at z_l, $l = 1, \ldots, m$. Let

$$G(z) = \frac{H(z)F(z)}{\prod_{l=1}^m (z - z_l)^q}. \tag{2}$$

Then G is an entire function satisfying

$$|F^{(q)}(z_1)| = |G(z_1)|\,\Big|\frac{q! \prod_{l=2}^m (z_1 - z_l)^q}{H(z_1)}\Big|. \tag{3}$$

by equation (1).

By Sterling's Formula (i.e., $\lim_{q\to\infty} \frac{q!}{\sqrt{2\pi} q^{(q+1)/2} e^{-q} = 1}$), we have the second factor of the right hand side of equation (3) satisfies

$$\Big|\frac{q! \prod_{l=2}^m (z_1 - z_l)^q}{H(z_1)}\Big| < \tau_3 q^q. \tag{4}$$

for some $\tau_3 = \tau_3(q)$. So we are reduced to determining an upper bound for $|G(z_1)|$. By the special case of the Maximum Modulus Principle 74.3,

$$|G(z_1)| \leq \max_{|z|=R} |G(z)| \text{ for all } R > |z_1|.$$

Assuming that $R \geq 1 + \max_l |z_l|$, we have the denominator of the righthand side of (2) satisfies

$$\Big|\prod_{l=1}^m (z - z_l)^q\Big| \geq (DR)^{mq}, \tag{5}$$

for some constant $D > 0$ independent of R.

As $|c_{ij}| \leq \delta^N N^{(\frac{1}{2}+\eta)N}$ by Lemma 74.7 and $h_i, h_i f_i$, $i = 1, 2$, are all of order $\leq \rho$, there exists a constant C_1 such that for z satisfying $|z| = R$, we have

$$|H(z)F(z)| \leq M^2 \delta^N N^{(\frac{1}{2}+\eta)N} (C_1 e^{R^\rho})^{2M} \leq \tau_4 q^{(\frac{1}{2}+\eta)q} e^{2MR^\rho} \tag{6}$$

for some $\tau_4 = \tau_4(q)$. Using equations (2)-(6), we see that

$$|F^{(q)}(z_1)| < \frac{\tau_5 q^{(\frac{3}{2}+\eta)q} e^{2MR^\rho}}{R^{mq}} \tag{7}$$

for some $\tau_5 = \tau_5(q)$. Set $R = q^{\frac{1}{2}\rho}$. Then for all sufficiently large $M > 0$, also N will be large enough so that with $q > N$ and all the above will hold. In particular, equation (7) becomes

$$|F^{(q)}(z_1)| < \tau_6 q^{(\frac{1}{2}+\eta-\frac{m}{2\rho})q}$$

for some $\tau_6 = \tau_6(q)$ as desired. □

PROOF. (of Schneider's Theorem 74.6). Under the hypotheses of Schneider's Theorem 74.6, by the lower and upper bounds on $|F^{(q)}(z_1)|$ established in the last two propositions with $M > 0$ sufficiently large, we can make N, hence $q \geq N$ sufficiently large so that both of these bounds hold simultaneously. Hence

$$\frac{3}{2} + \eta - \frac{m}{2\rho} \geq -(1+2\eta)(s-1),$$

i.e, $m \leq \rho(4s\eta - 2\eta + 2s + 1)$, as needed. □

We want to use Schneider's Theorem 74.6 to prove the Gelfond-Schneider Theorem that solves Hilbert's Seventh Problem. We shall show that the meromorphic functions in Schneider's Theorem 74.6 have further analytic properties. In particular, we want the following local conditions at a point z_0 to hold:

Conditions 74.13. Let$f_1, \dots, f_n$ be meromorphic functions and $z_0 \in \mathbb{C}$. Then $f_1, \dots, f_n$ satisfy all of the following:

(*i*) $K/\mathbb{Q}$ is a finite algebraic extension in $\mathbb{C}$.
(*ii*) $f_i(z)$, $i = 1, \dots n$, are defined and analytic on an open neighborhood of z_0.
(*iii*) $f_1(z_0), \dots, f_n(z_0)$ all lie in K.
(*iv*) The derivative d/dz maps the ring $K[f_1, \dots f_n]$ to itself, i.e.,

$$\frac{d}{dz} : K[f_1, \dots, f_n] \to K[f_1, \dots, f_n].$$

This will be applied in the following way.

Construction 74.14. Let $f(z)$ be a meromorphic function that satisfies a differential equation

$$\frac{d^n f}{dz^n} = f^{(n)}(z) = P(f^{(n-1)}, f^{(n-2)}, \dots, f', f) \tag{74.15}$$

with $P \in \Omega[t_1, \dots, t_n]$. Suppose that $f(z)$ is analytic at z_0 and satisfies $f(z_0), \dots, f^{n-1}(z_0)$ are all algebraic over $\mathbb{Q}$. Let $K/\mathbb{Q}$ be the algebraic number field generated by the (finitely many) coefficients of P and the $f^{(i)}(z_0)$, $i = 0, \dots, n-1$. Let U be an open neighborhood of z_0 on which f is analytic and defined.

We shall show that if $f_1, \dots, f_n$ satisfy Conditions 74.13 at z_0, i.e., such a K/F finite exists, then f_i, $i = 1, \dots, n$, satisfy conditions (1), (2), and (3) in Schneider's Theorem 74.6 at z_0.

Note that Condition 74.13(*iv*) says that there exist $P_1, \dots, P_n \in K[t_1, \dots, t_n]$ such that

$$f_i' = \frac{df_i}{dz} = P_i(f_1, \dots, f_n) \text{ for each } i = 1, \dots, n. \tag{74.16}$$

This is where Condition 74.13(iv) is crucial. Upon taking derivatives, we only want algebraic numbers arising as coefficients. For example, if $f_1 = e^{\alpha z}$ with α algebraic, then $P \in K[t]$ and K will have to contain α. We now how show we use Condition 74.13(iv) in this construction.

A map $D : A \to A$ of an R-algebra is called a *derivation* if it is an R-module map that satisfies Leibnitz's Rule $D(ab) = D(a)b + aD(b)$ for all $a, b \in A$. For example, Condition 74.13(iv) implies that

$$D : K[t_1, \dots, t_n] \to K[t_1, \dots, t_n] \text{ given by } Dt_i = P_i$$

is a well-defined K-derivation. In particular,

$$\text{if } Q \in K[t_1, \dots, t_n], \text{ then } DQ = \sum_{i=1}^{n} \frac{\partial Q}{\partial t_i} P_i. \tag{74.17}$$

Therefore, if

$$g(z) = Q\big(f_i(z), \dots, f_n(z)\big), \tag{74.18}$$

we have

$$g^{(k)} = \frac{d^k}{dz^k} g = D^k Q(f_1, \dots, f_n) \text{ for } l = 0, \dots, n. \tag{74.19}$$

Condition 74.13(iii) says $g^{(k)}(z_0) \in K$ for all $k \geq 0$ for some $K/\mathbb{Q}$ finite. If $g = f_i$ with $i = 1$ or 2, the meromorphic functions in Schneider's Theorem 74.6, then this is precisely condition (1) of Schneider's Theorem 74.6.

Let $d = \max(\deg Q, \deg P_1, \dots, \deg P_n)$ (where degree is the total degree) and $b > 0$ is an integer such that for any α, a coefficient of Q or a coefficient of any the P_i with $e_1 + \cdots + e_n \leq d$, then $b\big(\alpha f_1^{e_1} \cdots f_n^{e_n}(a_0)\big)$ lies in $\mathbb{Z}_K$. It follows by induction that

$$b^{k+1} D^k Q\big(f_1(z_0), \dots, f_n(z_0)\big) = b^{k+1} g^{(k)}(z_0) \text{ lies in } \mathbb{Z}_K \text{ for all } k \geq 0.$$

For $g = f_i$, $i = 1$ or 2, in Schneider's Theorem 74.6, this is precisely condition (2).

We turn to showing that condition (3) of Schneider's Theorem 74.6 holds, i.e.,

$$\overline{\left| g_l^{(k)}(z_0) \right|} = O(k^{\eta k}),$$

if Conditions 74.13 holds. To do this, we need a new relation.

Definition 74.20. Let $R = \sum_{i_1, \dots, i_n} r_{i_1, \dots, i_n} t_1^{i_1} \cdots t_n^{i_n} \in \mathbb{C}[t_1, \dots, t_n]$ and $S = \sum_{i_1, \dots, i_n} s_{i_1, \dots, i_n} t_1^{i_1} \cdots t_n^{i_n} \in \mathbb{R}[t_1, \dots, t_n]$. We write $R << S$ if $|r_{i_1, \dots, i_n}| \leq s_{i_1, \dots, i_n}$ for all $i_1, \dots, i_n$.

Remark 74.21. The relation $<<$ preserves addition and multiplication.

Computation 74.22. Suppose in the definition of $<<$ that $\deg R = r$. Then $R << C(1 + t_1 + \cdots + t_n)^r$ for some constant $C > 0$. In particular, if the $P_1, \dots, P_n$ are as in Construction 74.14, there exists a constant $B > 0$ and an $h \in \mathbb{Z}^+$ satisfying

$$P_i << B(1 + t_1 + \dots t_n)^{h+1}$$

for all $i = 1, \dots, n$ using Remark 74.21. We also have

$$\frac{\partial P_i}{\partial t_i} << rC(1 + t_1 + \cdots t_n)^{r-1}.$$

By Remark 74.21 and equation (74.17), we have

$$DR << nrCB(1 + t_1 + \cdots t_n)^{r+h}.$$

Hence by induction on k, we have

$$\begin{aligned} D^k R &<< Cn^k B^k \big((r)(r+h)\cdots(r+(k-1)h)\big)(1 + t_1 + \cdots + t_n)^{r+kh} \\ &<< C_R^k k!(1 + t_1 + \cdots + t_n)^{r+kh}. \end{aligned}$$

for some constant $C_R > 0$.

An easy application of Computation 74.22 upon an appropriate evaluation, now shows that condition (3) of Schneider's Theorem 74.6 holds under the assumption that Conditions 74.13 hold. In particular, if we assume that condition (4) of Schneider's Theorem 74.6 holds, we have the following

Theorem 74.23. (Lang) *Let $f_1(z), \dots, f_n(z)$ be meromorphic functions on the whole complex plane of order $\leq \rho$ and $K/\mathbb{Q}$ a finite extension of degree s with $\dfrac{d}{dz}$ a K-derivation on $K[f_1, \dots, f_n]$. Let $z_1, \dots, z_m$ be distinct points in $\mathbb{C}$ none of which are poles of f_i, $i = 1, \dots, n$ with $f_i(z_l) \in K$ for $i = 1, \dots, n$ and $l = 1, \dots, m$. Suppose at least two of $f_1, \dots, f_n$ are algebraically independent. Then $m \leq \rho(6s - 1)$.*

Proof. Our hypothesis includes condition (4) of Schneider's Theorem 74.6. Let R be as in Computation 74.22. In equations (74.17) – (74.19) of Construction 74.14, suppose that $R = Q$. Then

$$\begin{aligned} \overline{\left| g^{(k)}(z_0) \right|} &= \overline{\left| D^k Q(f_1(z_0), \dots, f_n(z_0)) \right|} \\ &\leq C_Q^k k!(1 + \overline{\left| f_1(z_0) \right|} + \cdots + \overline{\left| f_n(z_0) \right|}\,)^{r+kh} \leq C^k k!. \end{aligned}$$

for some constant C. Since $k! \leq k^k$, we have

$$\overline{\left| g^{(k)}(z_0) \right|} = O(k^{\eta k}) \text{ for all } \eta > 1.$$

This shows that condition (3) of Schneider's Theorem 74.6 holds. Since condition (3) holds for all $\eta > 1$, it follows that $m \leq \rho(6s - 1)$. $\square$

Lang's Theorem 74.23, shows if that if the conditions of Schneider's Theorem 74.6 holds, that there is a strict bound on the the number of points solving all the hypotheses of the theorem. In particular, if Conditions 74.13 hold, but algebraic independence condition in the theorem, e.g., condition (4) of Schneider's Theorem 74.6 does not hold, can lead to a transcendental element. So to apply Lang's Theorem 74.23, we need a property that will allow us to determine a transcendental element. Such a condition is given by the following theorem.

Theorem 74.24. *Let $K/\mathbb{Q}$ be a finite extension and $f_1(z), \dots, f_n(z)$, $n \geq 2$, be meromorphic functions on the whole complex plane of order $\leq \rho < \infty$. Suppose that*

$$\frac{d}{dz} : K[f_1, \dots, f_n] \to K[f_1, \dots, f_n]$$

is a K-derivation with at least two of the $f_1, \dots, f_n$ are algebraically independent over K. Suppose, in addition, that there exists $R_i \in K(t_1, \dots, t_n, t'_1, \dots, t'_n)$ (i.e., rational functions), where $t_1, \dots, t_n, t'_1, \dots, t'_n$ are $2n$ independent variables, satisfying for all $x, y \in \mathbb{C}$ and $i = 1, \dots, n$ the equation

$$(*) \qquad f_i(x+y) = R_i\big(f_1(x), \dots, f_n(x), f_1(y), \dots, f_n(y)\big).$$

[We allow the case that $f_i(x+y) = \infty$.] Let $0 \neq \alpha \in K$ not be a pole of any of the f_i, $i = 1, \dots, n$. Then at least one of $f_1(\alpha), \dots, f_n(\alpha)$ is transcendental over K.

PROOF. We first note that if $x \in K$. Then the minimal polynomial $m_{\mathbb{Q}}(x) \in \mathbb{Q}[t]$ has at most $\deg m_{\mathbb{Q}}(x)$ distinct roots in K. Now suppose that all the $f_i(\alpha)$ are algebraic over $\mathbb{Q}$, Enlarging K, we may assume without loss of generality that $f_i(\alpha) \in K$ for all i. By (*), it follows that $f_i(\lambda\alpha) \in K$ for all $i = 1, \dots, n$ and all $\lambda \in \mathbb{Z}^+$. It follows by the beginning note that the denominators of R_i, $i = 1, \dots, n$, have at most finitely many zeros, hence finitely many poles. Omitting all such poles, the result follows by Lang's theorem. $\square$

For example, if we want to apply the theorem to the example $K = \mathbb{Q}$, $f_1(z) = z$, $f_2(z) = \cos z$, $f_3(z) = \sin z$, we know that $R_1 = t_1 + t'_1$, $R_2 = t_2 t'_2$, and $R_3 = t_2 t'_3 + t_3 t'_2$, then the theorem would apply for algebraic $\alpha \neq 0$, if at least one of $f_1(\alpha), f_2(\alpha), f_3(\alpha)$, was transcendental over $\mathbb{Q}$ if we knew that two of these f_1, f_2, f_3 were algebraically independent. (This is, in fact, true.)

We show an easier case (whose conclusion is really Lindemann's Theorem).

Corollary 74.25. *Let $K/\mathbb{Q}$ be finite, $f_1(z) = z$, $f_2(z) = e^z$, $R_1 = t_1 + t'_1$, and $R_2 = t_2 t'_2$. Then for any nonzero α, one of α and e^α is transcendental over K.*

PROOF. We need to show that z and e^z are algebraically independent. Suppose this is false. Then there exists an identity

$$(*) \qquad z^n P_0(e^z) + z^{n-1} P_1(e^z) + \cdots P_n(e^z) = 0$$

with $P_i \in K[t]$ and P_0 not zero. Choose α satisfying $P_0(e^\alpha)$ is not zero. Let $z_k = \alpha + 2\pi\sqrt{-1}\,k$, $k = 0, 1, 2, \dots$. Evaluating (*) at z_k shows that

$$z_k^n P_0(e^\alpha) + z_k^{n-1} P_1(e^\alpha) + \cdots + P_n(e^\alpha) = 0$$

for all k. In particular, $t^n P_0(e^\alpha) + t^{n-1} P_1(e^\alpha) + \cdots + P_n(e^\alpha) = 0$ has infinitely many roots, which is impossible. Therefore, by Theorem 74.24, for any nonzero α, at least one of α, e^α is transcendental. $\square$

Examples 74.26. 1. Setting $\alpha = 1$ in the corollary, we see that e is transcendental over $\mathbb{Q}$, proving Hermite's Theorem.

2. Let β be algebraic and nonzero. Since $\beta = e^{\log \beta}$, any nonzero value of $\log \beta$ is transcendental over $\mathbb{Q}$. In particular, setting $\beta = 1$, we see that $2\pi n\sqrt{-1}$ is transcendental over $\mathbb{Q}$. for all non-zero n. In particular, π is transcendental over $\mathbb{Q}$ That was Lindemann's application of his theorem.

3. Let $\beta \in \Omega$ be nonzero. Suppose that $S \subset \mathbb{C}$ and φ is an arbitrary function defined on S satisfying $\varphi(z)$ and $e^{\beta z}\big|_S$ are algebraically dependent over Ω. Then $\varphi(\alpha)$ is transcendental for all nonzero α in $S \cap \Omega$. Indeed, suppose that $\varphi(\alpha) \in \Omega$. Then we have an equation
$$(e^{\beta z})^n P_0\big(\varphi(z)\big) + (e^{\beta z})^{n-1} P_1\big(\varphi(z)\big) + \cdots + P_n\big(\varphi(z)\big) = 0$$
for all $z \in S$ with $P_0, \ldots, P_n \in \Omega[t]$ relatively prime. As $\varphi(\alpha) \in \Omega$, we must have $P_i\big(\varphi(\alpha)\big) \neq 0$ for at least one i as the P_i are relatively prime in $\Omega[t]$. Setting $z = \alpha$, we have that $e^{\beta\alpha}$ lies in Ω. But either $\beta\alpha$ or $e^{\beta\alpha}$ is transcendental, a contradiction.

In particular, setting $\beta = \sqrt{-1}$, we have $\cos\alpha$ is transcendental over $\mathbb{Q}$ for all nonzero $\alpha \in \Omega$, as $(e^{\sqrt{-1}\,x})^2 + 1 - 2e^{\sqrt{-1}\,x}\cos x = 0$. for all $x \in \mathbb{R}$ with $\varphi(z) = \cos z$. Similarly, $\sin\alpha$ is transcendental for all nonzero $\alpha \in \Omega$.

Theorem 74.27. *Let β be a non-rational algebraic integer, $K = \mathbb{Q}(\beta)$, $f_1(z) = e^{\beta z}$, $f_2(z) = e^z$. Suppose that $R_1 = t_1 t_1'$ and $R_2 = t_2 t_2'$ and $e^{\beta z}$ and e^z are algebraically independent. Then at least one of $e^{\beta\alpha}, e^{\alpha}$ is transcendental over $\mathbb{Q}$ for all nonzero $\alpha \in \mathbb{C}$.*

PROOF. The proof is similar to the proof of Corollary 74.25. Suppose that $e^{\beta t}, e^t$ are algebraically dependent. Then we have an equation
$$(e^{\beta z})^n P_0(e^z) + (e^{\beta z})^{n-1} P_1(e^z) + \cdots P_n(e^z) = 0$$
with $P_i \in K[t]$ and P_0 not zero. Let $\gamma \in \mathbb{C}$ satisfy $P_0(e^{\beta\gamma}) \neq 0$. Since β is not rational, the numbers $e^{\beta(\gamma + 2k\pi\sqrt{-1})} = e^{\beta\gamma} e^{2k\beta\pi\sqrt{-1}}$, $k = 1, 2, \ldots$ are all distinct. Then $z_k = \gamma + 2k\pi\sqrt{-1}$ are infinitely many roots of
$$t^n P_0(e^{\gamma}) + t^{n-1} P_1(e^{\gamma}) + \cdots P_n(e^{\gamma}) = 0$$
which is impossible. The result follows by Theorem 74.24. $\square$

Example 74.28. Applying the theorem to $e^z, e^{\sqrt{-1}\,z}$, we see that e^{π} is transcendental. Similarly, $e^{-\pi}$ is transcendental.

The major application of the Theorem 74.27 is the solution of Hilbert's seventh problem.

Theorem 74.29. (Gelfond-Schneider) *Let α, β be algebraic numbers with $\alpha \neq 0, 1$ and β not rational, then α^{β} is transcendental over $\mathbb{Q}$.*

PROOF. By the theorem, if $\alpha \neq 0$ and β is algebraic and not rational, at least one of $e^{\beta\gamma}$, e^{γ} must be transcendental over $\mathbb{Q}$. Setting $\alpha = e^{\gamma}$ gives the result. $\square$

We have shown that if $\alpha \neq 0, 1$ and α^{β} are both algebraic, then $\beta = \log(\alpha^{\beta})/\log\alpha)$ must be a rational number or transcendental over $\mathbb{Q}$. So the Gelfond-Schneider Theorem says:

Corollary 74.30. *If α and γ are both nonzero algebraic numbers and $\alpha \neq 1$, then $\beta = \log\gamma/\log\alpha$ must be rational or transcendental over $\mathbb{Q}$.*

It also follows that

Corollary 74.31. *Let $\beta \in \mathbb{C}$ be irrational and $\alpha \in \mathbb{C}$ nonzero. Then at least one of $\beta, e^{\alpha}, e^{\beta\alpha}$ is transcendental.*

In fact, these two corollaries are equivalent to the Gelfond-Schneider Theorem. The most famous examples of the Gelfond-Schneider theorem are $2^{\sqrt{2}}$ and $\sqrt{2}^{\sqrt{2}}$. Note also that $e^{\pi} = (-1)^{\sqrt{-1}}$ and $e^{-\frac{\pi}{2}} = \sqrt{-1}^{\sqrt{-1}}$.

Remark 74.32. Theorem 74.24 is applicable to the analogue for the tangent, hyperbolic sine, hyperbolic cosine, and hyperbolic tangent functions. We leave these as exercises. Theorem 74.24 is also applicable to the Jacobi elliptic sine, cosine, and delta amplitude which are meromorphic in the whole complex plane and of order $\rho \leq 3$.

Using Using these methods, Lang proved (which we will not) the following theorem.

Theorem 74.33. (The Six Exponentials Theorem) *Let γ_1, γ_2, and γ_3 be linearly independent complex numbers over the rationals and that β_1 and β_2 are linearly independent complex numbers over the rationals. Then at least one of the numbers $e^{\gamma_i \beta_j}$, $i = 1, 2, 3$ and $j = 1, 2$, is transcendental.*

Exercise 74.34. 1. If P is a polynomial of degree d, determine the order of the function $e^{P(z)}$.

2. Prove that if f is an entire function of order $\leq \rho$, then its derivative f' has order $\leq \rho$.
3. Show that $\tan\alpha$ is transcendental over $\mathbb{Q}$ if α is a nonzero algebraic number.
4. Show that $\sinh\alpha$, $\cosh\alpha$, and $\tanh\alpha$ are transcendental over $\mathbb{Q}$ if α is a nonzero algebraic number.
5. Show that the statement if α and β are nonzero algebraic numbers with $\log\alpha$ and $\log\beta$ linearly independent over $\mathbb{Q}$, then they are linearly independent over Ω is equivalent to the Gelfond-Schneider Theorem.
6. Show that the statement if α and β are algebraic numbers linearly independent over $\mathbb{Q}$, then for any $0 \neq x \in \mathbb{C}$, at least one of $e^{x\alpha}$ and $e^{x\beta}$ is transcendental over $\mathbb{Q}$. is equivalent to the Gelfond-Schneider Theorem.

CHAPTER XIV

The Theory of Formally Real Fields

In this chapter, we develop the theory of formally real fields including the standard results from Artin-Schreier Theory. In particular, we determine the general form of the Fundamental Theorem of Algebra and solve Hilbert's seventeenth problem.

75. Orderings

Definition 75.1. A field F is called *formally real* if -1 is not a sum of squares in F, i.e., the polynomial $t_1^2 + \cdots + t_n^2$ has no nontrivial zero over F for any (positive) integer n. A formally real field is called *real closed* if it has no proper algebraic extension that is formally real.

If F is formally real, then the characteristic of F must be zero for if $\operatorname{char} F = p > 0$, then -1 is a sum of $p - 1$ squares.

Notation 75.2. For a commutative ring R, let

$$\sum R^2 := \{x \in R \mid x \text{ is a sum of squares in } R\}.$$

A field F of characteristic different from two is not formally real if and only if $F = \sum F^2$. In general, $\sum F^2$ is closed under addition and multiplication while $\sum (F^2)^\times := \sum (F^2) \setminus \{0\}$ is a multiplicative group.

Definition 75.3. Let R be a commutative ring, $P \subset R$ a subset. We say that P is a *preordering* of R if P satisfies all of the following:

(1) $P + P \subset P$.
(2) $P \cdot P \subset P$.
(3) $-1 \notin P$.
(4) $\sum R^2 \subset P$.

Let

$$\mathcal{Y}(R) := \{P \mid P \subset R \text{ is a preordering }\}.$$

[Of course, $\mathcal{Y}(R)$ may very well be empty.] A preordering $P \in \mathcal{Y}(R)$ is called an *ordering* if, in addition,

$$R = P \cup -P \text{ and } P \cap -P = \{0\}$$

where $-P := \{x \mid -x \in P\}$.

Let P be a preordering on R. We shall show below that if P is a maximal preordering (rel $\subset$), then $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field, this means that P is a maximal preordering if and only if P is an ordering. Note: We have not excluded 0 from lying in P.

Let

$$\mathcal{X}(R) := \{P \in \mathcal{Y}(R) \mid P \text{ is an ordering of } R\}.$$

By definition, a field F is formally real if and only if $\sum F^2$ is a preordering. In general, the preordering $\sum F^2$ in a formally real field F is not an ordering. We shall see that $\sum F^2$ is an ordering if and only if $|\mathcal{X}(F)| = 1$. For example, this is the case when a formally real field F is *euclidean*, i.e., if every element in F is a square or the negative of a square. For such a field, we even have $\sum F^2 = F^2$. For example, the real numbers or the real constructible numbers are euclidean fields. Of course, in general $F^2 \neq \sum F^2$. When a field of characteristic different from two satisfies $F^2 = \sum F^2$, it is called *pythagorean*. If F is not formally real (and $\operatorname{char} F \neq 2$), then F is pythagorean if and only if it is *quadratically closed*, i.e., $F = F^2$.

Lemma 75.4. *Let R be a commutative ring, $P \in \mathcal{Y}(R)$.*

(1) *If $a, b \in R$ satisfy $ab \in P$ then either $P + aP \in \mathcal{Y}(R)$ or $P - bP \in \mathcal{Y}(R)$.*
(2) *If $P \in \mathcal{Y}(R)$ is maximal (relative to $\subset$), then $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field then $P \in \mathcal{X}(R)$.*
(3) *If R is a field, then P is a maximal preordering if and only if $P \in \mathcal{X}(R)$.*
(4) *There exists a $\widetilde{P} \in \mathcal{Y}(R)$ containing P and satisfying $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field, there exists an ordering $\widetilde{P} \in \mathcal{X}(R)$ containing P.*

Proof. (1): We need only show that -1 cannot lie in both $P + aP$ and $P - bP$. If it does then there exist $x, y, z, w \in P$ satisfying $-1 = x + ay = z - bw$, so

$$(ay)(-bw) = (-1 - x)(-1 - z) = 1 + x + z + xz,$$

hence $-1 = x + z + xz + abyw \in P$, a contradiction.

(2): If $a \in R$ then $a^2 \in P$, so by (1), either $P + aP \in \mathcal{Y}(R)$ or $P - aP \in \mathcal{Y}(R)$. By maximality, either $a \in P$ or $-a \in P$, so $R = P \cup -P$. Certainly, $P \cap -P$ is closed under addition. If $x \in P \cap -P$ and $y \in R$ then $xy = (-x)(-y) \in P \cap -P$, so $P \cap -P$ is an ideal. Suppose that $ab = (-a)(-b) \in P \cap -P$. If $a \notin P \cap -P$, we may assume that $a \in -P \setminus P$. Maximality and (1) imply that $-b \in P$. As $a(-b) \in P \cap -P$ also, the same argument shows that $b \in P$.

(3): If $P \in \mathcal{X}(R)$, then $[R^\times : P \setminus \{0\}] = 2$, so P must be a maximal preordering.

(4) follows from (2) and Zorn's Lemma. □

Proposition 75.5. *Let F be a formally real field and $P \in \mathcal{Y}(F)$. Then*

$$P = \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}.$$

Proof. The inclusion $P \subset \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}$ is clear. Conversely, suppose that $x \notin P$. By definition, $-1 \notin P$. We show that $P - xP \in \mathcal{Y}(F)$. To do this, it suffices to show that $-1 \notin P - xP$. If this is false, write $-1 = y - xz$ with $y, z \in P$. Then $z \neq 0$ so

$$x = 1 \cdot z^{-1} + yz^{-1} = \frac{z}{z^2} + \frac{yz}{z^2} \in P,$$

a contradiction. By the lemma, there exists $\widehat{P} \in \mathcal{X}(F)$ such that $P - xP \subset \widehat{P}$. As $x \notin P - xP$, lest $-x^2 \in P - xP$, it follows that $-x \in \widehat{P}$, so $x \notin \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}$. $\square$

Corollary 75.6. (Artin-Schreier) *Suppose that F is formally real. Then $\sum F^2 = \bigcap_{\mathcal{X}(F)} P$. In particular, $\mathcal{X}(F) \neq \emptyset$.*

PROOF. As F is formally real, $\sum F^2 \in \mathcal{Y}(F)$. $\square$

Remark 75.7. Let F be a formally real field and $P \in \mathcal{X}(F)$. An element $0 \neq x \in F$ is called *positive* (respectively, *negative*) *rel* P and also written $x >_P 0$ (respectively, $x <_P 0$) if $x \in P$ (respectively, $-x \in P$). If $x \in \bigcap_{\mathcal{X}(F)} Q$, i.e., x is positive at every ordering of F then x is called *totally positive* (and $-x$ is called *totally negative*). In particular, we have

$$\bigcap_{\mathcal{X}(F)} Q \setminus \{0\} = (\sum F^2)^{\times},$$

i.e., $0 \neq x$ is totally positive if and only if it is a sum of squares.

If $P \in \mathcal{X}(F)$, let $\leq_P$ on F denote the total ordering given by $x \leq_P y$ if $y - x \in P$. Of course, if $x \leq_P y$ and $z \in F$ then $x + z \leq_p y + z$ and if, in addition, $0 \leq_P z$ then $xz \leq_P yz$. We let $>_P$ and $\geq_P$ have the obvious meaning.

Exercises 75.8.

1. Prove that the field of real constructible numbers C is euclidean and $C(\sqrt{-1})$ is quadratically closed.
2. Show that a formally real field is euclidean if and only if it is pythagorean and has a unique ordering.
3. Show the intersection of pythagorean fields is pythagorean.
4. Let F be a formally real field and $\widetilde{F}$ an algebraic closure of F. A pythagorean field L is called a *pythagorean closure* of F if either $L = F$ or whenever $F \subset K < L$, then K is not pythagorean. Show that if K is a field in $\widetilde{F}$ containing F and is the intersection of all pythagorean fields containing F in $\widetilde{F}$, then K is pythagorean. In particular, a pythagorean closure of F in $\widetilde{F}$ exists and is unique. It is denoted by F_{pyth}.
5. Let F be a formally real field and $\widetilde{F}$ an algebraic closure. Call a square root tower $F \subset F_0 \subset F_1 \subset \cdots \subset F_n$ admissible if for each i, $1 \leq i \leq n-1$, there exist x_i, y_i in F_{i-1} satisfying $F_i = F_{i-1}(\sqrt{x_i^2 + y_i^2})$. Show the union of all admissible square towers over F is the pythagorean closure F_{pyth} of F. In particular, if F is a formally real field, then the pythagorean closure of F in $\widetilde{F}$ exists.

76. Extensions of Ordered Fields

Definition 76.1. Let K/F be a field extension with K formally real. Let $P \in \mathcal{X}(K)$. The pair (K, P) is called an *ordered field.* If $P_0 \in \mathcal{X}(F)$ satisfies $P_0 = P \cap F$, then $(K, P)/(F, P_0)$ is called an *extension* of ordered fields and P is called an *extension* of P_0. If there exists no proper algebraic extension L of F with $Q \in \mathcal{X}(L)$ satisfying $Q \cap F = P_0$, we say that (F, P_0) is *real closed (rel P_0).* Let $(K, P)/(F, P_0)$ be an extension of ordered fields. If (K, P) real closed (rel P) and K/F is algebraic, (K, P) is called a *real closure of* (F, P_0) (*rel* P_0).

We shall show that given $P \in \mathcal{X}(F)$, there exists a real closure of (F, P) and it is unique up to a (unique) F-isomorphism. We begin with the following useful result characterizing extensions of orderings under certain conditions.

Theorem 76.2. *Let (F, P) be an ordered field.*

(1) Let $d \in F$ and $K = F(\sqrt{d})$. Then there exists an extension of P to K if and only if $d \in P$.
(2) If K/F is finite of odd degree then there exists an extension of P to K.

Proof. (1): Suppose $d \in P$. We may assume that $d \in P \setminus F^2$. Let

$$S = \{\sum x_i y_i^2 \mid x_i \in P,\ y_i \in K \text{ for all } i\}.$$

We show that $S \in \mathcal{Y}(K)$. Certainly, S is closed under addition and multiplication and contains $\sum K^2$. Thus it suffices to show that $-1 \notin S$. If $-1 \in S$, we can write $-1 = \sum x_i(a_i + b_i\sqrt{d})^2$ for some $a_i, b_i \in F$ and $x_i \in P$. As $\{1, \sqrt{d}\}$ is an F-basis for K, we arrive at the contradiction that $-1 = \sum x_i(a_i^2 + b_i^2 d) \in P$. This proves that $S \in \mathcal{Y}(K)$. By Lemma 75.4, there exists $\widetilde{P} \in \mathcal{X}(K)$ satisfying $P \subset S \subset \widetilde{P}$. As P is an ordering, $P = \widetilde{P} \cap F$. Suppose that $d \notin P$. If $\widetilde{P} \in \mathcal{X}(K)$ contains P then $d = (\sqrt{d})^2 \in \widetilde{P} \cap F = P$, a contradiction.

(2): As the char F is zero, by the Primitive Element Theorem, $K = F(x)$, for some $x \in K$. Let f be the minimal polynomial of x and suppose that

$$S = \{\sum a_i y_i^2 \mid a_i \in P,\ y_i \in K \text{ for all } i\} \notin \mathcal{Y}(K).$$

As $K \cong F[t]/(f)$, we have an equation,

$$-1 \equiv \sum a_i g_i(t)^2 \mod (f)$$

for some $a_i \in P$ and $0 \neq g_i \in F[t]$ with $\deg g_i < \deg f$ for all i. Lift this to an equation

$$-1 = \sum a_i g_i(t)^2 + qf$$

in $F[t]$. Since $-1 \notin P$, evaluating at x shows that there exists an i with $\deg g_i > 0$. Let $m = \max_i(\deg g_i) > 0$. Then the coefficient of t^{2m} in $\sum a_i g_i(t)^2$ cannot be zero as it is of the form $\sum a_m b^{2m} >_P 0$. Thus $\deg qf = 2m$. As $\deg f = n$ is odd, so is $\deg q$. By construction, $\deg q < \deg f$. Let $h \mid q$ in $F[t]$ with $h \in F[t]$ irreducible of odd degree. Then $-1 \equiv \sum a_i g_i^2 \mod (h)$, so we may repeat the argument. By induction, this reduces to the case when f is linear where the result is false. □

The first part of the theorem has the immediate corollary.

Corollary 76.3. *Let (F, P_0) be an ordered field. Suppose that there exists an element d in $P_0 \setminus F^2$. Then there exists an proper extension of ordered fields $(F(\sqrt{d}), P)/(F, P_0)$ with $P_0 + dP_0 \subset P$. In particular $F(\sqrt{d})$ is formally real.*

We also conclude that

Corollary 76.4. *(F, P) is real closed (rel P) if and only if F is real closed. Moreover, if this is the case, then $P = F^2$.*

PROOF. Suppose that (F, P) is real closed. Let $d \in P$. By the theorem, P extends to $F(\sqrt{d})$ so $d \in F^2$ by hypothesis. As (F, P) is real closed, we must have $P = F^2 \in \mathcal{X}(F)$. Since $F^2 \subset Q$ for all $Q \in \mathcal{X}(F)$, we have $\mathcal{X}(F) = \{F^2\}$. Suppose that L/F is algebraic and L formally real. Let $Q \in \mathcal{X}(L)$. Then $(L, Q)/(F, F^2)$ is an extension of ordered fields as $F^2 \subset Q$, hence $L = F$ and F must be real closed.

If F is real closed, then $F(\sqrt{d})/F$ cannot be a nontrivial extension of formally real fields. It follows by the Corollary 76.3 and Lemma 75.4 that $F^2 \in \mathcal{Y}(F)$ is an ordering hence $\mathcal{X}(F) = \{F^2\}$. If L/F is a proper algebraic extension, then $-1 \in L^2$ so $F \cap L^2$ cannot be a preordering on F. It follows that (F, F^2) is real closed. $\square$

There exist fields such that $\mathcal{X}(F) = \{F^2\}$ but F not real closed, i.e., euclidean fields that are not real closed. The real constructible numbers is such an example. The theorem only implies that a euclidean field has no formally real quadratic extensions. It also means that $F(\sqrt{-1})$ is quadratically closed. Indeed if $\alpha = a + b\sqrt{-1}$ in $F(\sqrt{-1})$ with $a, b \in F$, then there exist $c, d \in F$ satisfying

$$c^2 = \frac{a + \Delta}{2} \quad \text{and} \quad d^2 = \frac{-a + \Delta}{2}$$

where $\Delta = \sqrt{a^2 + b^2}$ is the positive square root in F^2 in euclidean F.

Proposition 76.5. *Every ordered field (F, P) has a real closure $(\widetilde{F}, \widetilde{F}^2)$.*

PROOF. Let $\widehat{F}$ be an algebraic closure of F. Let $P \in \mathcal{X}(F)$. The statement follows from a Zorn's Lemma argument on

$$\{\widehat{F}/K/F \mid (K, Q)/(F, P) \text{ is an extension}\}. \qquad \square$$

Because of the last results, if $(K, P)/(F, P_0)$ is a real closure, then $P = K^2$ and K is real closed so we just say K is a real closure of F rel P.

Corollary 76.6. *Let F be a formally real field and $P \in \mathcal{X}(F)$. Then there exists a real closure of (F, P).*

Theorem 76.7. (Fundamental Theorem of Algebra) *The following are equivalent:*

(1) F is real closed.
(2) F is euclidean and every polynomial $f \in F[t]$ of odd degree has a root in F.
(3) F is not algebraically closed but $F(\sqrt{-1})$ is.

PROOF. (1) $\Rightarrow$ (2): We have seen the hypothesis implies that F is euclidean. Let $p \in F[t]$ be irreducible of odd degree. As $K = F[t]/(p)$ is an extension of odd degree any ordering of F would extend, hence p is linear. (2) easily follows.

(2) $\Rightarrow$ (3): This is the essentially the same proof given for the usual Fundamental Theorem of Algebra 57.12. Let $K = F(\sqrt{-1})$. Then K is quadratically closed. Let E/K be an algebraic extension and L/F the normal closure of K/F. If H is a 2-Sylow subgroup of the Galois group G of L/F, then $L^H = F(\theta)$ for some θ by the Primitive Element Theorem, so we must have $L^H = F$ by hypothesis. Consequently, $G = H$ and hence must be trivial as K is quadratically closed.

(3) $\Rightarrow$ (1): It suffices to show that F is formally real, i.e., $-1 \notin \sum F^2$. As $\sqrt{-1} \notin F$, it suffices to show F is pythagorean. Let $a, b \in F^\times$. We need only show that $a^2 + b^2 \in F^2$.

Let $f = t^4 - 2at^2 + (a^2 + b^2) \in F[t]$. Then $f = \big(t^2 - (a + b\sqrt{-1})\big)\big(t^2 - \big(a - b\sqrt{-1}\big)\big)$ in $F(\sqrt{-1})$. As $F(\sqrt{-1})$ is quadratically closed, there exist $\alpha, \beta \in F(\sqrt{-1})$ satisfying

$$\alpha^2 = a + b\sqrt{-1} \quad \text{and} \quad \beta^2 = a - b\sqrt{-1},$$

hence

$$f = (t - \alpha)(t + \alpha)(t - \beta)(t + \beta) \quad \text{in} \quad F(\sqrt{-1})[t].$$

As $f \in F[t]$ cannot be irreducible and $ab \neq 0$, either $(t - \alpha)(t - \beta)$ or $(t - \alpha)(t + \beta)$ is an irreducible factor of f over F. Thus $\alpha\beta \in F$, hence

$$a^2 + b^2 = (a + b\sqrt{-1})(a - b\sqrt{-1}) = \alpha^2\beta^2$$

lies in F^2. □

It follows that over a real closed field, the only irreducible polynomials are linear polynomials and irreducible quadratic polynomials of the form $at^2 + bt + c$ with negative discriminant (in the unique ordering).

Artin and Schreier proved that the conditions of the Fundamental Theorem of Algebra were also equivalent to: F is not algebraically closed and its algebraic closure is a finite extension of it. We shall come back to this below. Before showing this, we turn to the question of uniqueness.

We first need some facts and notation about symmetric bilinear forms over a field of characteristic different from two. Details are given Appendix E.

Remark 76.8. Let V be a finite dimensional vector space over a field F of characteristic different from two and $B : V \times V \to F$ a symmetric bilinear form, i.e., linear in each variable and $B(x, y) = B(y, x)$ for all $x, y \in V$. We call (V, B) a *bilinear space*. We say x, y in V are *orthogonal* (rel B) if $B(x, y) = 0$. We call subspaces W_1, W_2 of V *orthogonal* if $B(x, y) = 0$ for all $x \in W_1$, $y \in W_2$ and then denote $W_2 + W_2$ by $W_1 \perp W_2$. Every (V, B) has an *orthogonal basis* $\mathcal{C}$, i.e., $V = \perp_{v \in \mathcal{C}} Fv$. If $\mathcal{B} = \{v_1, \dots v_n\}$ is an ordered basis for V, then we write $[B]_{\mathcal{B}}$ for the *matrix representation* $\big((B(v_i, v_j)\big)$ of B. So $\mathcal{B}$ an orthogonal basis if and only if $[B]_{\mathcal{B}}$ is diagonal. If $\mathcal{C}$ is another basis, then $[1_V]^t_{\mathcal{B},\mathcal{C}}[B]_{\mathcal{B}}[1_V]_{\mathcal{B},\mathcal{C}} = [B]_{\mathcal{C}}$. We also have $B(x+y, x+y) = B(x, x) + 2B(x, y) + B(y, y)$, so B is determined by $B(x, x)$, $x \in V$ (or v in a basis for V).

Let A be a finite dimensional commutative F-algebra. If $a \in A$ let $\lambda_a : A \to A$ be the F-linear map given by $x \mapsto ax$. The *trace* trace : $A \to F$ is the trace of a matrix representation of λ_a relative to a fixed basis. It is independent of the choice of basis. If A is a finite separable field extension of F of characteristic different from two, then this is just $\mathrm{Tr}_{A/F}$ (cf. Exercise 60.26(6).) Define the *trace form* of A

$$\varphi = \varphi_A : A \times A \to F$$

to be the symmetric bilinear form given by $\varphi(x, y) = \mathrm{trace}(\lambda_{xy})$.

Let B be a symmetric bilinear form over a formally real field F and $P \in \mathcal{X}(F)$. As an ordering on F is induced by the real closure relative to this ordering, the results of Appendix E hold. In particular, we have the following: [Cf. Proposition E.12 and Corollary E.16 in Appendix E.]

If V is the underlying vector space for B, then V has an *orthogonal decomposition* $V = V_+ \oplus V_- \oplus V_0$ where if $0 \neq v \in V$ then

$$B(v,v) \;=\; \begin{cases} >_P 0 & \text{if } v \in V_+ \\ <_P 0 & \text{if } v \in V_- \\ = 0 & \text{if } v \in V_0 \end{cases}$$

with the dimensions of $V_0,\ V_+$, and V_- independent of such an orthogonal decomposition. The integer

$$\operatorname{sgn}_P B \;:=\; \dim V_+ \,-\, \dim V_-$$

is an invariant of B, called the *signature* of B at P, so $\operatorname{sgn}_P B = \operatorname{sgn}_P[B]_{\mathcal{B}}$ for any basis $\mathcal{B}$ of V.

If K is an extension field of F and $\mathcal{B}$ a basis for V, then we let V_K be the vector space over K on basis $\mathcal{B}$. By restricting scalars to F, we see it is also a vector space over F and there exists a (natural) linear embedding of V into V_K by sending $\mathcal{B}$ to $\mathcal{B}$. We call this *extension of scalars*. [Cf. making a real vector space into a complex one by extending scalars.] If B is the bilinear form above, define $B_K : V_K \times V_K \to K$ to be the map additive in each variable satisfying $B_K(\alpha v, \beta w) = \alpha\beta B_K(v,w)$ for all α and β in K and $v, w \in \mathcal{B}$. Then, if $(K,Q)/(F,P)$ is an extension, $\operatorname{sgn}_P B = \operatorname{sgn}_Q(B_K)$.

Computation 76.9. Let (F,P) be an ordered field, $f \in F[t] \setminus F$ and $A = F[t]/(f)$. Let $^- : F[t] \to A$ be the canonical epimorphism and $\varphi : A \times A \to F$ the trace form.

Case 1. $f = (t-a)^n$ in $F[t]$:

Let $\mathcal{B} = \{1, \bar{t}-a, \dots, (\bar{t}-a)^{n-1}\}$, an F-basis for A. As $(\bar{t}-a)$ is nilpotent, $\operatorname{trace}(\lambda_{(\bar{t}-a)^i}) = 0$ for all $i \geq 1$. It follows that the matrix representation of φ in the $\mathcal{B}$ basis is given by the diagonal matrix

$$[\varphi]_{\mathcal{B}} \quad=\quad \big(\operatorname{trace}(\lambda_{(\bar{t}-a)^i(\bar{t}-a)^j})\big) \quad=\quad \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

Consequently, $\operatorname{sgn}_P \varphi = \operatorname{sgn}[\varphi]_{\mathcal{B}}) = 1$.

Case 2. $f = (t^2 + at + b)^n$ in $F[t]$ with $a^2 - 4b <_P 0$:

Let $\mathcal{B} = \{1, \bar{t}, \bar{t}^2 + 1, \bar{t}(\bar{t}^2+1), (\bar{t}^2+1)^2, \bar{t}(\bar{t}^2+1)^2, \dots, (\bar{t}^2+1)^n, \bar{t}(\bar{t}^2+1)^n\}$, an F-basis for A. Computation shows the matrix representation of φ in the $\mathcal{B}$ basis is given by the diagonal matrix

$$[\varphi]_{\mathcal{B}} \quad=\quad \begin{pmatrix} 2n & 0 & \cdots & & 0 \\ 0 & -2n & & & \\ \vdots & & 0 & & \\ & & & \ddots & \\ 0 & \cdots & & & 0 \end{pmatrix}.$$

Consequently, $\operatorname{sgn}_P \varphi = 0$

The next result, due to Sylvester, is the main ingredient in proving the uniqueness of real closures relative to a fixed ordering. It replaces the more common use of Sturm's Theorem.

Lemma 76.10. (Sylvester) *Let (F, P) be an ordered field and K a real closure of (F, P). Suppose that f is a non-constant polynomial in $F[t]$ and $A = F[t]/(f)$. Let $\varphi : A \times A \to F$ be the trace form. Then*

$$\operatorname{sgn}_P \varphi = \text{ the number of roots of } f \text{ in } K.$$

In particular, the number of roots of f in K depends only on P and is independent of the real closure K.

Proof. Let φ be the trace form on $A = F[t]/(f)$ and $f = up_1^{e_1} \cdots p_r^{e_r}$ a factorization of f into monic irreducibles in $K[t]$ with $u \in K^\times$. By the Chinese Remainder Theorem, we have a natural ring (and K-algebra) isomorphism.

$$A_K := K[t]/(f) \cong \prod K[t]/(p_i^{e_i}) = \prod A_i,$$

where $A_i = K[t]/(p_i^{e_i})$. We view this as an identification of K-algebras. It is easy to check that $\varphi_{A_K}(x_i, x_j) = 0$ if $x_i \in A_i$ and $x_j \in A_j$ with $i \neq j$ (by writing down an appropriate K-basis), hence we can restrict φ_A to each A_i to obtain $\operatorname{sgn}_P \varphi = \operatorname{sgn}_{K^2}(\varphi_{A_K}) = \sum_{i=1}^r \operatorname{sgn}_{K^2} \varphi_{A_i}$. Since K is real closed, it follows from the Fundamental Theorem of Algebra that p_i is either linear or quadratic. Therefore, by the computation, we see that

$$\operatorname{sgn}_P \varphi = \sum_{i=1}^r \operatorname{sgn}_{K^2} \varphi_{A_i} = \sum_{p_{i_j} \text{ linear}} e_{i_j}.$$

The result follows. $\square$

Lemma 76.11. *Let (F, P) be an ordered field and K a real closure of (F, P). Suppose that $E = F(\alpha)$ is an algebraic extension of F with $\varphi = \varphi_E : E \times E \to F$ the trace form. Suppose that $r = \operatorname{sgn}_P \varphi > 0$. Then the minimal polynomial f of α has r roots in K. Let $\alpha_1, \alpha_2, \ldots, \alpha_r$ be these roots. Then there exist at most r distinct and at least one extension $\widetilde{P}_i$ of P to E. These are induced by the r distinct F-homomorphisms $\sigma_i : E \to K$ given by $\alpha \mapsto \alpha_i$ with $\widetilde{P}_i = \sigma_i^{-1}(K^2)$.*

Proof. Every F-embedding of E into an algebraic closure of K must take α to a root of f. By Lemma 76.10, the polynomial f has r distinct roots in K. Clearly, $\widetilde{P}_i = \sigma_i^{-1}(K^2) \supset P$. Suppose that $Q \in \mathcal{X}(E)$ extends P with $Q \neq \widetilde{P}_i$ for $1 \leq i \leq r$. Then there exists an $a_i \in E$ satisfying $a_i \in Q$ but $\sigma_i(a_i) \notin K^2$ for all i. By repeated application of Theorem 76.2, there exists $\widetilde{Q}_i \in \mathcal{X}\big(E(\sqrt{a_1}, \ldots, \sqrt{a_r})\big)$ extending Q. By the Primitive Element Theorem, there exists β such that $F(\beta) = E(\sqrt{a_1}, \ldots, \sqrt{a_r})$. Let L be a real closure of $(F(\beta), \widetilde{Q})$ hence of (F, P). The minimal polynomial $g \in F[t]$ of β has root $\beta \in L$, hence a root in K by Lemma 76.10. Therefore, there exists an F-homomorphism $\tau : F(\beta) \to K$. As the F-homomorphism $\tau|_E : E \to K$ must satisfy $\tau|_E = \sigma_i$ for some i, we have $\sigma_i(a_i) = \tau(a_i) = \big(\tau(\sqrt{a_i})\big)^2$ lies in K^2, a contradiction. $\square$

Theorem 76.12. *Let (F, P) be an ordered field.*

(1) Let $(E, Q)/(F, P)$ be an algebraic extension of ordered fields and K a real closure of (F, P). Then there exists an order preserving F-homomorphism $E \to K$.
(2) Let K_1 and K_2 be two real closures of (F, P). Then there exists a unique F-homomorphism $K_1 \to K_2$ and it is an order preserving isomorphism.
(3) In Lemma 76.11 and its notation, all the $\widetilde{P}_i$ are distinct. In particular, there exist precisely r extensions of P to E.

Proof. (1): Apply Zorn's Lemma to

$$\{M \mid E/M/F \text{ and there exists an order preserving } F\text{-homomorphism } \psi_M : (M, Q \cap M) \to (K, K^2)\}$$

to obtain a maximal such M_0. Suppose $M_0 \neq E$ and $x \in E \setminus M_0$. Let L be a real closure of (E, Q). Then f has a root in L so $\operatorname{sgn}_{Q \cap M_0(x)} \varphi_{M_0(x)} > 0$. By Lemma 76.11, we can extend $\psi_{M_0} : (M_0, Q \cap M_0) \to (K, K^2)$ to $\psi_{M_0(x)} : \big(M_0(x), Q \cap M_0(x)\big) \to (K, K^2)$, a contradiction. So $M_0 = E$.

(2): By (1) there exists an F-homomorphism $\sigma : K_1 \to K_2$. Let $\tau : K_1 \to K_2$ be another F-homomorphism. Let $\alpha \in K_1$ and $\alpha_1, \ldots, \alpha_r$ be the (distinct) roots of the minimal polynomial f of α in K_1. We may assume that

$$\alpha_1 <_{K_1^2} \cdots <_{K_1^2} \alpha_r$$

in K_1. As $\sigma(K_1^2) \subset K_2^2$ and $\tau(K_1^2) \subset K_2^2$ and $\mathcal{X}(K_i) = \{K_i^2\}$ for $i = 1, 2$ both σ and τ are order preserving. Thus

$$\sigma(\alpha_1) <_{K_2^2} \cdots <_{K_2^2} \sigma(\alpha_r) \quad \text{and} \quad \tau(\alpha_1) <_{K_2^2} \cdots <_{K_2^2} \tau(\alpha_r) \quad \text{in} \quad K_2.$$

As these are the roots of f in K_2, we have $\sigma(\alpha_i) = \tau(\alpha_i)$ for all i. In particular, $\sigma(\alpha) = \tau(\alpha)$. Hence $\sigma = \tau$.

(3): Suppose in Lemma 76.11 that we have $\widetilde{P}_i = \widetilde{P}_j$ with $i \neq j$. Let $\widetilde{P} = \widetilde{P}_i$ and M a real closure of $(E, \widetilde{P})$. By (1), the embeddings $\sigma_i \neq \sigma_j$ can be extended to F-homomorphisms $M \to K$. This contradicts (2). $\square$

Exercise 76.13. Show that the field of real constructible numbers C has algebraic extensions that are not euclidean. In particular, $C(\sqrt{-1})$ is not algebraically closed.

77. Characterization of Real Closed Fields

We turn to the proof of the generalization of the Fundamental Theorem of Algebra for fields. [We shall prove a version of the usual statement of the Fundamental Theorem of Algebra for division rings in Theorem 108.12.] The proof is rather intricate in order to take care of the positive characteristic case. The proof revolves around the study of primitive p^rth roots of unity for a prime p. This is necessary in order to reduce to $\sqrt{-1}$, a primitive fourth root of unity in characteristic zero.

To prove the last Artin-Schreier characterization of real closed fields, we need a few facts from field theory. These are

Facts 77.1. *Let F be a field.*

(1) *Let p be a prime. Then $\mathbb{Z}/p^r\mathbb{Z}$ is cyclic unless $p = 2$ in which case $(\mathbb{Z}/2^r\mathbb{Z})^\times$ is not cyclic if $r \geq 3$. (Cf. Section F.)*
(2) *If $\operatorname{char} F = p > 0$ and $a \in F \setminus F^p$, then $t^{p^e} - a \in F[t]$ is irreducible for all $e \geq 1$. (Cf. Example 50.15(8.))*
(3) *If $\operatorname{char} F = p > 0$ and $a \in F$ satisfies $a \neq u^p - u$ for any $u \in F$, then $t^p - t - a \in F[t]$ is irreducible. (Cf. Exercise 56.22(10).)*
(4) *If $\operatorname{char} F = p > 0$ and $a \in F$ satisfies $a \neq u^p - u$ for any $u \in F$ and E is a splitting field of $t^p - t - a$ over F, then there exists an extension K/E of degree p.*
(5) *Let p be a prime different from $\operatorname{char} F$. If $t^p - 1$ splits in $F[t]$ and K/F is a cyclic extension of degree p, then $K = F(r)$ for some r satisfying $r^p \in F$. (Cf. Theorem 60.20.)*
(6) *Suppose p is a prime different from $\operatorname{char} F$ and F contains a primitive pth root of unity. If E/F is cyclic of degree p, then there exists an $x \in E$ such that $E = F(x)$ and the minimal polynomial of x over F is $t^p - a \in F[t]$. (Same as the previous.)*
(7) *If $\operatorname{char} F = p > 0$ and E/F is cyclic of degree p, then there exists an $x \in E$ such that $E = F(x)$ and $x^p - x \in F$. (Same as previous.)*

We prove Fact 77.1(4), leaving the others as exercises.

PROOF. (of Fact 77.1(4)) Let $E = F(u)$ with $m_F(u) = t^p - t - a$, a cyclic extension of degree p with F-basis $\mathcal{B} = \{1, u, \dots, u^{p-1}\}$. We show that the element au^{p-1} is not of the form $v^p - v$, $v \in E$. If this is not so, say $au^{p-1} = v^p - v$. Write $v = v_0 + v_1 u + \cdots + v_{p-1}u^{p-1}$ for some $v_0, \dots, v_{p-1}$ in E. Since $u^p = u + a$, we have

$$\begin{aligned} au^{p-1} = v^p - v = v_0^p + v_1^p(u+a) + v_2^p(u+a)^2 + \cdots + v_{p-1}^p(u+a)^{p-1} \\ - v_0 - v_1 u - \cdots - v_{p-1}u^{p-1}. \end{aligned}$$

As $\mathcal{B}$ is a basis, we see that $v_{p-1}^p - v_{p-1} = a$, a contradiction. By Fact 77.1(3), we know that $t^p - t - au^{p-1} \in E[t]$ is irreducible, hence its splitting field is of degree p over E. $\square$

Theorem 77.2. (Artin-Schreier) *Let C be algebraically closed and R a proper subfield with C/R finite. Then R is real closed and $C = R(\sqrt{-1})$.*

PROOF. Let $C' = R(\sqrt{-1})$. As C is the algebraic closure of C', it suffices to show $C = C'$. We have:

$$\text{If } C/E/C', \text{ then } [E : C'] \leq [C : C'] < \infty. \tag{77.3}$$

Claim 77.4. C' is perfect:

Suppose not, then $\operatorname{char} F = p > 0$ and there exists an element $a \in C' \setminus (C')^p$. By Fact 77.1(2), for each $n \geq 1$, there exists E_n/C' of degree p^n contradicting (77.3). Therefore C' is perfect.

By Claim 77.4, we conclude that C/C' is finite Galois, since C is algebraically closed. Let $G = G(C/C')$. We may assume that $C' < C$. In particular, there exists a prime number p dividing $|G|$ and a cyclic subgroup $H \subset G$ of order p. Let E be the fixed field C^H, so

$$C/E \text{ is cyclic of degree } p.$$

Claim 77.5. *E contains a primitive pth root of unity and $p \neq \operatorname{char} E$:*

Suppose to the contrary that $p = \operatorname{char} E$. By Fact 77.1(7), there exists an $x \in C$ such that $C = E(x)$ is the splitting field of the irreducible polynomial $t^p - t - a \in E[t]$ over E. Hence $a = x^p - x$. By Fact 77.1(4), there exists a field extension K/C of degree p which is impossible.

Let $\operatorname{char} E = l \geq 0$. We have shown that $p \neq l$. As $t^p - 1 = (t-1)(t^{p-1} + \cdots + 1)$ splits in $C[t]$ and each irreducible factor has degree at most $p-1$, the polynomial $t^p - 1$ must split in $E[t]$, hence E contains a primitive pth root of unity. As $\operatorname{char} E \neq p$, we have $C = E(r)$ where the irreducible polynomial of r is $t^p - a \in E[t]$ by Fact 77.1(5). Let $\zeta \in C$ be a primitive p^2-root of unity. As $t^{p^2} - a$ splits in C, we can write

$$t^{p^2} - a = \prod_{i=0}^{p^2-1} (t - \zeta^i s) \text{ with } a = s^{p^2}$$

in $C[t]$.

Claim 77.6. *There exists a primitive p^2-root of unity ε in C such that $C = E(\varepsilon)$:*

If $\zeta^i s \in E$ for some i, then $(\zeta^i s)^p \in E$. This would mean that $\big((\zeta^i s)^p\big)^p = s^{p^2} = a$ and $t^p - a$ has a root in E contradicting the irreducibility of $t^p - a \in E[t]$. Consequently, $\zeta^i s \notin E$ for all i. This means that every irreducible factor of $t^{p^2} - a \in E[t]$ must have degree p. Let f be such an irreducible factor, say the minimal polynomial of $\zeta^i s$. The constant term of f must be $b = \zeta^n s^p \in E$ for some n. As $s^p \notin E$, we have

$$C = E(s^p) = E(b\zeta^{-n}) = E(\zeta^n).$$

Since E contains a primitive pth root of unity, $\varepsilon = \zeta^n$ must be a primitive p^2-root of unity as needed.

Let Δ be the prime subfield in C.

Claim 77.7. *There exists an r such that $\Delta(\varepsilon)$ contains a primitive p^rth root of unity but not a p^{r+1}th primitive root of unity:*

Let ζ_{p^r} be a primitive p^rth root of unity in C. Suppose that $\Delta \cong \mathbb{Q}$. Then $[\mathbb{Q}(\zeta_{p^r}) : \mathbb{Q}] = p^{r-1}(p-1) \to \infty$ as $r \to \infty$. As $[\mathbb{Q}(\varepsilon) : \mathbb{Q}] < \infty$, Claim 77.7 holds in this case. Suppose that $\Delta \cong \mathbb{Z}/l\mathbb{Z}$. By Claim 77.5, we know that $l \neq p$. As $t^{p^r} - 1$ has no repeated roots in C, we have $|\Delta(\zeta_{p^r})| \geq p^r \to \infty$. As $|\Delta(\varepsilon)| < \infty$, Claim 77.7 also holds if Δ has positive characteristic.

Let r be as in Claim 77.7. As $\varepsilon \in F(\varepsilon)$ and ε is a primitive p^2-root of unity, we have $r \geq 2$. Let ω be a primitive p^{r+1}th root of unity in the algebraically closed field C. Let N be the Galois group of $\Delta(\omega)/\Delta$. If Δ is a finite field, then N is cyclic. If $\operatorname{char} \Delta = 0$, then $N \cong (\mathbb{Z}/p^{r+1}\mathbb{Z})^\times$, which is cyclic unless $p = 2$ and $r \geq 3$ by Fact 77.1(1). Moreover, if N is cyclic, it contains a unique subgroup of order p, hence $\Delta(\omega)$ contains a unique subfield over which it has degree p.

Claim 77.8. *$\Delta(\omega)$ contains two subfields over which it has degree p. In particular, $p = 2$ and $\operatorname{char} E = 0$:*

Let $f \in E[t]$ be the minimal polynomial of ω. Since $\varepsilon \notin E$ also $\omega \notin E$. Hence $C = E(\omega)$ and $\deg f = p$. We also have $f \mid t^{p^{r+1}} - 1$ in $E[t]$ and $t^{p^{r+1}} - 1 = \prod_{i=0}^{p^{r+1}-1}(t - \omega^i)$ in $C[t]$. Let $K = \Delta(\omega) \cap E$. Then by the above, $f \in K[t]$ so f is the minimal polynomial of ω over K. Consequently, $[\Delta(\omega) : K] = p$ and K is one of the desired subfields.

Let $z = \omega^p$, a primitive p^rth root of unity and $K' = \Delta(z)$. As $\Delta(\omega)$ contains a primitive p^rth root of unity and hence all such, we have $K' = \Delta(z) \subset \Delta(\omega)$. We show that $[\Delta(\omega) : \Delta(z)] = [\Delta(\omega) : K'] = p$. As ω is a root of $t^p - z \in K'[t]$, it suffices to show $t^p - z \in K'[t]$ is irreducible. By Fact 77.1(5), it suffices to show that $t^p - z$ has no roots in K'. Since $z \in K'$, certainly K' contains all the primitive pth roots of unity. In particular, if $t^p - z$ has a root in K', it splits in $K'[t]$ and $\omega \in K'$. But this means that $\Delta(\omega) = K' \subset \Delta(\varepsilon)$ contradicting Claim 77.7. Thus $[\Delta(\omega) : K'] = p$.

To prove Claim 77.8, we need only show that $K \neq K'$. Suppose that $K = K'$. Then $z \in K' = K$ means that K contains a primitive p^rth root of unity. As $K = \Delta(\omega) \cap E$, we have $z \in E$, i.e., E contains a primitive p^rth root of unity for some $r \geq 2$, hence it contains a p^2-root of unity. Thus $\varepsilon \in E$ which is a contradiction. We conclude that $K \neq K'$ and Claim 77.8 is established.

Thus we have shown that $\operatorname{char} \Delta = 0$ and $C = E(\varepsilon)$ with ε a primitive 2^2 root of unity, i.e, $\varepsilon = \pm\sqrt{-1}$. But $\sqrt{-1} \in E$ by assumptionThis proves the theorem. $\square$

Let F be a field and C an algebraic closure. Then

$$F_{sep} := \{x \mid x \in C \text{ separable over } x\},$$

is called the *separable closure* of F. It is the maximal separable field extension of F in C. The Galois group $G(F_{sep}/F)$ is called the *absolute Galois group* of F.

Corollary 77.9. *Let F be a field. If the absolute Galois group of F contains a nontrivial element of finite order, then F is formally real and the element is an involution.*

Proof. Let C be an algebraic closure of F and F_{sep} the separable closure of F in C. Let σ lie in $G(F_{sep}/F)$ be of order $r > 1$. Then $G(F_{sep}/F_{sep}^{<\sigma>})$ is finite with $[F_{sep} : F_{sep}^{<\sigma>}] = r$. By Theorem 77.2 the result follows if F is perfect, so we may assume that $\operatorname{char} F = p > 0$. Suppose that $C/K/F_{sep}$ is an intermediate field and $\sigma : K \to C$ an embedding. Then σ induces a unique isomorphism $K[t]/(t^{p^n} - a) \to \sigma(K)[t]/\big(t^{p^n} - \sigma(a)\big)$ taking the unique root of $t^{p^n} - a$ in $K[t]/(t^{p^n} - a)$ in C to the corresponding root of $t^{p^n} - \sigma(a)$ in $\sigma(K)[t]/\big(t^{p^n} - \sigma(a)\big)$ for all n and all $a \in K$. It follows that if $\sigma : F_{sep} \to F_{sep}$ lifts to a unique F-automorphism $\widehat{\sigma}$ of C of order r. Then $C/C^{<\widehat{\sigma}>}$ is a finite extension of order $r > 1$. The result follows by Theorem 77.2. $\square$

Corollary 77.10. *If the absolute Galois group of a field is finite, then it is isomorphic to 1 or $\mathbb{Z}/2\mathbb{Z}$.*

Of course, if F is a field with $\operatorname{char} F = p > 0$ and $\widetilde{F}$ an algebraic closure, it is possible for the absolute Galois group of F to be trivial, but $\widetilde{F}/F$ infinite.

Exercise 77.11. Prove Facts 77.1.

78. Hilbert's 17th Problem

In this section, we present a solution to Hilbert's 17th Problem on positive functions over the reals. We establish the Lang Homomorphism Theorem to solve this problem. We present the proof of the Lang Homomorphism Theorem based on Sylvester's Lemma 76.10. To use it, we shall need the fact (which we do not prove) that the trace form of an algebra over a field of characteristic zero can be diagonalized, i.e., has a diagonal matrix representation. This is in fact true for all symmetric bilinear forms over fields of characteristic not two. We prove this for the special case that we need.

Lemma 78.1. *Let E be a field of characteristic zero and K/E be a finite field extension of degree n. Then there exists an E-basis $\mathcal{B}$ for K such that the matrix representation of the trace form $\varphi : K \times K \to E$ in the basis $\mathcal{B}$ is diagonal.*

PROOF. We must produce an E-basis $\{x_1, \ldots, x_n\}$ for K satisfying $\varphi(x_i, x_j) = \delta_{ij}$ for all $i, j = 1, \ldots, n$. As K is a field and E perfect, the trace form is just the non-degenerate field trace $\operatorname{Tr}_{K/E} : K \times K \to E$ by Lemma 80.1. In particular, if x is a nonzero element in E, then there exists a y in E with $\operatorname{Tr}_{K/E}(xy)$ nonzero. As

$$\operatorname{Tr}_{K/E}(xy) = \tfrac{1}{2}\operatorname{Tr}_{K/E}\left((x+y)^2\right) - \operatorname{Tr}_{K/E}(x^2) - \operatorname{Tr}_{K/E}(y^2)$$

one of $\operatorname{Tr}_{K/E}\left((x+y)^2\right)$, $\operatorname{Tr}_{K/E}(x^2)$, $\operatorname{Tr}_{K/E}(y)^2)$ is non-zero, i.e., the nondegeneracy of $\operatorname{Tr}_{K/E}$ insures the existence of an element z in K with $\operatorname{Tr}_{K/E}(z^2)$ nonzero. (Of course, we know this is true for the element $z = 1$.) As $\operatorname{Tr}_{K/E}(zx) = 0$ for x in K implies that $x = 0$, the restriction of $\operatorname{Tr}_{K/E}$ to $(Fz)^{\perp} = \{x \mid \operatorname{Tr}_{K/E}(zy) = 0\}$, gives an 'orthogonal' decomposition $K = Ez \perp (Ez)^{\perp}$. Repeating the argument yields the result. $\square$

Theorem 78.2. (Lang Homomorphism Theorem) *Let F be a real closed field and $K = F(x_1, \ldots, x_n)$ a proper finitely generated formally real field extension of F (hence not algebraic). Then there exists infinitely many F-algebra homomorphisms $F[x_1, \ldots, x_n] \to F$, i.e., if R is an affine F-algebra that is a domain, then there exist infinitely many F-algebra homomorphisms $R \to F$.*

PROOF. Let $m = \operatorname{tr} \deg K/F$. By hypothesis, $m > 0$. We induct on m.

Suppose that $m = 1$. We may assume that $x = x_1$ is transcendental over F. As $\operatorname{char} F = 0$, we can write $K = F(x)[y]$ for some $y \in K$ and furthermore we may assume that y is integral over $F[x]$ (cf. Remark 79.14(1)).

Claim 78.3. There exist (infinitely many) F-algebra homomorphisms $\sigma : F[x, y] \to F$:

Let $f = f(x, t) = t^r + a_{r-1}(x)t^{r-1} + \cdots + a_0(x)$ with the $a_i(x) \in F[x]$ be the minimal polynomial of integral y over $F[x]$. It suffices to show that there exist (infinitely many) $(a, b) \in F^2$ such that $f(a, b) = 0$, since for each such (a, b) the map $\sigma : F[x, y] \to F$ via $x \mapsto a$ and $y \mapsto b$ is well-defined. Let $P \in \mathcal{X}(K)$ and let $\widetilde{K}$ be a real closure of (K, P). So we have $P \cap K = K^2$. Let φ be the trace form of the $F(x)$-algebra K. Since f has a root y in $\widetilde{K}$, by Sylvester's Lemma 76.10, we have $\operatorname{sgn}_P \varphi > 0$. As $r = [K : F(x)]$ and φ can be diagonalized over $F(x)$, we may assume that there is an $F(x)$-basis $\mathcal{B}$ for K such that

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} h_1(x) & 0 & \\ 0 & \ddots & \\ & & h_r(x) \end{pmatrix}$$

is diagonal for some $h_i(x) \in F(x)$. Moreover, modifying this basis, we may also assume that all the $h_i(x) \in F[x]$ and are square-free. As F is real closed, each of the $h_i(x)$ can be factored as a product of a constant in F, monic linear polynomials, and monic irreducible quadratic polynomials in x. Each irreducible monic polynomial $x^2 + ax + b = (x + \frac{a}{2})^2 + (b - \frac{a^2}{4})$, a sum of two squares in $F(x)$ hence totally positive as is $c^2 + ac + b$ for all $c \in F$ by substitution. Thus each h_i is a product of a totally positive or totally negative element and finitely many positive factors $x - a$ relative to P and finitely many negative factors $x - b$ relative to P for some $a, b \in F$.

Among all the finitely many a's occurring in all the h_i, let a_0 be the maximum relative to the ordering $F^2 = P \cap F$ and among all the finitely many b's occurring in all the h_i, let b_0 be the minimum relative to the ordering $P \cap F$. (If there are no a's (respectively, b's) choose a_0 (respectively, b_0) such that $a_0 <_P b_0$ in F.) Thus we have $x - a_0 \leq_P x - a$ for all such a and $x - b \leq_P x - b_0$ for all such b. As $x - b_0 <_P x - a_0$, we must also have $a_0 <_P b_0$. Thus there exist infinitely many $c \in F$ with $a_0 <_P c <_P b_0$ (e.g., $a_0 + \dfrac{b_0 - a_0}{2^N}$ for $N > 0$). For each such c, we have $c - a >_P 0$ and $x - a >_P 0$ for all the a's and $c - b <_P 0$ and $x - b <_P 0$ for all the b's. It follows that

$$\operatorname{sgn}_P \varphi = \operatorname{sgn}_{P\cap F} [\varphi]_{\mathcal{B}} = \operatorname{sgn}_{P\cap F} \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}$$

Assertion. Let $c \in F$ satisfy $a_0 <_P c <_P b_0$ and φ_c be the trace form of the F-algebra $A = F[t]/\big(f(c,t)\big)$. Then, except for finitely many exceptions, there exists an F-basis $\mathcal{C}$ for A satisfying

$$(*) \qquad [\varphi_c]_{\mathcal{C}} = \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}.$$

In particular, for each such c, we have $\operatorname{sgn}_{P\cap F} \varphi_c > 0$.

As $f(c,t)$ has a root in K hence $\widetilde{K}$, if (*) holds then $\operatorname{sgn}_{P\cap F} \varphi_c > 0$ by Sylvester's lemma 76.10. So we need only show the first part.

Let $B = \big(b_{ij}(x)\big)$ be the matrix representation of the trace form φ relative to the basis $\{1, y, \dots, y^{r-1}\}$ of K over $F(x)$. As y is integral over $F[x]$, each $b_{ij}(x) \in F[x]$. There exists a change of basis matrix $C = \big(c_{ij}(x)\big) \in \mathrm{GL}_r(K)$ such that

$$C^t BC = \begin{pmatrix} h_1(x) & 0 & \\ 0 & \ddots & \\ & & h_r(x) \end{pmatrix}.$$

If c is not a zero of the denominator of any of the $c_{ij}(x)$'s or a zero of $\det C$, then

$$C|_{x=c}^{t} B|_{x=c} C|_{x=c} \quad = \quad \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}$$

as desired. As there exist only finitely many c not satisfying these conditions, the assertion follows.

As F is real closed, for each c in the assertion satisfying $\operatorname{sgn}_{P\cap F} \varphi_c > 0$, there exists a $b \in F$ such that $f(c, b) = 0$ by Sylvester's Lemma 76.10. Thus there exist infinitely many (c, b) such that $f(c, b) = 0$ and the Claim is established.

Next we show that almost all of the F-algebra homomorphisms $\sigma : F[x, y] \to F$ extend to $F[x, y, x_2, \dots, x_n] \to F$. As the transcendence degree $m = \text{tr deg}_F K = 1$ and $F(x)[y] = F(x_1, \dots x_n)$, we can write

$$x_i = \frac{g_i(x, y)}{q_i(x)} \text{ for some } g_i \in F[x, t], \quad 0 \neq q_i \in F[x].$$

Let $q = \prod_i q_i$. As $q(c) \neq 0$ for almost all of the c constructed above, $\sigma\big(q(x)\big) \neq 0$ for almost all the σ constructed above. Let σ be any such one. By properties of localization with $S = \{t^i \mid i \geq 0\}$, any such σ extends to an F-algebra homomorphism $\sigma : F[x, y][\frac{1}{q}] \to F$ by Exercise 29.4(2). Since $F[x_1, \dots, x_n] \subset F[x_1, y, x_2, \dots, x_n] \subset F[x, y][\frac{1}{q}]$, the $m = 1$ case is completed.

Suppose that $m > 1$. Choose $K/E/F$ satisfying $\text{tr deg}_E K = 1$. Let $\widetilde{E}$ be the field of algebraic elements over E lying in a real closure $\widetilde{K}$ of K (called the algebraic closure of E in $\widetilde{K}$). By the Fundamental Theorem of Algebra, $\widetilde{E}$ is also real closed. As $\text{tr deg}_{\widetilde{E}} \widetilde{E}(x_1, \dots, x_n) = 1$, there exist infinitely many $\widetilde{E}$-algebra homomorphisms $\sigma : \widetilde{E}(x_1, \dots, x_n) \to \widetilde{E}$ by the $m = 1$ case. Let σ be any one of these. If $\sigma(x_i) \in F$ for all i, then $\sigma|_{F[x_1,\dots,x_n]} : F[x_1, \dots, x_n] \to F$ is an F-algebra homomorphism. So suppose that there exists an i with $\sigma(x_i) \notin F$. As $\sigma(x_i) \in \widetilde{E}$ for all i, we have $\text{tr deg}_F F\big(\sigma(x_1), \dots, \sigma(x_n)\big) \leq \text{tr deg}_F \widetilde{E} = \text{tr deg}_F E < m$. By induction there exist infinitely many F-algebra homomorphisms $\tau : F[\sigma(x_1), \dots, \sigma(x_n)] \to F$. Then the compositions $\tau \circ \sigma|_{F[x_1,\dots,x_n]} : F[x_1, \dots, x_n] \to F$ give infinitely many F-algebra homomorphisms. □

A commutative ring is called *semi-real* if -1 is not a sum of squares in R and is called *formally real* if $\sum x_i^2 = 0$ in R implies that $x_i = 0$ for all i. More generally, an ideal $\mathfrak{A}$ in a commutative ring is called *semi-real* (respectively, *formally real*) if $R/\mathfrak{A}$ is semi-real (respectively, formally real). We note that if R is a domain, then it is formally real if and only if its quotient field is and a field is semi-real if and only if formally real.

Lemma 78.4. *Let R be a commutative ring. Then the following are equivalent*

(1) *R is semi-real.*
(2) *There exists $P \in \mathcal{Y}(R)$ such that $P \cap -P$ is a prime ideal.*

(3) There exists a formally real prime ideal $\mathfrak{p}$ in R.
(4) There exists a ring homomorphism $R \to F$ for some formally real field F.

Proof. (1) $\Rightarrow$ (2): As R is semi-real, we have $\sum R^2 \in \mathcal{Y}(R)$. Thus (2) follows by Lemma 75.4.
(2) $\Rightarrow$ (3): The prime ideal $P \cap -P$ in (2) is clearly semi-real and excludes the multiplicative set $S = 1 + \sum R^2$. By choice $0 \notin S$ and $S + \sum R^2 \subset S$. By Zorn's lemma there exists an ideal $\mathfrak{p}$ containing $P \cap -P$ and excluding S. It is prime (as is well-known or easy to check). The quotient field of $R/\mathfrak{p}$ is checked to be semi-real hence formally real. It follows that $R/\mathfrak{p}$ is formally real.
(3) $\Rightarrow$ (4): The quotient field of $R/\mathfrak{p}$ works.
(4) $\Rightarrow$ (1): As a quotient of R is semi-real so is R. $\square$

Corollary 78.5. *Let F be a real closed field and R a domain that is an affine F-algebra. If R is semi-real the there exists an F-algebra homomorphism $R \to F$.*

Proof. There exists a formally real prime ideal $\mathfrak{p}$ in R. Then $R/\mathfrak{p}$ is a real F-affine ring that is a domain so there exists an F-algebra homomorphism $R/\mathfrak{p} \to R$ by the Lang Homomorphism Theorem 78.2. The composition $R \to R/\mathfrak{p} \to R$ now works. $\square$

Corollary 78.6. *Let F be a real closed field and R a real affine F-algebra that is a domain. Let $f_1, \dots, f_n \in R \setminus \{0\}$. Then there exists an F-algebra homomorphism $\varphi : R \to F$ so that $\varphi(f_i) \neq 0$ for all i.*

Proof. $R[f_1^{-1}, \dots f_n^{-1}]$ is a real affine F-algebra that is a domain. $\square$

Note that $R = \mathbb{R}[t_1, \dots, t_n]/(t_1^2 + \cdots + t_n^2)$ is a semi-real affine $\mathbb{R}$-algebra that is a domain but it is not formally real since any $\mathbb{R}$-algebra homomorphism $R \to \mathbb{R}$ must take each $t_i \mapsto 0$.

Corollary 78.7. *Let F be a real closed field and R an affine F-algebra. Let $f_1, \dots, f_r$, $g_1, \dots, g_s$ lie in R. Suppose that there exists a maximal preordering $P \in \mathcal{Y}(R)$ such that $0 \neq f_i \in P$ for all i and $g_j \in P$ for all j. Then there exists an F-algebra homomorphism $\varphi : R \to F$ so that $0 \neq \varphi(f_i) \in F^2$ and $\varphi(g_j) \in F^2$ for all i and j.*

Proof. Let $\mathfrak{p} = P \cap -P$, a prime ideal in R. Replacing R by $R/\mathfrak{p}$ we may assume that $P \cap -P = \mathfrak{p} = 0$. In particular, R is a domain and P is an ordering on R. Clearly, P extends to an ordering of the quotient field K of R. Also call this extension P. We have $f_i >_P 0$ and $g_j \geq_P 0$ for all i and j. Let $E = K(\sqrt{f_1}, \dots, \sqrt{f_r}, \sqrt{g_1}, \dots, \sqrt{g_s})$. By induction on $r + s$, the ordering P extends to an ordering Q on E. Since the domain $A = R[f_1^{-1}, \dots, f_r^{-1}, \sqrt{f_1}, \dots, \sqrt{f_r}, \sqrt{g_1}, \dots, \sqrt{g_s}]$ is a real affine F-algebra, there exists an F-algebra homomorphism $A \to F$. This homomorphism has the desired properties. $\square$

If $\mathfrak{A}$ be an ideal in $F[t_1, \dots, t_n]$ with F infinite, let
(1) $Z_F(\mathfrak{A}) = \{x \in F^n \mid f(x) = 0 \text{ for all } f \in \mathfrak{A}\}$.
(2) $A_F(\mathfrak{A}) = F[t_1, \dots, t_n]/\mathfrak{A}$.

Suppose that $\mathfrak{p} = \mathfrak{A}$ is a prime ideal in $F[t_1, \dots, t_n]$. So $A_F(\mathfrak{p})$ is a domain. As usual we say a rational function f in the quotient field of $A_F(\mathfrak{p})$ is *defined* at $x \in Z_F(\mathfrak{p})$ if $f = g/h$ for some $g, h \in A_F(\mathfrak{A})$ with $h(x) \neq 0$. If F is real closed, a rational function f in the quotient field of $A_F(\mathfrak{p})$ is called *positive semi-definite* if $f(x) \geq 0$ for all points $x \in A_F(\mathfrak{p})$ where f is defined.

Corollary 78.8. (Weak Real Nullstellensatz) *If F is real closed and $\mathfrak{A}$ is an ideal in $F[t_1, \dots, t_n]$, then $\mathfrak{A}$ is semi-real if and only if $Z_F(\mathfrak{A}) \neq \emptyset$.*

Proof. Let $A_F(\mathfrak{A}) = F[x_1, \dots, x_n]$. If $(a_1, \dots, a_n) \in Z_F(\mathfrak{A})$, then $f : A_F(\mathfrak{A}) \to F$ by $x_i \to a_i$ defines an F-algebra homomorphism and $\mathfrak{A}$ is semi-real by Lemma 78.4. Suppose that $\mathfrak{A}$ is semi-real. By Lemma 78.4 there exists a prime ideal $\mathfrak{p}$ in $F[t_1, \dots, t_n]$ such that $\mathfrak{p}/\mathfrak{A}$ is a formally real prime ideal in $A_F(\mathfrak{A})$. It follows that $F[t_1, \dots, t_n]/\mathfrak{p}$ is a formally real domain so by the Lang Homomorphism Theorem 78.2 there exists an F-algebra homomorphism $\sigma : F[t_1, \dots, t_n]/\mathfrak{p} \to F$. Then $(a_1, \dots, a_n) \in Z_F(\mathfrak{A})$ with $a_i = \sigma(t_i + \mathfrak{p})$. $\square$

Theorem 78.9. (Artin) *Let F be a real closed field and $\mathfrak{p}$ a formally real prime ideal in $F[t_1, \dots, t_n]$. Let K be the quotient field of $A_F(\mathfrak{p})$. Let $f \in K$. If f is positive, then f is a sum of squares in K.*

Proof. Suppose that f is not a sum of squares in K. Then there exists $P \in \mathcal{X}(K)$ such that $f <_P 0$ by Corollary 75.6. Thus P extends to $E = K(\sqrt{-f})$ by Theorem 76.2. Let $f = g/h$, with $g, h \in A_F(\mathfrak{p})$ and let $^- : F[t_1, \dots, t_n] \to A_F(\mathfrak{p})$ be the canonical map. By the Lang Homomorphism Theorem, there exists an F-algebra homomorphism $\sigma : F[\bar{t}_1, \dots \bar{t}_n, \sqrt{-g}, \dfrac{1}{gh}] \to F$. For each i, let $a_i = \sigma(\bar{t}_i)$. If $q \in \mathfrak{p}$ then

$$q(a_1, \dots, a_n) = q\big(\sigma(\bar{t}_1), \dots, \sigma(\bar{t}_n)\big) = \sigma\big(q(\bar{t}_1, \dots, \bar{t}_n)\big) = 0,$$

so $(a_1, \dots, a_n) \in Z_F(\mathfrak{p})$. As $\sigma(g) \neq 0$ and $\sigma(h) \neq 0$, the rational function f is defined at $(a_1, \dots, a_n)$ and is not zero at it. Since $\sigma(-f) \in F^2$, we must have $\sigma(f) <_P 0$. But $\sigma(f) = f(a_1, \dots, a_n) >_P 0$, a contradiction. $\square$

Artin's Theorem immediately answers Hilbert's 17th problem whether every positive semi-definite function f in $\mathbb{R}(t_1, \dots, t_n)$ is a sum of squares in K in the affirmative.

Corollary 78.10. (Hilbert's 17th Problem) *Any positive semi-definite function in $\mathbb{R}(t_1, \dots t_n)$ is a sum of squares in $\mathbb{R}(t_1, \dots, t_n)$.*

The Motzkin polynomial $t_1^4 t_2^2 + t_1^2 t_2^4 - 3t_1^2 t_2^2 + 1$ in $\mathbb{R}[t_1, t_2]$ is positive semi-definite but not a sum of squares in $\mathbb{R}[t_1, t_2]$. [It is a sum of four squares, but not three squares, in $\mathbb{R}(t_1, t_2)$. Cf. Lam's book *Introduction to quadratic forms over fields* [35] for more detail.]

Part 6

Commutative Algebra and Algebraic Number Theory

CHAPTER XV

Dedekind Domains

The study of fields arose from the desire to understand the roots of polynomials with rational coefficients. In a similar way, algebraic number theory arose from the desire to understand the roots of monic polynomials with integer coefficients. In this chapter, we give an introduction to algebraic number theory. Since the generalization to studying Dedekind domains – an appropriate generalization of the integers – is not too difficult, we take this more general approach.

79. Integral Elements

We begin with the study of monic polynomials over (nonzero) commutative rings.

Definition 79.1. Let B be a nonzero commutative ring and A a subring of B. We shall write this as B/A and call it a *ring extension* mimicking how we wrote field extensions. An element x in B is called *integral over* A if x is a root of a monic polynomial f in $A[t]$.

Examples 79.2. 1. Let K/F be an extension of fields. Then an element x in K is integral over F if and only if it is algebraic over F.

2. If A is a nonzero commutative ring, then every element x in A is integral over A as it is a root of $t - x$ in $A[t]$.
3. Let d be a square-free integer and $x = a + b\sqrt{d}$ with a and b in $\mathbb{Q}$. Then x is a root of $t^2 - 2at + (a^2 - b^2 d)$. In particular, one checks that x is integral over $\mathbb{Z}$ if and only if $2a$ and $a^2 - b^2 d$ are integers. For example, the complex numbers $\sqrt{2}$ and $(-1 + \sqrt{-3})/2$ are integral over $\mathbb{Z}$.
4. Every nth root of unity in $\mathbb{C}$, $n > 0$, is integral over $\mathbb{Z}$.
5. Let $C/B/A$ be ring extensions. If x in C is integral over A, then it is integral over B.

Because of the first example, we expect that some of our field theoretic ideas and results should generalize. This is true. First, we must find a ring theoretic test for an element to be integral. It is given by the following basic result.

Proposition 79.3. *Let B/A be an extension of nonzero commutative rings and x an element in B. Then x is integral over A if and only if $A[x]$ is a finitely generated A-module.*

PROOF. ($\Rightarrow$): Let x be a root of the monic polynomial f in $A[t]$. By the General Division Algorithm 34.4, we can write

$$g = fq + r \ \text{ with } \ q, r \in A[t] \ \text{ and } \ r = 0 \ \text{ or } \ \deg r < \deg f.$$

Therefore, if $n = \deg f$, we have

$$g(x) = r(x) \text{ lies in } \sum_{i=0}^{n-1} Ax^i, \text{ so } A[x] = \sum_{i=0}^{n-1} Ax^i.$$

($\Leftarrow$): We prove a more general statement.

Claim 79.4. Let M be an $A[x]$-module that is finitely generated as an A-module. If $\operatorname{ann}_{A[x]} M = 0$ (i.e., if r in $A[x]$ satisfies $rm = 0$ for all m in M, then $r = 0$), then x is integral over A:

Suppose that $M = Am_1 + \cdots + Am_n$. By assumption, M is an $A[x]$-module, i.e.,

$$\lambda_x : M \to M \text{ given by } m \mapsto xm$$

is an A-endomorphism. For each i, $i = 1, \dots, n$, we can write

$$xm_i = \sum_{j=1}^{n} a_{ij} m_j \text{ some } a_{ij} \in A,\ 1 \le i, j \le n.$$

Therefore,

$$\sum_{j=1}^{n} (\delta_{ij} x - a_{ij}) m_j = 0 \text{ for } i = 1, \dots, n.$$

Let Δ be the determinant of the $n \times n$ matrix $(\delta_{ij} x - a_{ij})$. By Cramer's Rule, $\Delta m_i = 0$ for $i = 1, \dots, n$, so $\Delta M = 0$, i.e., Δ lies in $\operatorname{ann}_{A[x]} M = 0$. It follows that x is a root of the monic polynomial $\det((\delta_{ij} t - a_{ij}))$ in $A[t]$. This proves the claim.

Let $M = A[x]$, a finitely generated A-module by assumption. As $A[x]$ is a subring of B it has a one (not zero). Consequently, $\operatorname{ann}_{A[x]} A[x] = 0$, so the claim yields the desired result. □

Corollary 79.5. *Let B/A be an extension of nonzero commutative rings, $x_1, \dots, x_n$ elements of B. Then $x_1, \dots, x_n$ are all integral over A if and only if $A[x_1, \dots, x_n]$ is a finitely generated A-module.*

Proof. ($\Rightarrow$): The case for $n = 1$ follows from the proposition. By induction, $A[x_1, \dots, x_{n-1}]$ is a finitely generated A-module, say with generating set S. As x_n is integral over A, it is a root of a monic polynomial in $A[t]$, say of degree d. Then $\bigcup_{i=0}^{d-1} \{sx^i \mid s \in S\}$ is a generating set for $A[x_1, \dots, x_n]$.

($\Leftarrow$): Since $\operatorname{ann}_{A[x_i]} A[x_1, \dots, x_n] = 0$ for $i = 1, \dots, n$, this follows from Claim 79.4. □

In the above, we could also have used the following lemma that we leave as an exercise. We shall need it below.

Lemma 79.6. *Let B/A be an extension of nonzero commutative rings. If M is a finitely generated B-module and B a finitely generated A-module, then M is a finitely generated A-module.*

Note the corollary above is analogous with the result that when K/F is a field extension, $x_1, \dots, x_n$ elements in K, then $F(x_1, \dots, x_n)/F$ is finite if and only if $x_1, \dots, x_n$ are all algebraic over F. In a similar way we have the following analogue of another field theoretical result.

Corollary 79.7. *Let B/A be an extension of nonzero commutative rings, x and y elements of B both integral over A. Then $x \pm y$, and xy are integral over A. In particular,*

$$\{x \in B \mid x \text{ integral over } A\}$$

is a subring of B.

Proof. If z is any element in $A[x, y]$, then $A[x, y, z] = A[x, y]$ is a finitely generated A-module, so z is integral over A. □

Of course, if B/A is an extension of commutative rings with x in B a unit in B, it may be integral over A but its inverse not. For example, $1/2$ is not integral over $\mathbb{Z}$.

Definition 79.8. Let B/A be an extension of nonzero commutative rings and

$$C = \{x \in B \mid x \text{ integral over } A\}.$$

The ring C is called the *integral closure* of A in B. If $C = B$, i.e., every element of B is integral over A, we say B/A is *integral* and if $C = A$, we say that A is *integrally closed* in B.

Therefore, B/A being integral is the analogue of an algebraic extension in field theory. Analogous to the field case, we have:

Corollary 79.9. *Let C/B and B/A be extensions of nonzero commutative rings. Then C/B and B/A are integral extensions if and only if C/A is an integral extension.*

Proof. We need only show if C/B and B/A are integral extensions so is C/A. Let c be an element in C. As c is integral over B, it satisfies an equation $c^n + b_{n-1}c^{n-1} + \cdots + b_0 = 0$ in B for some $b_0, \dots, b_{n-1}$ in B. Let $B_0 = A[b_0, \dots, b_{n-1}]$. As B/A is integral, it follows that $A[c, b_0, \dots, b_{n-1}]$ is a finitely generated B_0-module by Proposition 79.3 and B_0 is a finitely generated A-module by Corollary 79.5. The result now follows using Lemma 79.6. □

Corollary 79.10. *Let $C/B/A$ be extensions of nonzero commutative rings with B/A integral. Then the integral closure of A in C is the same as the integral closure of B in C.*

We shall mostly be interested in the case that our rings are domains. For this case, we shall use the following notation:

Notation 79.11. If A is a domain and K any field containing A, we shall denote the integral closure of A in K by A_K, i.e.,

$$A_K := \{x \in K \mid x \text{ is integral over } A\}.$$

Definition 79.12. Let A be a domain. We say that A is *integrally closed* or a *normal domain* if A is integrally closed in its quotient field, i.e., if F is the quotient field of A, then $A = A_F$.

An important example of integrally closed domains is given by:

Proposition 79.13. *A UFD is integrally closed. In particular, any PID is integrally closed.*

PROOF. Let A be a UFD and a and b be elements in A with b nonzero. Suppose that a/b is integral over A. As A is a UFD, we may assume that a and b are relatively prime. By definition, a/b satisfies an equation

$$(\frac{a}{b})^n + c_1(\frac{a}{b})^{n-1} + \cdots + c_n = 0$$

for some $c_1, \dots, c_n$ in A. Therefore, we have the equation

$$a^n = -b(c_1 a^{n-1} + \cdots + b^{n-1}c_n) \text{ in } A.$$

It follows that $b \mid a^n$ in the UFD A, hence b is a unit in A and a/b lies in A. □

Remarks 79.14. Let A be a domain, F its quotient field, and $L/K/F$ algebraic field extensions.

1. If x lies in L, then there exists a nonzero element c in A satisfying cx is integral over A, i.e., lies in A_L. Indeed if x satisfies
$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$
in L with $a_0, \dots, a_n$ elements of A and a_n nonzero, then multiplying this equation by a_n^{n-1} shows that $a_n x$ is integral over A.
2. We have $A_L \cap K = A_K$.
3. The quotient field of A_K is K by the first remark.
4. If E/F is a field extension and $E/K/F$ with K the maximal algebraic field extension of F in E, then $A_K = A_L$.

The case in which we are most interested is when A is the ring of integers. If $K/\mathbb{Q}$ is a finite field extension, then K is called an *algebraic number field* and $\mathbb{Z}_K$ is called the *ring of algebraic integers in* K. Let Ω be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Then algebraic number theory is the study of $\mathbb{Z}_\Omega$. As Ω is the union of number fields, we have $\mathbb{Z}_K = \mathbb{Z}_\Omega \cap K$ for any number field K. Since $\mathbb{Q}$ is perfect, we shall not have to worry about separability that in the general case can cause serious problems.

Exercises 79.15.

1. Let d be a square-free integer and $x = a + b\sqrt{d}$ with a and b in $\mathbb{Q}$. Show that x is integral over $\mathbb{Z}$ if and only if $2a$ and $a^2 - b^2 d$ are integers.
2. Let $\varphi : A \to B$ be a ring homomorphism and M a finitely generated B-module. If B is a finitely generated A-module via the pullback, then M is a finitely generated A-module. In particular, Lemma 79.6 is valid.
3. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. We say that φ is *integral* if $B/\varphi(A)$ is integral. Show that a composition of integral homomorphisms is integral.

4. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. We say that φ is *finite* if B is a finitely generated A-module (via φ) and of *finite type* if there exists a ring epimorphism $A[t_1, \ldots, t_n] \to B$ for some n lifting φ. Show that φ is finite if and only if φ is integral and of finite type.
5. Let A be a domain with quotient field F and K/F a field extension. If S is a multiplicative set (not containing zero) in A, show that $(S^{-1}A)_K = S^{-1}(A_K)$, i.e., taking localization and integral closures commute. In particular, this applies when $S = A \setminus \mathfrak{p}$ with $\mathfrak{p}$ a prime ideal in A.

80. Integral Extensions of Domains

In this section, we shall investigate the integral closure of a domain in a separable extension of its quotient field. Separability makes life much nicer, and since we are mostly interested in applications to algebraic number theory, this will not be an impairment.

When we investigated cyclic extensions of fields, we needed to use the norm of such an extension. Although we introduced the notion of trace at the same time, except for exercises, we did not use it. Although the norm and trace of an arbitrary, i.e., not necessarily separable, finite extension can be defined, this is not useful for the trace. Indeed, the trace from a finite inseparable extension turns out to be the zero map. However, for the separable case, the trace is very useful. This was because of Exercise 60.26(5). Because this is important to our current investigation, we now prove this exercise.

Lemma 80.1. *Let K/F be a finite separable extension of fields of degree n, $K^* := \mathrm{Hom}_F(K, F)$, the F-linear dual space of K. Then the map $T : K \to K^*$ defined by $x \mapsto \mathrm{tr}_x : y \mapsto \mathrm{Tr}_{K/F}(xy)$ is an F-linear isomorphism. In particular, if $\mathcal{B} = \{w_1, \ldots, w_n\}$ is an F-basis for K, then there exists an F-basis $\{w'_1, \ldots, w'_n\}$ for K satisfying $\mathrm{Tr}_{K/F}(w_i w'_j) = \delta_{ij}$ for all $i, j = 1, \ldots, n$.*

Remark 80.2. The basis $\{w'_1, \ldots, w'_n\}$ in the lemma is called the *complementary basis* to $\mathcal{B}$

Proof. By Dedekind's Lemma 54.3, the map $\mathrm{Tr}_{K/F} : K \to F$ is nontrivial, so there exists an element z in K with $\mathrm{Tr}_{K/F}(z)$ nonzero. Therefore, $\mathrm{Tr}_{K/F}(xx^{-1}z)$ is not zero for every nonzero x in K. It follows that tr_x is nonzero for all nonzero x in K, i.e., $T : K \to K^*$ is injective, hence an isomorphism by dimension count. Let $\{f_1, \ldots, f_n\}$ be the dual basis to $\mathcal{B}$ and choose w'_i in K to satisfy $T(w'_i) = f_i$ for $i = 1, \ldots, n$. Then $\{w'_1, \ldots, w'_n\}$ works. $\square$

Suppose that K_i/F is a field extension for $i = 1, 2$ and $\sigma : K_1 \to K_2$ is an F-isomorphism. Let A be a domain lying in F. If x in K_1 is integral over A then $\sigma(x)$ is also integral over A, i.e., $\sigma(A_{K_1}) \subset A_{K_2}$. In particular, if $K = K_1 = K_2$ then $\sigma(A_K) = A_K$. This means that we can use Galois theory. We shall need the results about the trace (and norm) for finite separable extensions as stated in Remark 60.18 as well as Exercise 60.26(3) which we now prove.

Lemma 80.3. *Let K/F be a finite separable extension of fields. Then $\mathrm{N}_{K/F} = \mathrm{N}_{E/F} \circ \mathrm{N}_{K/E}$ and $\mathrm{Tr}_{K/F} = \mathrm{Tr}_{E/F} \circ \mathrm{Tr}_{K/E}$ for any intermediate field $K/E/F$.*

Proof. Let L/F be a finite Galois extension with L/K. Let $\sigma_1, \ldots, \sigma_n : E \to L$ denote all the F-homomorphisms and $\tau_1, \ldots, \tau_m : K \to L$ all the E-homomorphisms. Extend each σ_i to $\widehat{\sigma}_i$ in $G(L/F)$. If $\mu : K \to L$ is an F-homomorphism, then $\mu|_E = \sigma_i$ for some i and $\widehat{\sigma}_i^{-1}\mu = \tau_j$ for some j, hence $\mu = \widehat{\sigma}_i\tau_j$. As in the proof of Proposition 56.11, we see that these are all the distinct E-homomorphisms $K \to L$. It follows that if x lies in K, then

$$\mathrm{Tr}_{K/F}(x) = \sum \widehat{\sigma}_i\tau_j(x) = \sum \widehat{\sigma}_i(\mathrm{Tr}_{K/E}(x)) = \mathrm{Tr}_{E/F}(\mathrm{Tr}_{K/E}(x))$$

and similarly for the norm. □

Let K/F be finite separable closure with L/K a normal closure of K/F and $\sigma_1, \ldots, \sigma_n : K \to L$ all the F-homomorphisms. If x is an element of K, then any elementary symmetric function in $\sigma_1(x), \ldots, \sigma_n(x)$ is fixed by $G(L/F)$ hence lies in F. It also follows that $\mathrm{N}_{K/F(x)}(x) = x^{[K:F(x)]}$.

Proposition 80.4. *Let A be a domain with quotient field F and K/F a finite separable extension. If x in K is integral over A, then the minimal polynomial $m_F(x)$ of x lies in $A_F[t]$ and both $\mathrm{Tr}_{K/F}(x)$ and $\mathrm{N}_{K/F}(x)$ lie in A_F. In particular, if A is integrally closed, then $m_F(x)$ lies in $A[t]$ and $\mathrm{Tr}_{K/F}(x)$ and $\mathrm{N}_{K/F}(x)$ lie in A.*

Proof. Let L/F be a finite Galois extension satisfying L/K and $\sigma_1, \ldots, \sigma_n : K \to L$ all the distinct F-homomorphisms. If x is an element of K, then any elementary symmetric function in $\sigma_1(x), \ldots, \sigma_n(x)$ is fixed by $G(L/F)$ hence lies in F. Therefore, $m_F(x) = \prod_{i=1}^n (t - \sigma_i(x))$ lies in $A_L[t] \cap F[t]$. It follows that $m_F(x)$ lies in $A_F[t]$. By the lemma and Property 60.17(3) (cf. Remark 60.18), we have

$$\mathrm{N}_{F(x)/F}(\mathrm{N}_{K/F(x)}(x)) = (\mathrm{N}_{F(x)/F}(x))^{[K:F(x)]},$$

hence lies in $A_L \cap F = A_F$, and similarly for the trace. □

The result for which we are aiming can now be established.

Theorem 80.5. *Let A be an integrally closed Noetherian domain with quotient field F. Let K/F be a finite separable extension. Then the integral closure of A in K is a finitely generated A-module. In particular, A_K is also an integrally closed Noetherian domain.*

Proof. By Theorem 40.7, we know that any finitely generated A-module is a Noetherian A-module, since A is a Noetherian ring. In particular, it suffices to show there exists a finitely generated A-module M with A_K a submodule of M. Let $n = [K : F]$ and $\mathcal{B} = \{w_1, \ldots, w_n\}$ be an F-basis for K. As K is the quotient field of A_K, by clearing denominators, we may assume that $\mathcal{B}$ lies in A_K. As K/F is a finite separable extension, there exists a complementary basis $\mathcal{B}' = \{w_1', \ldots, w_n'\}$ to $\mathcal{B}$ for K by Lemma 80.1, i.e., $\mathrm{Tr}_{K/F}(w_i w_j') = \delta_{ij}$ for all i and j. For each j, $j = 1, \ldots, n$, there exists a nonzero c_j in A such that $c_j w_j'$ lies in A_K by Remark 79.14(1). Let $c = c_1 \cdots c_n \neq 0$. To finish, it suffices to show the following:

Claim 80.6. A_K is a submodule of $\sum A c^{-1} w_i$:

Let z be an element of A_K. We can write $z = \sum_{i=1}^n b_i w_i$ for some $b_1, \dots, b_n$ in F. Then for each j, $1 \leq j \leq n$, we have

$$\mathrm{Tr}_{K/F}(czw'_j) = c\,\mathrm{Tr}_{K/F}(zw'_j) = cb_j.$$

As z and each cw'_j lies in A_K, we conclude that cb_j lies in $A_F = A$ for every j, i.e., $b_j \in c^{-1}A$ for each j. The result follows. $\square$

Remark 80.7. The result is false if we drop the separability assumption. Indeed there exist counterexamples with A a PID with precisely one nonzero prime ideal.

Corollary 80.8. *Let A be a PID with quotient field F and K/F a finite separable field extension. Then A_K is a finitely generated free A-module of rank $[K : F]$.*

PROOF. By the theorem, A_K is a finitely generated A-module. Since A_K is a domain, it is a torsion-free A-module, hence A-free by Corollary 44.16 to the Fundamental Theorem of Finite Generated Modules over a PID. Therefore, we need only compute the rank of A_K. We showed in the proof above that there exists an F-basis for K lying in A_K, so the rank of A_K is at most $[K : F]$ using Proposition 44.1. But any A-linearly independent set in A_K is F-linearly independent, so we must have A_K be of rank $[K : F]$. $\square$

Let A be a domain with quotient field F and K/F a finite field extension. If A_K is a free A-module of rank $[K : F]$, then an A-basis for is called an *integral basis*. The corollary applies to any ring of algebraic numbers $\mathbb{Z}_K$. It says that it is a free abelian group of rank $[K : \mathbb{Q}]$, hence has an integral basis. The theorem also applies to the case that $A = F[t]$ with F a field, and $K/F(t)$ a finite separable extension. Then $F[t]_K$ is a free $F[t]$-module of rank $[K : F(t)]$. These are the primary examples of our next study, when we further restrict our domain.

Exercise 80.9. Let K be a quadratic extension of $\mathbb{Q}$. Write $K = \mathbb{Q}(\sqrt{D})$ with D a square-free integer. Show every element α of K can be written as $\alpha = (a + b\sqrt{D})/2$ with a and b integers. In particular, $Tr_{K/\mathbb{Q}}(\alpha) = a$ and $N_{K/\mathbb{Q}}(\alpha) = (a^2 + b^2 D)/4$. Show

(*i*) If $\alpha = (a + b\sqrt{D})/2$ with a and b integers then $\alpha \in \mathbb{Z}_K$ if and only if

$$\begin{array}{llll} a \equiv b & \bmod 2 & \text{when} & D \equiv 1 \mod 4 \\ a \equiv b \equiv 0 & \bmod 2 & \text{when} & D \equiv 2, 3 \mod 4 \end{array}$$

(*ii*) Find a $\mathbb{Z}$-basis for $\mathbb{Z}_K$.

81. Dedekind Domains

We know that PIDs are UFDs, but rings of algebraic integers are not always UFDs, e.g., $\mathbb{Z}_{\mathbb{Q}[\sqrt{-5}]}$. We wish to study the appropriate generalization of PIDs that allows us to replace the property of UFDs, which is a property about elements, with an analogous property about ideals. We shall begin using the theory that we have already developed. We need the following observation:

Lemma 81.1. *Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. If $\mathfrak{P}$ is a prime ideal in B, then $\varphi^{-1}(\mathfrak{P})$ is a prime ideal in A. In particular, if φ is the inclusion of rings, then $\mathfrak{P} \cap A$ is a prime ideal in A.*

PROOF. The homomorphism φ induces a monomorphism of rings $\overline{\varphi} : A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$. As $B/\mathfrak{P}$ is a domain so is $A/\varphi^{-1}(\mathfrak{P})$ and the result follows. □

We come to the main subject of this chapter.

Definition 81.2. Let A be a domain, not a field. Then A is called a *Dedekind domain* if A is a Noetherian integrally closed domain in which every nonzero prime ideal is a maximal ideal. Let $\mathrm{Max}(A) = \{\mathfrak{p} \mid \mathfrak{p} \text{ a maximal ideal}\}$. Then $\mathrm{Spec}(A) = \{0\} \cup \mathrm{Max}(R)$ in a Dedekind domain A.

Examples 81.3. The verifications of the following are left as exercises:

1. Let A be a PID that is not a field. Then A is a Dedekind domain. In particular, $\mathbb{Z}$ and $F[t]$, with F a field, are Dedekind domains.
2. Let A be a Dedekind domain and S a multiplicative set not containing zero. Then the localization $S^{-1}A$ of A is a Dedekind domain or a field. In particular, if $\mathfrak{p}$ is a maximal ideal in A, then the localization $A_\mathfrak{p} = S^{-1}A$ with $S = A \setminus \mathfrak{p}$ is a Dedekind domain.
3. Recall that a *local ring* is a commutative ring with a unique maximal ideal. A Dedekind domain with a unique maximal ideal, so a local integrally closed, Noetherian domain that is not a field, is called a *discrete valuation ring*. If A is a Dedekind domain and $\mathfrak{p}$ a maximal ideal in A, then the localization $A_\mathfrak{p}$ of A is a discrete valuation ring, e.g., $\mathbb{Z}_{(p)}$ is a discrete valuation ring where p is a prime in $\mathbb{Z}$ as is $F[t]_{(f)}$ with F a field and f an irreducible polynomial in $F[t]$.

Remark 81.4. If A is a nonzero commutative ring, we define the *(Krull) dimension* of A by

$$\dim A = \max\{n \mid \text{there exists a chain of prime ideals } \mathfrak{p}_0 < \cdots < \mathfrak{p}_n \text{ in } A\},$$

if this a finite number (and to be infinite if not). (The trivial ring is said to have (Krull) dimension -1 (or $-\infty$.) With this definition, a Dedekind domain is a Noetherian integrally closed domain of dimension one. The geometric analogue of a Dedekind domain turns out to be a smooth affine curve. The points of the curve correspond to the maximal ideals in a Dedekind domain.

Our extension theory allows us to conclude:

Theorem 81.5. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Then A_K is a Dedekind domain.*

PROOF. A_K is integrally closed by definition and Noetherian by Theorem 80.5. So it suffices to show if $\mathfrak{P}$ is a nonzero prime ideal in A_K, then it is a maximal ideal, equivalently, $A_K/\mathfrak{P}$ is a field. By the lemma, $\mathfrak{P} \cap A$ is a prime ideal. If x is a nonzero element in $\mathfrak{P}$, it satisfies an equation

$$(*) \qquad x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0 \text{ for some } a_0, \dots, a_{n-1} \text{ in } A.$$

As A_K is a domain, we may assume that a_0 is nonzero. Since a_0 lies in $\mathfrak{P} \cap A$, the prime ideal $\mathfrak{p} = \mathfrak{P} \cap A$ is nonzero hence maximal, so $A/\mathfrak{p}$ is a field. But A_K is a finitely generated A-module by Theorem 80.5, so the domain $A_K/\mathfrak{P}$ is a finite dimensional $(A/\mathfrak{p})$-vector space. It follows that $A_K/\mathfrak{P}$ is a field by Corollary 48.19. □

Using equation (*) above, we easily see that the following is true:

Corollary 81.6. *Let A be a domain with quotient field F and K/F a finite field extension. If $\mathfrak{A}$ is a nonzero ideal in A_K, then $\mathfrak{A} \cap A_F$ is nonzero.*

Corollary 81.7. *Let A be a PID with quotient field F and K/F a finite separable extension. Then A_K is a Dedekind domain. Moreover, A_K has an integral basis.*

Remarks 81.8. 1. The corollary shows that every ring of algebraic integers is a Dedekind domain and is a finitely generated free abelian group.

2. The corollary shows that $F[t]_K$ for any field F and finite separable extension K of $F(t)$ is also a Dedekind domain and is a finitely generated free $F[t]$-module. It turns out that the theory about rings of algebraic integers and $F[t]_K$ is very similar when F is a finite field. This case can be viewed as the geometric analogue of the arithmetic case of the ring of algebraic integers. Although one must worry about separability, it is easier in general. A field that is either a finite extension of the rationals or a finite separable extension of $F(t)$ with F a finite field is called a *global field*. Since the expectation that study of the rings $\mathbb{Z}_K$ and $F[t]_K$ in the appropriate global field K are similar, this often leads to the search and study for the appropriate analogue. For example, the solution of the analogue of the Riemann Hypothesis (the Weil Conjecture) for the geometric case by Deligne is one of the major theorems in mathematics proven in the twentieth century.
3. In general, if A is a Dedekind domain with quotient field F and K/F a finite separable field extension, then A_K is not a free A-module.
4. Let A be a domain with quotient field F and K/F a finite field extension. In general, A_K will not be a finitely generated A-module. However, the Krull-Akizuki Theorem 96.17 below says if A is a Noetherian domain of dimension at most one, then B has the same properties for any domain B satisfying $A \subset B \subset K$. In particular, if A is a Dedekind domain, then A_K is also a Dedekind domain. In particular, if $K/F(t)$ is a finite extension, then $F[t]_K$ is a Dedekind domain. Therefore, any finite extension of $F(t)$ is also called a global field.
5. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. If $\mathfrak{p}$ is a maximal ideal in A, then $(A_\mathfrak{p})_K = (A_K)_\mathfrak{p}$, the localization at $A \setminus \mathfrak{p}$ by Exercise 79.15(5). In particular, $(A_\mathfrak{p})_K$ is a Dedekind domain. Although $A_\mathfrak{p}$ is a discrete valuation ring, $(A_\mathfrak{p})_K$ may not be. However, by Exercise 81.14 (6), it does have only finitely many maximal ideals and both it and $A_\mathfrak{p}$ are PIDs.

We next prove the defining property of Dedekind domains that also achieves our desired generalization of a UFD.

Theorem 81.9. *Let A be a Dedekind domain. Then every nonzero nonunit ideal in A is a product of nonzero prime ideals, unique up to order.*

Proof. Let $\mathfrak{A} < A$ be a nonzero ideal and F the quotient field of A. We prove the result in a number of steps.

The first step uses the fact that A is Noetherian.

Step 1. There exist nonzero prime ideals (hence maximal ideals) $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ in A, not necessarily distinct, satisfying $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset \mathfrak{A}$ (Cf. Exercise 30.22 (19)):

If this is not true, by the Maximal Principle, we may assume that $\mathfrak{A}$ is a maximal counterexample, i.e., $\mathfrak{A}$ does not satisfy Step 1, but any nonunit ideal properly containing $\mathfrak{A}$ does. Of course, $\mathfrak{A}$ cannot be a prime ideal, so there exist ideals $\mathfrak{B}_i$, $i = 1, 2$ satisfying $\mathfrak{A} < \mathfrak{B}_i < A$ with $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$ by Lemma 26.20. As each $\mathfrak{B}_i$ contains a product of prime ideals by the maximality condition, so does $\mathfrak{A}$, a contradiction.

The second step uses the fact that nonzero prime ideals in the Noetherian domain A are maximal.

Step 2. Let $\mathfrak{p}$ be a nonzero prime ideal in A. Set

$$\mathfrak{p}^{-1} := \{x \in F \mid x\mathfrak{p} \subset A\},$$

an A-submodule of F. Then $A < \mathfrak{p}^{-1}$:

Clearly, $A \subset \mathfrak{p}^{-1} \subset F$ and $\mathfrak{p}^{-1}$ is an A-module. Let a be a nonzero element of $\mathfrak{p}$. If $\mathfrak{p} = (a)$, then a^{-1} lies in $\mathfrak{p}^{-1} \setminus A$, so we may assume not. By Step 1, there exist nonzero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ satisfying $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (a) \subset \mathfrak{p} < A$. Moreover, we may assume that we have chosen these prime ideals so that $r > 1$ is minimal. Since $\mathfrak{p}$ is a prime ideal, $\mathfrak{p}_i \subset \mathfrak{p}$ for some i, and as $\mathfrak{p}_i$ is a maximal ideal, we, in fact, must have $\mathfrak{p}_i = \mathfrak{p}$. Changing notation if necessary, we may assume that $\mathfrak{p} = \mathfrak{p}_1$. By the minimality of r, there exists an element b in $\mathfrak{p}_2 \cdots \mathfrak{p}_r \setminus (a)$. We then have

$$b\mathfrak{p} = b\mathfrak{p}_1 \subset \mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (a).$$

In particular, $a^{-1}b\mathfrak{p}$ lies in A, i.e., $a^{-1}b$ lies in $\mathfrak{p}^{-1}$. By the choice of b, we have $a^{-1}b \notin A$.

The third step uses the fact that the dimension one Noetherian domain A is integrally closed.

Step 3. Let $\mathfrak{p}$ be a nonzero prime ideal in A. Then for every nonzero ideal $\mathfrak{B}$ in A, we have $\mathfrak{B} < \mathfrak{B}\mathfrak{p}^{-1}$. In particular, $\mathfrak{p}\mathfrak{p}^{-1} = A$:

Suppose that $\mathfrak{B} = \mathfrak{B}\mathfrak{p}^{-1}$. By Step 2, there exists an element a in $\mathfrak{p}^{-1} \setminus A$ satisfying $a\mathfrak{B} \subset \mathfrak{B}$. Hence $\lambda_a : \mathfrak{B} \to \mathfrak{B}$ is an A-homomorphism. Therefore, $\mathfrak{B}$ is an $A[a]$-module. As $A[a]$ is a nonzero ring, $\operatorname{ann}_{A[a]} \mathfrak{B} = 0$ (as $\mathfrak{B} \subset F$ is $A[a]$-torsion free). Moreover, $\mathfrak{B}$ is finitely generated, since A is a Noetherian ring. Therefore, a is integral over A by Claim 79.4. As A is integrally closed, we conclude that a lies in A, a contradiction. Finally, if $\mathfrak{B} = \mathfrak{p}$, then $\mathfrak{p} < \mathfrak{p}\mathfrak{p}^{-1}$ forces $\mathfrak{p}\mathfrak{p}^{-1}$ to be A, as $\mathfrak{p}$ is a maximal ideal.

Step 4. Finish:

We first show that $\mathfrak{A}$ is a product of nonzero prime ideals. If not, as in the proof of Step 1, we may assume that $\mathfrak{A}$ is a maximal counterexample. Since $\mathfrak{A}$ cannot be a prime ideal, there exists a prime ideal $\mathfrak{p}$ in A with $\mathfrak{A} < \mathfrak{p}$. In particular, $\mathfrak{A}\mathfrak{p}^{-1} \subset \mathfrak{p}\mathfrak{p}^{-1} = A$. By Step 3 and the maximality condition, there exist nonzero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{A}\mathfrak{p}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$. Applying Step 4 again shows that $\mathfrak{A} = \mathfrak{p}\mathfrak{p}_1 \cdots \mathfrak{p}_r$, a contradiction.

Next we show uniqueness. If we have

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{A} = \mathfrak{P}_1 \cdots \mathfrak{P}_s$$

with all the $\mathfrak{p}_i$ and $\mathfrak{P}_j$ maximal ideals, then as in Step 1, we see that $\mathfrak{p}_i = \mathfrak{P}_j$ for some i and j. Changing notation, we may assume that $1 = i = j$. Multiplying the above equation by $\mathfrak{p}^{-1}$ leads to our conclusion as $\mathfrak{p}_1\mathfrak{p}_1^{-1} = A$ by Step 3. $\square$

The converse of the theorem is true, i.e., a domain in which every nonzero nonunit ideal is a product of prime ideals, unique up to order, is a Dedekind domain. This is left as an exercise.

Let A be a Dedekind domain and $\mathfrak{A}$ a nonzero nonunit ideal in A. By the theorem, there exist unique nonzero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ in A and unique positive integers $e_1, \dots, e_r$ satisfying

$$(*) \qquad \mathfrak{A} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$$

(unique up to order). We call (*) a *factorization* of $\mathfrak{A}$. For each nonzero prime ideal $\mathfrak{p}$ and we define

$$v_{\mathfrak{p}}(\mathfrak{A}) = \begin{cases} e_i & \text{if } \mathfrak{p} = \mathfrak{p}_i, \text{ some } i = 1, \dots, r \\ 0 & \text{otherwise.} \end{cases}$$

Then (*) can be written

$$\mathfrak{A} = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(\mathfrak{A})}.$$

Let $\mathfrak{A}$ and $\mathfrak{B}$ be two nonzero ideals in a Dedekind domain A. Then, by unique factorization of ideals, $\mathfrak{A} \subset \mathfrak{B}$ if and only if $v_{\mathfrak{p}}(\mathfrak{A}) \geq v_{\mathfrak{p}}(\mathfrak{B})$ for all maximal ideals $\mathfrak{p}$ in A. Mimicking the case for elements, we say $\mathfrak{B}$ *divides* $\mathfrak{A}$ and write $\mathfrak{B} \mid \mathfrak{A}$. In particular, $\mathfrak{p}^n \mid \mathfrak{A}$ if and only if $n \leq v_{\mathfrak{p}}(\mathfrak{A})$. We define the *greatest common divisor* of $\mathfrak{A}$ and $\mathfrak{B}$ in A to be the largest ideal $\mathfrak{D}$ of A containing both $\mathfrak{A}$ and $\mathfrak{B}$. As $\mathfrak{A}$ and $\mathfrak{B}$ lie in $\mathfrak{A} + \mathfrak{B}$, it follows that $\mathfrak{A} + \mathfrak{B} = \mathfrak{D} = \prod_{\mathfrak{p}} \mathfrak{p}^{\min\left(v_{\mathfrak{p}}(\mathfrak{A}), v_{\mathfrak{p}}(\mathfrak{B})\right)}$. We say $\mathfrak{A}$ and $\mathfrak{B}$ are *relatively prime* if $\mathfrak{D} = A$, i.e., $A = \mathfrak{A} + \mathfrak{B}$.

We have shown that if A is a Dedekind domain, then any nonzero prime ideal $\mathfrak{p}$ has an 'inverse', viz., $\mathfrak{p}^{-1}$ as $\mathfrak{p}\mathfrak{p}^{-1} = A$. We next generalize this to a notion of inverses of ideals in Dedekind domains.

Definition 81.10. Let A be a domain with quotient field F. A nonzero A-submodule $\mathfrak{A}$ of F is called a *fractional ideal* of A if there exists a nonzero element x in A satisfying $x\mathfrak{A} \subset A$. If $\mathfrak{A}$ is a fractional ideal, we define

$$\mathfrak{A}^{-1} := \{x \in F \mid x\mathfrak{A} \subset A\}.$$

As this is clearly a submodule of F and $a\mathfrak{A}^{-1} \subset A$ for any a in $\mathfrak{A}$, the module $\mathfrak{A}^{-1}$ is also a fractional ideal. Set

$$I_A := \{\mathfrak{A} \mid \mathfrak{A} \text{ a fractional ideal of } A\}.$$

We say that $\mathfrak{A}$ in I_A is *invertible* if $\mathfrak{A}\mathfrak{A}^{-1} = A$.

As every maximal ideal in a Dedekind domain A is invertible and there exist unique factorization of nonzero nonunit ideals in A, we expect that every fractional ideal in A is invertible. This is in fact true, as we shall now show.

Corollary 81.11. *Let A be a Dedekind domain. Then every fractional ideal of a Dedekind domain A is invertible. Moreover, I_A is a free abelian group on basis the set of maximal ideals in A.*

Proof. Let $\mathfrak{A}$ be a fractional ideal. Then there exists a nonzero element a in A satisfying $a\mathfrak{A} \subset A$. By the theorem, we have factorizations say $a\mathfrak{A} = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ and $(a) = \mathfrak{p}_1^{f_1} \cdots \mathfrak{p}_s^{f_s}$. As nonzero prime ideals are invertible, we have $\mathfrak{A} = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r} \mathfrak{p}_1^{-f_1} \cdots \mathfrak{p}_s^{-f_s}$. Cancelling appropriate $\mathfrak{P}_i$'s and $\mathfrak{p}_j$'s gives a factorization that is clearly unique. It follows that I_A is a free abelian group on the set of maximal ideals with unity A. □

Of course, if A is a Dedekind domain and $\mathfrak{A}$ a fractional ideal, we have $\mathfrak{A} = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(\mathfrak{A})}$, where $v_{\mathfrak{p}}(\mathfrak{p}^n) = n$ for any integer n.

Definition 81.12. Let A be a Dedekind domain. Set

$$P_A := \{xA \mid x \in F^\times\}$$

a subgroup of I_A called the *group of principal fractional ideals* of A. The quotient group

$$Cl_A = I_A/P_A \text{ is called the } \textit{class group of } A.$$

It measures the obstruction of A from being a PID. It also measures when A is a UFD as we shall now show.

Corollary 81.13. *Let A be a Dedekind domain. Then A is a UFD if and only if A is a PID if and only if Cl_A is trivial.*

Proof. We know that a domain is a UFD if and only if every nonzero prime ideal contains a prime element by Kaplansky's Theorem 31.1. As A is a Dedekind domain, this is equivalent to every maximal ideal being principal which, by Theorem 81.9, is equivalent to every ideal being principal. □

In general, a Dedekind domain may not be a PID, e.g., $\mathbb{Z}[\sqrt{-5}]$. Also the class group Cl_A may be infinite. In fact, L. Claburn has shown that any abelian group can be realized as the class group of some Dedekind domain. However, an important theorem in number theory is that $Cl_{\mathbb{Z}_K}$ is a finite group for any ring of algebraic integers $\mathbb{Z}_K$. This is usually proven by using what is called Minkowski Theory.

Exercises 81.14.

1. Define the least common multiple of two nonzero nonunit ideals in a Dedekind domain and show that it is equal to the intersection of these ideals.
2. Show that $\mathbb{Z}[\sqrt{-5}]$ is a Dedekind domain but not a PID.
3. Let A be a discrete valuation ring, i.e., a Dedekind domain with a unique maximal ideal. Show that A is a PID.
4. Let A be a Dedekind domain and S a multiple set not containing zero. Show that the localization $S^{-1}A$ is either a field or a Dedekind domain. In particular, if $\mathfrak{p}$ is a maximal ideal in A then the localization of A at $\mathfrak{p}$, i.e., $A_{\mathfrak{p}} = S^{-1}A$ where $S = A \setminus \mathfrak{p}$, is a discrete valuation ring.
5. Show that every Dedekind domain with finitely many prime ideals is a PID.

6. Let A be a Dedekind domain with quotient field F and having finitely many prime ideals. Show if K/F is a finite separable field extension, then A_K is a Dedekind domain with finitely many prime ideals.
7. Let R be an arbitrary domain: Show the following:
   (a) Any invertible fractional ideal in R is finitely generated.
   (b) A product of fractional ideals in R is invertible if and only if each of the fractional ideals is invertible.
   (c) If an ideal in R is a product of invertible prime ideals, then the prime ideals are unique up to order.
8. Let R be a domain in which every nonzero ideal is a product of prime ideals. Show that every nonzero prime ideal is invertible and maximal.
9. Show that a domain, not a field, is a Dedekind domain, if and only if every nonzero ideal is a product of prime ideals.
10. Let A be a Dedekind domain and $\mathfrak{p}$ a prime ideal in A. Show the following:
    (i) $\mathfrak{p}^m = \mathfrak{p}^n$ if and only if $m = n$.
    (ii) If $\pi \in \mathfrak{p} \setminus \mathfrak{p}^2$, then $\mathfrak{p}^m = (\pi)^m + \mathfrak{p}^n$ for any positive integers $n \geq m$.
    (iii) The only proper ideals in $A/\mathfrak{p}^n$ are $\mathfrak{p}/\mathfrak{p}^n, \dots \mathfrak{p}^{n-1}/\mathfrak{p}^n$
11. Let A be a Dedekind domain and $\mathfrak{A}$ a nonzero ideal in A. Using the previous exercise show the following:
    (i) Every ideal in $A/\mathfrak{A}$ is principal.
    (ii) Every ideal in A can be generated by two elements.
12. Let A be a Dedekind domain. Prove the following:
    (i) Let $\mathfrak{A}$ be a fractional ideal of A and $\mathfrak{B} \subset A$ an ideal. Then there exists an element $a \in \mathfrak{A}$ such that $\mathfrak{A}^{-1}a + \mathfrak{B} = A$.
    (ii) Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two fractional ideals of A. Then the A-module $\mathfrak{A}_1 \coprod \mathfrak{A}_2$ is isomorphic to $A \coprod \mathfrak{A}_1\mathfrak{A}_2$.
    (iii) Let $\mathfrak{A}_1$ be a fractional ideal of A. Then there exists an A-isomorphism $\mathfrak{A} \coprod \mathfrak{A}^{-1} \cong A^2$. In particular, every fractional ideal over A is A-projective (cf. Exercises 39.12(126.1), (126.3), (126.5)).
    (iv) Every finitely generated torsion-free A-module is isomorphic to a finite direct sum of ideals in A. (Cf. Exercise 44.24(2).)
    (v) Let M be a finitely generated torsion-free A-module. Then $M \cong A^n \coprod \mathfrak{A}$ for some n and some ideal $\mathfrak{A}$ in A. [There is a uniqueness statement. Can you guess what it is?]
13. Let A be a local noetherian domain of dimension one with $\mathfrak{m}$ its maximal ideal. Let $k := A/\mathfrak{m}$, called the *residue class field* of A. Show the following are equivalent:
    (i) A is a discrete valuation ring.
    (ii) A is integrally closed.
    (iii) $\mathfrak{m}$ is a principal ideal.
    (iv) $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = 1$.
    (v) Every non-zero ideal is a power of $\mathfrak{m}$.
    (vi) There exists $x \in A$ such that every non-zero ideal is of the form (x^n), $n \geq 0$.

(vii) A is a *valuation ring*, i.e., if a nonzero x is an element in the quotient field of A, then either x is an element of A or its multiplicative inverse is.
(viii) A is a Dedekind domain.

14. Let A be a noetherian domain. Then the following are equivalent:
 (i) A is a Dedekind domain.
 (ii) $A_{\mathfrak{m}}$ is a discrete valuation ring for all maximal ideals $\mathfrak{m}$ in A.
 (iii) Every non-zero fractional ideal in A is invertible.
 (iv) Every integral ideal in A factors uniquely into prime ideals.
15. Let A be a Dedekind domain and $\mathfrak{A}$ an nonzero ideal in A with factorization $\mathfrak{A} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ into prime ideals in A. Prove all of the following:
 (i) There exists a ring isomorphism $A/\mathfrak{A} \cong A/\mathfrak{p}_1^{e_1} \times \cdots A/\mathfrak{p}_r^{e_r}$.
 (ii) If M is an A-module, then there exist A-modules M_i for $i = 1, \ldots, r$ and an A-module isomorphism $M \cong \coprod_{i=1}^r M_i$ with $A/\mathfrak{p}_i^{k_j} M_i = 0$ for every $j \neq i$ and each $i = 1, \ldots, r$. In particular, each M_i is an $A/\mathfrak{p}_i^{e_i}$-module.
16. Let A be a Dedekind domain, $\mathfrak{p}$ a nonzero prime ideal in A, and M a finitely generated torsion A-module with *annihilator*, $\operatorname{ann}_A M := \{x \in A \mid xA = 0) = \bigcap_{x\in A} \operatorname{ann}_A x$. Show all of the following:
 (i) Let $A_{\mathfrak{p}}$ be the localization of A at $\mathfrak{p}$, a discrete valuation ring hence a PID. Then $A/\mathfrak{p}^e \cong A_{\mathfrak{p}}/(\mathfrak{p}^e A_{\mathfrak{p}})$ as rings.
 (ii) If M satisfies $\mathfrak{p}^e = \operatorname{ann}_A(M)$, then M is isomorphic to a coproduct of A-modules $A/\mathfrak{p}^k$ for various $1 \le k \le e$.
 (iii) If $\operatorname{ann}_A(M) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, then M is isomorphic to a coproduct of A-modules $A/\mathfrak{p}_i^{k_{ij}}$ for various $1 \le k_{ij} \le e_i$ for $i = 1, \ldots, r$. [There is a uniqueness statement. What is it?]
17. Let A be a Dedekind domain and M a finitely generated A-module. Then $M = M_{tf} \oplus M_t$ with the submodules M_{tf} torsion-free and M_t torsion, respectively. In particular, by (12) and (16) of these Exercises (81.14), $M \cong R^n \coprod \mathfrak{A} \coprod \coprod_{i,j} A/\mathfrak{p}_i^{k_j}$, for some ideal $\mathfrak{A}$ in A and for various $A/\mathfrak{p}_i^{k_{ij}}$, $1 \le k_{ij} \le e_i$, where $\operatorname{ann}_A(M) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$. (There is also a uniqueness statement. What is it?)

82. Extension of Dedekind Domains

Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. We have shown that A_K is also a Dedekind domain. We wish to investigate how nonzero prime ideals in A factor into a product of primes ideals in A_K. We begin by showing any proper ideal in A remains a proper ideal in A_K. This follows from the following (which is a special case of the Cohen-Seidenberg Theorem 93.14(2) below).

Claim 82.1. Let $\mathfrak{p}$ be a nonzero prime ideal in the Dedekind domain A above. Then $\mathfrak{p}A_L < A_L$ for any field extension L/F.

We know that $\mathfrak{p}^2 < \mathfrak{p}$. Let $\pi \in \mathfrak{p} \setminus \mathfrak{p}^2$. It follows that $\pi A = \mathfrak{p}\mathfrak{A}$ for some ideal $\mathfrak{A}$ of A with $\mathfrak{p} \not| \mathfrak{A}$, hence $\mathfrak{p} + \mathfrak{A} = A$. Write $1 = p + a$ with $p \in \mathfrak{p}$ and $a \in \mathfrak{A}$. It follows that $a \notin \mathfrak{p}$ and $a\mathfrak{p} \subset \mathfrak{p}\mathfrak{A} = \pi A$. Suppose that $\mathfrak{p}A_L = A_L$. Then $aA_L = a\mathfrak{p}A_L \subset \pi A_L$, hence $a = \pi x$, for some $x \in A_L \cap F = A$. This implies that a lies in $\mathfrak{p}$, a contradiction. This proves the claim.

In particular, if $\mathfrak{p}$ is a nonzero prime ideal in A, then we have a factorization

$$(*) \qquad \mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r} \text{ in } A_K$$

called the *splitting behavior of* $\mathfrak{p}$ in A_K. We say a prime ideal $\mathfrak{P}$ in A_K *lies over* $\mathfrak{p}$, if $\mathfrak{P} \cap A = \mathfrak{p}$. If $\mathfrak{P}$ is such a prime ideal then $\mathfrak{p}A_K \subset (\mathfrak{P} \cap A)A_K \subset \mathfrak{P} < A_K$, so $\mathfrak{P} = \mathfrak{P}_i$ for some i, as $\mathfrak{P}$ is a maximal ideal in A_K. Conversely if $\mathfrak{P} = \mathfrak{P}_i$, then $\mathfrak{p}A_K \cap A \subset \mathfrak{P} \cap A < A$, so $\mathfrak{P} \cap A = \mathfrak{p}$, as $\mathfrak{p}$ is a maximal ideal. Thus

$$\{\mathfrak{P}_1, \dots, \mathfrak{P}_r\} = \{\mathfrak{P} \mid \text{with } \mathfrak{P} \text{ a prime ideal in } A_K \text{ satisfying } \mathfrak{P} \mid \mathfrak{p}A_K\},$$

the set of prime ideals in A_K lying over $\mathfrak{p}$ in A_K. We shall write $\mathfrak{P} \mid \mathfrak{p}$ for $\mathfrak{P} \mid \mathfrak{p}A_K$. Let $\mathfrak{P}$ be a nonzero prime ideal in A_K. Then the composition of the inclusion $A \subset A_K$ and the canonical epimorphism $^- : A_K \to A_K/\mathfrak{P}$ shows that the field $A_K/\mathfrak{P}$ is a field extension of $A/\mathfrak{P} \cap A$. We define *ramification index* $e(\mathfrak{P}/\mathfrak{p})$ *of* $\mathfrak{P}$ *over* $\mathfrak{p}$ and if $\mathfrak{P}$ lies over $\mathfrak{p}$, the *inertia index* $f(\mathfrak{P}/\mathfrak{p})$ *of* $\mathfrak{P}$ *over* $\mathfrak{p}$ by

$$\begin{aligned} e(\mathfrak{P}/\mathfrak{p}) &= v_{\mathfrak{P}}(\mathfrak{p}A_K) \\ f(\mathfrak{P}/\mathfrak{p}) &= [A_K/\mathfrak{P} : A/\mathfrak{p}]. \end{aligned}$$

respectively. If $e(\mathfrak{P}/\mathfrak{p}) > 1$ or if $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is not separable, we say that $\mathfrak{p}$ *ramifies* in A_K. That the inertia index is a finite number follows from the following important result:

Theorem 82.2. *Let A be a Dedekind domain with quotient field F, K/F a finite separable field extension. If $\mathfrak{p}$ is a nonzero prime ideal in A, then $f(\mathfrak{P}/\mathfrak{p})$ is finite for all $\mathfrak{P} \mid \mathfrak{p}$ and*

$$[K : F] = \sum_{\mathfrak{P} \mid \mathfrak{p}} e(\mathfrak{P}/\mathfrak{p}) f(\mathfrak{P}/\mathfrak{p}).$$

PROOF. Let $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ be a factorization of $\mathfrak{p}$ in A_K. So $e(\mathfrak{P}_i/\mathfrak{p}) = e_i$ and $f(\mathfrak{P}_i/\mathfrak{p}) = f_i$ for each i. Let $n = [K : F]$ and $^- : A_K \to A_K/\mathfrak{p}A_K$ be the canonical epimorphism, so $\overline{A} = A/\mathfrak{p}$. By the Chinese Remainder Theorem

$$(*) \qquad A_K/\mathfrak{p}A_K = A_K/\mathfrak{P}_1^{e_1} \times \cdots \times A_K/\mathfrak{P}_r^{e_r}.$$

The ring $\overline{A_K} = A_K/\mathfrak{p}A_K$ is a finite dimensional $\overline{A}$-vector space as A_K is a finitely generated A-module, hence so are the $A_K/\mathfrak{P}_i^{e_i}$'s. It follows from this that the $f(\mathfrak{P}_i/\mathfrak{p})$ are all finite.

Step 1. $n = [K : F] = \dim_{\overline{A}} \overline{A_K}$:

Choose $\mathcal{B} = \{x_1, \dots, x_m\}$ in A_K, so that $\overline{\mathcal{B}} = \{\overline{x}_1, \dots, \overline{x}_m\}$ is an $\overline{A}$-basis for $\overline{A_K}$.

We shall show that $\mathcal{B}$ is an F-basis for K establishing the claim. We first show $\mathcal{B}$ is linearly independent. Suppose not, then clearing denominators if necessary, we see that we have an equation

$$a_1x_1 + \cdots + a_mx_m = 0 \text{ in } A_K$$

for some $a_1, \dots, a_m$ in A, not all zero. Let $\mathfrak{A}$ be the nonzero ideal $(a_1, \dots, a_m)$. As $\mathfrak{A}\mathfrak{A}^{-1} = A$, we can choose a nonzero element a in $\mathfrak{A}^{-1}$ such that not every aa_i lies in $\mathfrak{p}$. As $aa_i \in A$ for every i, we have

$$\overline{aa}_1\overline{x}_1 + \cdots + \overline{aa}_m\overline{x}_m = 0 \text{ in } \overline{A_K}$$

contradicting $\overline{\mathcal{B}}$ is $\overline{A}$-linearly independent. Thus $\mathcal{B}$ is linearly independent.

We next show $\mathcal{B}$ spans. Let $B = Ax_1 + \cdots + Ax_m$, a submodule of A_K. As $\overline{\mathcal{B}}$ is a basis for $\overline{A_K}$, if $b \in B$, there exists an $x \in A_K$ satisfying $x - b$ lies in $\mathfrak{p}A_K$, i.e., $A_K = B + \mathfrak{p}A_K$. Let $M = A_K/B$, a finitely generated A-module. Then $M/\mathfrak{p}M = 0$, since each x_i maps to zero in it so a zero dimensional $\overline{A}$-vector space. Consequently, $M = \mathfrak{p}M$. Set $M = Ay_1 + \cdots + Ay_s$. Then we have equations

$$y_i = \sum_{j=1}^{s} p_{ij}y_j \text{ for some } p_{ij} \in \mathfrak{p},\ 1 \leq i, j \leq s,$$

hence

$$\sum_{j=1}^{s}(\delta_{ij} - p_{ij})y_j = 0 \text{ for } i = 1, \dots s.$$

Let Δ be the determinant of the $s \times s$ matrix $(\delta_{ij} - p_{ij})$. Then $\Delta y_i = 0$ for every i by Cramer's Rule, so $\Delta M = 0$. It follows that $\Delta A_K \subset B$, hence also that $\Delta K \subset Fx_1 + \cdots Fx_m$. But $\Delta \equiv 1 \mod \mathfrak{p}$, so Δ is nonzero. Therefore, we conclude that

$$K = \Delta K = Fx_1 + \cdots + Fx_m.$$

Therefore, $\mathfrak{B}$ is an F-basis for K and $m = n$. This completes Step 1.

Step 2. Finish:

In view of (*), it suffices to show that $\dim_{\overline{A}} A_K/{\mathfrak{P}_i}^{e_i} = e_if_i$ for each i, $1 \leq i \leq r$.

We have a descending chain of $\overline{A}$-vector spaces

$$A_K/\mathfrak{P}_i^{e_i} \supset \mathfrak{P}_i/\mathfrak{P}_i^{e_i} \supset \cdots \supset \mathfrak{P}_i^{e_i-1}/\mathfrak{P}_i^{e_i} \supset 0,$$

so

$$\dim_{\overline{A}} A_K/\mathfrak{P}_i^{e_i} = \sum_{m=0}^{e_i-1} \dim_{\overline{A}} \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}.$$

(Why?) As $\mathfrak{P}_i^{m+1} < \mathfrak{P}_i^m$ by unique factorization of ideals, we can choose an element a in $\mathfrak{P}_i^m \setminus \mathfrak{P}_i^{m+1}$ for each m. Then $\mathfrak{P}_i^m = aA_K + \mathfrak{P}_i^{m+1}$, the greatest common divisor of $\mathfrak{P}_i^{m+1}$ and aA_K and $\mathfrak{P}_i^{m+1} = aA_K \cap \mathfrak{P}_i^{m+1}$ (the least common multiple of $\mathfrak{P}_i^{m+1}$ and aA_K, cf. Exercise 81.14(1)). Let $\lambda_a : A_K \to \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}$ be the homomorphism defined by $x \mapsto ax + \mathfrak{P}_i^{m+1}$. The kernel of λ_a is $\mathfrak{P}_i$ and λ_a is onto, as

$$\operatorname{im} \lambda_a = (aA_K + \mathfrak{P}_i^{m+1})/\mathfrak{P}_i^{m+1} = \mathfrak{P}_i^m/(aA_K \cap \mathfrak{P}_i^{m+1}) = \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}.$$

Therefore $f_i = \dim_{\overline{A}} A_K/\mathfrak{P}_i = \dim_{\overline{A}} \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}$ for any $m > 0$ and $\dim_{\overline{A}} A_K/\mathfrak{P}_i^{e_i} = e_if_i$ as desired. □

Remark 82.3. If $\mathfrak{P} \mid \mathfrak{p}$ in the theorem and $(A_K/\mathfrak{P})_{sep}$ is the separable closure of $A/\mathfrak{p}$ in $A_K/\mathfrak{P}$, then $f(\mathfrak{P}/\mathfrak{p}) = [(A_K/\mathfrak{P}/\mathfrak{p})_{sep} : A/\mathfrak{p}]\, p^s$ for some s where $p = \operatorname{char}(A/\mathfrak{p})$. In particular, if $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is normal, then $f(\mathfrak{P}/\mathfrak{p}) = |G\big(A_K/\mathfrak{P})/(A/\mathfrak{p})\big)|\, p^s$, so if $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is also separable, then $f(\mathfrak{P}/\mathfrak{p}) = |G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)|$.

Remark 82.4. If $A = \mathbb{Z}$ in the theorem, Step 1 can be simplified. Indeed let $\{w_1, \dots, w_n\}$ be an integral basis for $\mathbb{Z}_K$ and $\mathfrak{A}$ any nonzero ideal in $\mathbb{Z}_K$. We know that $\mathfrak{A} \cap \mathbb{Z} = (a)$ for some nonzero integer a. Set $S = \{\sum r_i w_i \mid 0 \leq r_i < a\}$. Then using the division algorithm, one shows that S is a system of representatives for the cosets $\mathbb{Z}_K/\mathfrak{A}$, so has a^n elements.

Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. Let $\mathfrak{p}$ be a nonzero prime ideal in A and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. We say that $\mathfrak{P}$ is *unramified* over A if $e(\mathfrak{P}/\mathfrak{p}) = 1$ and $A_k/\mathfrak{P}$ is a separable extension of $A/\mathfrak{p}$ and *ramified* otherwise (and *totally ramified* if in addition $f(\mathfrak{P}/\mathfrak{p}) = 1$). We say $\mathfrak{p}$ is *unramified* in A_K if every prime ideal in A_K lying over $\mathfrak{p}$ is unramified and the extension K/F (or A_K/A) is called *unramified* if every prime ideal in A is unramified in A_K. We say a nonzero prime ideal $\mathfrak{p}$ in A *splits completely* in A_K if $e(\mathfrak{P}/\mathfrak{p}) = 1 = f(\mathfrak{P}/\mathfrak{p})$ for every prime ideal $\mathfrak{P}$ lying over $\mathfrak{p}$. If K is a global field, with $A = \mathbb{Z}$ or $F[t]$, we do not have to worry about the separability condition of $A_K/\mathfrak{P}$ over $A/\mathfrak{p}$ as either $A_K/\mathfrak{P}$ is of characteristic zero or a finite field. A deep result in number theory states that there exist infinitely many nonzero prime ideals in A that split completely in A_K when K is a global field.

The Kummer-Dedekind Theorem gives a method for computing the splitting behavior of prime ideals for all but finitely many prime ideals in a given separable extension. We first define the set of prime ideals whose splitting behavior this theorem omits. Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. By the Primitive Element Theorem 57.9, $K = F(\alpha)$ for some α in K. We may assume that α lies in A_K. So $A[\alpha]$ is a subring of A_K. Define the *conductor* of $A[\alpha]$ in A_K to be the largest ideal $\mathfrak{f}$ of A_K that lies in $A[\alpha]$, i.e.,

$$\mathfrak{f} = \{x \in A_K \mid xA_K \subset A[\alpha]\}.$$

We know that A_K contains the F-basis $\{1, \alpha_1, \dots, \alpha^{n-1}\}$ for K, so it follows by Claim 80.6 that $\mathfrak{f} \neq 0$. Note that $\mathfrak{f} \subset A[\alpha]$ as 1 lies in A_K, so the conductor $\mathfrak{f}$ is an ideal of both A_K and $A[\alpha]$.

Kummer-Dedekind's result is the following:

Theorem 82.5. (Kummer-Dedekind) *Let A be a Dedekind domain with quotient field F and $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite separable field extension with $K = F(\alpha)$, $\alpha \in A_K$, and $\mathfrak{p}$ relatively prime to the conductor $\mathfrak{f}$ of $A[\alpha]$ in A_K. Let $^- : A[t] \to (A/\mathfrak{p})[t]$ be the canonical epimorphism and let $p_1, \dots, p_r$ be distinct monic polynomials in $A[t]$ such that*

$$\overline{m_F(\alpha)} = \overline{p_1}^{e_1} \cdots \overline{p_r}^{e_r}$$

is a factorization of the image of the minimal polynomial of α in $(A/\mathfrak{p})[t]$ into irreducibles. Set

$$\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K \ \ \textit{for each} \ \ i = 1, \dots, r.$$

Then $\mathfrak{P}_1, \dots, \mathfrak{P}_r$ are all of the distinct primes in A_K lying over $\mathfrak{p}$, $f(\mathfrak{P}_i/\mathfrak{p}) = f_i$, and $e_i = e(\mathfrak{P}_i/\mathfrak{p})$ for $i = 1, \dots, r$. In particular,

$$\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$$

is a factorization of $\mathfrak{p}$ in A_K.

PROOF. **Step 1**. $A_K/\mathfrak{p}A_K \cong A[\alpha]/\mathfrak{p}A[\alpha]$:

Since $\mathfrak{p}A_K$ and $\mathfrak{f}$ are relatively prime, $A_K = \mathfrak{p}A_K + \mathfrak{f}$. But $\mathfrak{f} \subset A[\alpha]$, so we have $A_K = \mathfrak{p}A_K + A[\alpha]$ and the natural map $A[\alpha] \to A_K/\mathfrak{p}A_K$ is an epimorphism with kernel $\mathfrak{p}A_K \cap A[\alpha]$. As $\mathfrak{f} \cap A$ and $\mathfrak{p}$ must also be relatively prime, we have $\mathfrak{p}A[\alpha] + \mathfrak{f} = A[\alpha]$, so

$$\mathfrak{p}A_K \cap A[\alpha] = (\mathfrak{p}A[\alpha] + \mathfrak{f})(\mathfrak{p}A_K \cap A[\alpha]) \subset \mathfrak{p}A[\alpha].$$

It follows that $\mathfrak{p}A_K \cap A[\alpha] = \mathfrak{p}A[\alpha]$ establishing Step 1.

Step 2. $\overline{A}[t]/(\overline{m_F(\alpha)}) \cong A[\alpha]/\mathfrak{p}A[\alpha]$:

The kernel of the natural epimorphism $A[t] \to \overline{A}[t]/(\overline{m_F(\alpha)})$ is generated by $\mathfrak{p}$ and $(m_F(\alpha))$, Therefore, the evaluation map $A[t] \to A[\alpha]$ induces the desired isomorphism as $A[t]/\mathfrak{p}A[t] \cong (A/\mathfrak{p})[t]$.

Step 3. Finish:

Let $B = \overline{A}[t]/(\overline{m_F(\alpha)})$ and $\widetilde{} : \overline{A}[t] \to B$ be the canonical epimorphism. By the Chinese Remainder Theorem,

$$B \cong \overline{A}[t]/(\overline{p_1}^{e_1}) \times \cdots \times \overline{A}[t]/(\overline{p_r}^{e_r}).$$

It follows that the nonzero prime ideals in B are the distinct principal ideals $(\widetilde{p_i})$ generated by the $p_i \mod \overline{m_F(\alpha)}$, i.e., the (p_i). We have $\deg p_i = \deg \overline{p_i} = [B/(\widetilde{p_i}) : \overline{A}]$ for $i = 1, \ldots, r$. Further,

$$0 = \big(\widetilde{\overline{m_F(\alpha)}}\big) = \bigcap_{i=1}^{r} (\widetilde{p_i}^{e_i}) \text{ in } B.$$

By Steps 1 and 2, we know that $B \cong A_K/\mathfrak{p}A_K$, with the isomorphism induced by evaluation at α, so by the Correspondence Principle, the prime ideals in B correspond to the prime ideals in $A_K/\mathfrak{p}A_K$. Let $\mathfrak{Q}_i$ in $A_K/\mathfrak{p}A_K$ be the prime ideal corresponding to the prime ideal $(\widetilde{p_i})$ in B for $i = 1, \ldots, r$. It follows that each $\mathfrak{Q}_i$ is a principal ideal generated by $\overline{p_i(\alpha)}$ with $\deg \overline{p_i} = [A_K/\mathfrak{Q}_i : A/\mathfrak{p}]$ for each i, and furthermore, $0 = \bigcap_{i=1}^{r} \mathfrak{Q}_i^{e_i}$. Let $\mathfrak{P}_i$ be the ideal in A_K that is the preimage of $\mathfrak{Q}_i$ under the natural epimorphism $A_K \to A_K/\mathfrak{p}A_K$. Then $\mathfrak{P}_i$ is a nonzero prime ideal by Lemma 81.1 and satisfies $\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K$ for $i = 1, \ldots, r$. It follows that $\mathfrak{P}_1, \ldots, \mathfrak{P}_r$ are all the prime ideals in A_K lying over $\mathfrak{p}$, and $f_i = [A_K/\mathfrak{P}_i : A/\mathfrak{p}] = \deg \overline{p_i} = \deg p_i$ for each i. Moreover, as

$$e_i = |\{\mathfrak{Q}_i^m \mid m \geq 0\}|, \text{ we have } \overline{\mathfrak{P}_i}^{e_i} = \mathfrak{Q}_i^{e_i},$$

so $\mathfrak{P}_i^{e_i}$ is the preimage of $\mathfrak{Q}_i^{e_i}$. Since $\prod_{i=1}^r \mathfrak{P}_i^{e_i} \subset \bigcap_{i=1}^r \mathfrak{P}_i^{e_i} \subset \mathfrak{p}A_K$, we have $v_{\mathfrak{p}}(\mathfrak{P}_i) \leq e_i$. We have $[K : F] = \deg m_F(\alpha) = \sum_{i=1}^r e_i \deg p_i = \sum_{i=1}^r e_i f_i$. If $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1'} \cdots \mathfrak{P}_r^{e_r'}$ is a factorization of $\mathfrak{p}A_K$, then $\sum_{i=1}^r e_i f_i = [K : F] = \sum_{i=1}^r e_i' f_i$ by Theorem 82.2. Hence $e_i = e_i'$ for $i = 1, \ldots, r$, and $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ is a factorization of $\mathfrak{p}A_K$. □

Examples 82.6. Let $K = \mathbb{Q}(\alpha)$ with $\alpha^3 = 2$. It can be shown that $\mathbb{Z}_K = \mathbb{Q}[\alpha]$. Then $t - 2 = (t-3)(t^2 + 3t - 1)$ in $\mathbb{Z}/5\mathbb{Z}[t]$. As $t^2 + 3t - 1$ is irreducible in $\mathbb{Z}/5\mathbb{Z}[t]$, then by the Kummer-Dedekind Theorem 82.5, we see that $5\mathbb{Z}_K$ splits completely, say as $5Z_K = \mathfrak{p}_1\mathfrak{p}_2$ with $f(\mathfrak{p}_1/5\mathbb{Z}_K) = 1$ and $f(\mathfrak{p}_2/5\mathbb{Z}_K) = 2$.

Exercises 82.7.

1. Complete the proof of Remark 82.4.

2. Let A be a Dedekind domain with quotient field F and $L/K/F$ finite separable extensions. If $\mathfrak{P}$ is a nonzero prime ideal in A_L, show
$$e(\mathfrak{P}/\mathfrak{P} \cap A) = e(\mathfrak{P}/\mathfrak{P} \cap A_K)e(\mathfrak{P} \cap A_K/\mathfrak{P} \cap A)$$
$$f(\mathfrak{P}/\mathfrak{P} \cap A) = f(\mathfrak{P}/\mathfrak{P} \cap A_K)f(\mathfrak{P} \cap A_K/\mathfrak{P} \cap A).$$
3. (Nakayama's Lemma) Let A be a *local ring*, i.e., a nonzero commutative ring with a unique maximal ideal, say $\mathfrak{m}$. Show if M is a finitely generated A-module satisfying $\mathfrak{m}M = M$, then $M = 0$.
4. Let A be a local ring with maximal ideal $\mathfrak{m}$ and $^- : M \to M/\mathfrak{m}M$ the canonical epimorphism. Let $\mathcal{C} = \{x_1, \dots, x_n\} \subset M$ and $\overline{\mathcal{C}} = \{\overline{x_1}, \dots, \overline{x_n}\}$. Show that $\mathcal{C}$ generates M if and only if $\overline{\mathcal{C}}$ spans $\overline{M}$ as an $A/\mathfrak{m}$-vector space and is a minimal generating set for M (obvious definition) if and only if $\overline{\mathcal{C}}$ is an $A/\mathfrak{m}$-basis for $\overline{M}$.
5. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. If α is an element of A_K, then for each maximal ideal not dividing the conductor $\mathfrak{f}$ of $A[\alpha]$ (hence all but finitely many maximal ideals) $(A_\mathfrak{p})_K = A_\mathfrak{p}[\alpha]$, where $A_\mathfrak{p}$ is the localization of A at $A \setminus \mathfrak{p}$.
6. Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. Then a prime ideal $\mathfrak{p}$ in A ramifies in A_K if and only if the localization $A_\mathfrak{p}$ of A ramifies in $(A_\mathfrak{p})_K$.

83. Hilbert Ramification Theory

Theorem 82.2 becomes even nicer when K/F is Galois. We need the following to prove this.

Proposition 83.1. *Let A be a Dedekind domain with quotient field F and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension, then the Galois group $G(K/F)$ acts transitively on the set of prime ideals in A_K lying over $\mathfrak{p}$.*

Proof. Let $\mathfrak{P}_i \mid \mathfrak{p}$ be prime ideals in A_K for $i = 1, 2$. Suppose that $\mathfrak{P}_2 \neq \sigma(\mathfrak{P}_1)$ for any $\sigma \in G(K/F)$. By the Chinese Remainder Theorem, there exists an element x in A_K satisfying $x \equiv 0 \mod \mathfrak{P}_2$ and $x \equiv 1 \mod \sigma(\mathfrak{P}_1)$ for all $\sigma \in G(K/F)$, i.e., $x \in \mathfrak{P}_2$ but $\sigma^{-1}(x) \notin \mathfrak{P}_1$ for any σ in the group $G(K/F)$. Since
$$N_{K/F}(x) = \prod_{G(K/F)} \sigma(x) \text{ lies in } \mathfrak{P}_2 \cap A_F = \mathfrak{p} = \mathfrak{P}_1 \cap A \subset \mathfrak{P}_1,$$
we have $\sigma(x) \in \mathfrak{P}_1$ for some $\sigma \in G(K/F)$, a contradiction. □

Proposition 83.2. *Let A be a Dedekind domain with F its quotient field and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension, then $e(\mathfrak{P}/\mathfrak{p}) = e(\mathfrak{P}'/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{p}) = f(\mathfrak{P}'/\mathfrak{p})$ for all prime ideals $\mathfrak{P}$ and $\mathfrak{P}'$ lying over $\mathfrak{p}$ in A_K. In particular, if $\mathfrak{P} \mid \mathfrak{p}$ in A_K is a prime ideal and $e = e(\mathfrak{P}/\mathfrak{p})$, $f = f(\mathfrak{P}/\mathfrak{p})$, then $[K : F] = efr$ where r is the number of primes in A_K lying over $\mathfrak{p}$.*

Proof. By the last proposition, there exists an element σ in $G(K/F)$ satisfying $\mathfrak{P}' = \sigma(\mathfrak{P})$, hence we have an isomorphism $A_K/\mathfrak{P} \to A_K/\sigma(\mathfrak{P})$ given by $x + \mathfrak{P} \mapsto$

$\sigma(x)+\sigma(\mathfrak{P})$. It follows that $f(\mathfrak{P}/\mathfrak{p}) = f(\mathfrak{P}'/\mathfrak{p})$ Further, if $\mathfrak{P}^m \mid \mathfrak{p}A_K$, then $\sigma(\mathfrak{P})^m \mid \mathfrak{p}A_K$, so $e(\mathfrak{P}/\mathfrak{p}) = e(\mathfrak{P}'/\mathfrak{p})$. The result now follows using Theorem 82.2. □

Proposition 83.2 allows us to use Galois Theory in the subject. This is called Hilbert Ramification Theory. We introduce the subject without going into it deeply (leaving some results as exercises).

For the rest of this section, let A be a Dedekind domain with quotient field F and K/F a finite Galois extension. Let $\mathfrak{p}$ be a nonzero prime ideal in A and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. As $G(K/F)$ acts on the set of prime ideals in A_K lying over $\mathfrak{p}$, we can look at the isotropy subgroup

$$G_{\mathfrak{P}} = \{\sigma \in G(K/F) \mid \sigma(\mathfrak{P}) = \mathfrak{P}\}$$

of $\mathfrak{P}$. It is called the *decomposition group* of $\mathfrak{P}$ and its fixed field

$$Z_{\mathfrak{P}} = \{x \in K \mid \sigma(x) = x \text{ for all } \sigma \in G_{\mathfrak{P}}\}$$

is called the *decomposition field* of $\mathfrak{P}$. As usual, we have $\sigma G_{\mathfrak{P}}\sigma^{-1} = G_{\sigma(\mathfrak{P})}$. We also have the index $[G : G_{\mathfrak{P}}]$ = number of primes in A_K lying over $\mathfrak{p}$ as $G(K/F)$ acts transitively on the primes in A_K lying over $\mathfrak{p}$. This means that $\mathfrak{p}$ splits completely in A_K if and only if $G_{\mathfrak{P}} = 1$ and $G(K/F) = G_{\mathfrak{P}}$ if and only if $\mathfrak{P}$ is the only prime ideal in A_K lying over $\mathfrak{p}$.

Proposition 83.3. *Let A be a Dedekind domain with quotient field F and K/F a finite Galois extension, $\mathfrak{P}$ a prime ideal in A_K with $\mathfrak{P}$ lying over $\mathfrak{p}$. Then*

(1) $\mathfrak{P}$ is the only prime ideal over $\mathfrak{P} \cap Z_{\mathfrak{P}}$.
(2) $e(\mathfrak{P}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = e(\mathfrak{P}/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = f(\mathfrak{P}/\mathfrak{p})$.
(3) $e(\mathfrak{P} \cap Z_{\mathfrak{P}}/\mathfrak{p}) = 1$ and $f(\mathfrak{P} \cap Z_{\mathfrak{P}}/\mathfrak{p}) = 1$.

PROOF. By Proposition 83.2, we have $[K : Z_{\mathfrak{P}}] = e'f'r'$ where e' is the ramification index of each prime ideal in A_K lying over $\mathfrak{P} \cap Z_{\mathfrak{P}}$, f' its inertia index, and r' is the number of prime ideals in A_K lying over $\mathfrak{P} \cap Z_{\mathfrak{P}}$. Similarly, we have $[Z_{\mathfrak{P}} : F] = e''f''r''$ where e'' is the ramification index of each prime ideal in $Z_{\mathfrak{P}}$ over $\mathfrak{p}$, f'' its inertia index, and r'' is the number of prime ideals in $A_{Z_{\mathfrak{P}}}$ lying over $\mathfrak{p}$, and $[K : F] = efr$ where e is the ramification index of each prime ideal in A_K lying over $\mathfrak{p}$, f is its inertia index, and $r = [G(K/F) : G_{\mathfrak{P}}]$ is the number of prime ideals in A_K lying over $\mathfrak{p}$. By the transitivity of ramification and inertia indices (cf. Exercise 82.7(2), we have $e = e'e''$, $f'f'' = f$, hence $r = r'r''$.

(1): By definition, $G_{\mathfrak{P}} = G(K/Z_{\mathfrak{P}})$, so the primes lying over $\mathfrak{P}_{Z_{\mathfrak{P}}}$ are $\sigma(\mathfrak{P})_{Z_{\mathfrak{P}}}$ by Proposition 83.1, each is $\mathfrak{P}$ and $r' = 1$.

(2), (3): As $r = [G(K/F) : G_{\mathfrak{P}}] = [Z_{\mathfrak{P}} : F] = e''f''r''$, we have $[K : Z_{\mathfrak{P}}] = |G_{\mathfrak{P}}| = ef$. By (1), we have, $e'e''f'f'' = ef = [K : Z_{\mathfrak{P}}] = e'f'$. Thus $e'' = 1 = f''$ and $e' = e$ and $f' = f$ □

We next relate the decomposition group $G_{\mathfrak{P}}$ of $\mathfrak{P}$ in A_K lying over $\mathfrak{p}$ and the Galois group of the field extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$. Let σ be an element of the $G_{\mathfrak{P}}$. Then σ induces an $(A/\mathfrak{p})$-isomorphism

$$\overline{\sigma} : A_K/\mathfrak{P} \to A_K/\mathfrak{P} \text{ given by } x + \mathfrak{P} \mapsto \sigma(x) + \sigma(\mathfrak{P}) = \sigma(x) + \mathfrak{P}.$$

Therefore, we get a group homomorphism

$$\text{(83.4)} \qquad {}^{-} : G_{\mathfrak{P}} \to G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big) \text{ given by } \sigma \mapsto \overline{\sigma}.$$

Proposition 83.5. *Let A be a Dedekind domain with F its quotient field and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension and $\mathfrak{P}$ a prime in A_K lying over $\mathfrak{p}$, then $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is a finite normal extension and the map ${}^{-} : G_{\mathfrak{P}} \to G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$ in (83.4) is surjective.*

Proof. Let ${}^{-} : A_K[t] \to (A_K/\mathfrak{P})[t]$ be the natural epimorphism. Since $f(\mathfrak{P} \cap A_{Z_{\mathfrak{P}}}/\mathfrak{p}) = 1$, we have $A/\mathfrak{p} = A_{Z_{\mathfrak{P}}}/(\mathfrak{P} \cap A_{Z_{\mathfrak{P}}})$, so we may assume that $F = Z_{\mathfrak{P}}$, hence $G(K/F) = G_{\mathfrak{P}}$. Let $x \in A_K/\mathfrak{P}$ and choose α in A_K satisfying $\overline{\alpha} = x$. Then we have

$$(*) \qquad m_{A/\mathfrak{p}}(x) \mid \overline{m_F(\alpha)} \text{ in } (A/\mathfrak{p})[t].$$

Since K/F is normal, $m_K(\alpha)$ splits over K, hence $\overline{m_K(\alpha)}$ splits over $A_K/\mathfrak{P}$. It follows that $m_{A/\mathfrak{p}}(x)$ splits over $A_K/\mathfrak{P}$, hence the finite extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is normal.

Next, using the Primitive Element Theorem 57.9, we can choose α in K such that $(A/\mathfrak{p})(\overline{\alpha})$ is the maximal separable extension of $A/\mathfrak{p}$ in $A_K/\mathfrak{P}$. We may assume that α lies in A_K. Suppose that $\overline{\sigma}$ lies in $G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. Then $\overline{\sigma}$ is completely determined by $\overline{\sigma}(\overline{\alpha})$. As $\overline{\sigma}(\overline{\alpha})$ is a root of $m_{A/\mathfrak{p}}(\overline{\alpha})$, it is a root of $\overline{m_F(\alpha)}$ by (*). Choose a root α' of $m_F(\alpha)$ satisfying $\overline{\alpha'} = \overline{\sigma}(\overline{\alpha})$. There exists an element σ in $G(K/F)$ satisfying $\sigma(\alpha) = \alpha'$ as $G(K/F)$ acts transitively on the roots of $m_F(\alpha)$. Then $\sigma \mapsto \overline{\sigma}$. □

The kernel of (83.4) is denoted by

$$I_{\mathfrak{P}} = \{\sigma \in G_{\mathfrak{P}} \mid \overline{\sigma} = 1_{A_K/\mathfrak{P}}\}$$

and called the *inertia group* of $\mathfrak{P}$, its fixed field

$$T_{\mathfrak{P}} = K^{I_{\mathfrak{P}}}$$

is called the *inertia field* of $\mathfrak{P}$. By Proposition 83.5, we have an isomorphism $G_{\mathfrak{P}}/I_{\mathfrak{P}} \cong G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. We know that $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is normal, but it is not Galois in general. However, if K is a global field, it is as $A_K/\mathfrak{P}$ is then a finite field so separable over $(A/\mathfrak{p})$.

If the extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is separable, then $I_{\mathfrak{P}} = 1$ if and only if $\mathfrak{p}$ is unramified and $G_{\mathfrak{P}} \cong G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. More generally, in this case of separability, we have $e(\mathfrak{P}/\mathfrak{p}) = |I_{\mathfrak{P}}| = [K : T_{\mathfrak{P}}]$ and $f(\mathfrak{P}/\mathfrak{p}) = [G_{\mathfrak{P}} : I_{\mathfrak{P}}] = [T_{\mathfrak{P}} : Z_{\mathfrak{P}}]$.

Suppose that K is a global field, so $F = \mathbb{Q}$ and $A = \mathbb{Z}_K$ or F a finite field and $A = F[t]_K$. Of course, if $F = \mathbb{Q}$, then $\mathfrak{p} = (p)$ in $\mathbb{Z}$ for some prime element p. Suppose that $\mathfrak{P}$ is unramified over $\mathfrak{p}$, then there exists a unique element $\sigma_{\mathfrak{P}}$ in $G(K/F)$ satisfying

$$\sigma_{\mathfrak{P}}(x) \equiv x^{f(\mathfrak{P}/\mathfrak{p})} \mod \mathfrak{P}.$$

The automorphism $\sigma_{\mathfrak{P}}$ is called the *Frobenius automorphism*. If K/F is abelian, then $G_{\tau(\mathfrak{P})} = \tau G_{\mathfrak{P}} \tau^{-1} = G_{\mathfrak{P}}$ for all τ in $G(K/F)$. As $G(K/F)$ acts transitively on the primes lying over $\mathfrak{p}$ all decomposition groups over $\mathfrak{p}$ are equal, so the Frobenius homomorphism $\sigma_{\mathfrak{P}}$ depends only on $\mathfrak{p}$. This was the first case, called 'class field theory', thoroughly studied in the first third of the twentieth century.

Exercises 83.6.

1. Let A be a Dedekind domain with quotient field F, $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite Galois extension and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. Show that $T_{\mathfrak{P}}/Z_{\mathfrak{P}}$ is a normal extension satisfying $G(T_{\mathfrak{P}}/Z_{\mathfrak{P}}) \cong G\big(A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$ and $G(K/T_{\mathfrak{P}}) \cong I_{\mathfrak{P}}$.
2. Let A be a Dedekind domain with quotient field F, $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite Galois extension and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$ and satisfying $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is separable. Assuming the previous exercise and Exercise 82.7(2) show
   - (*i*) $e(\mathfrak{P}/\mathfrak{p}) = |I_{\mathfrak{P}}| = [K : T_{\mathfrak{P}}]$ and $f(\mathfrak{P}/\mathfrak{p}) = [G_{\mathfrak{P}} : I_{\mathfrak{P}}] = [T_{\mathfrak{P}} : Z_{\mathfrak{P}}]$.
   - (*ii*) $e(\mathfrak{P}/\mathfrak{P} \cap T_{\mathfrak{P}}) = e(\mathfrak{P}/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{P} \cap T_{\mathfrak{P}}) = 1$.
   - (*iii*) $e(\mathfrak{P} \cap T_{\mathfrak{P}}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = 1$ and $f(\mathfrak{P} \cap T_{\mathfrak{P}}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = f(\mathfrak{P}/\mathfrak{p})$.

84. The Discriminant of a Number Field

In Section 65, we studied the discriminant of a polynomial f in $F[t]$. If K/F is a finite separable extension, then by the Primitive Element Theorem, $K = F(\alpha)$, for some α. Of course there are many such α's, so there is no unique such discriminant. Suppose that F is the quotient field of a Dedekind domain A. Then we know that we can choose α to lie in A. We now have two problems, α is still not unique, and more seriously, A_K is not necessarily $A[\alpha]$ for any α. We investigate what we must do to solve this problem, as the correct definition is crucial to finding the primes in A that ramify in A_K. Although we shall not prove the full result, this addendum should provide a suitable introduction.

Let K/F be a finite, separable field extension of degree n, L/F be a finite Galois extension with L/K, and $\sigma_1, \dots, \sigma_n : K \to L$ the distinct F-homomorphisms. Let $w_1, \dots, w_n$ be elements in K and define

$$\Delta(w_1, \dots, w_n) = \Delta_{K/F}(w_1, \dots, w_n) := \det\big(\operatorname{Tr}_{K/F}(w_i w_j)\big).$$

Then $\Delta(w_1, \dots, w_n)$ lies in F as each $\operatorname{Tr}(w_i w_j)$ does. Moreover, if every w_i lies in A_K, then $\Delta(w_1, \dots, w_n)$ lies in A. We have (cf. also Section 65):

Properties 84.1. Let K/F be a finite, separable field extension of degree n, L/F a Galois extension with L/K, and $\sigma_1, \dots, \sigma_n : K \to L$ the distinct F-homomorphisms. Let $\mathcal{B} = \{w_1, \dots, w_n\}$

(1) If $K = F(\alpha)$, then $\{1, \alpha, \dots, \alpha^{n-1}\}$ is an F-basis for K. Set $\alpha_i = \sigma_i(\alpha)$. Then

$$\Delta(1, \alpha, \dots, \alpha^{n-1}) = \prod_{i<j} (\alpha_i - \alpha_j)^2$$

is the discriminant of the minimal polynomial $m_F(\alpha)$. We denote it by $\Delta(\alpha)$.

(2) $\Delta(w_1, \dots, w_n)$ is not zero if and only if $\mathcal{B}$ is an F-basis for K.

(3) If $\mathcal{C}$ is an F-basis for K and $\{w_1', \dots, w_n'\}$ its complementary basis, i.e., $\operatorname{Tr}_{K/F}(w_i w_j') = \delta_{ij}$ for all $1 \le i, j \le n$, then

$$\Delta(w_1, \dots, w_n)\Delta(w_1', \dots, w_n') = 1.$$

(4) If $\mathcal{C}$ is an F-basis for K, then
$$\Delta(w_1, \dots, w_n) = \det\big((\sigma_i(w_j))\big)^2.$$

(5) If $\mathcal{B}$ and $\{v_1, \dots, v_n\}$ are two F-bases for K, then
$$\Delta(v_1, \dots, v_n) = (\det C)^2 \Delta(w_1, \dots, w_n)$$
with $C \in \mathrm{GL}_n(F)$, the change of basis matrix.

(6) If $F = \mathbb{Q}$ and $\mathcal{B}$ and $\{v_1, \dots, v_n\}$ are two integral bases for $\mathbb{Z}_K$, then
$$\Delta(v_1, \dots, v_n) = \Delta(w_1, \dots, w_n),$$
as the change of basis matrix for these two integral bases must lie in $\mathrm{GL}_n(\mathbb{Z})$. Denote the integer $\Delta(v_1, \dots, v_n)$, independent of integral basis, by d_K. It is called the *discriminant of* K.

(7) Suppose that $F = \mathbb{Q}$ and $\mathcal{B}$ is an integral basis for $\mathbb{Z}_K$ and $\{v_1, \dots, v_n\}$ a $\mathbb{Q}$-basis for K with each v_i lying in $\mathbb{Z}_K$. Then there exists an integer a such that $a^2 d_K = \Delta(v_1, \dots, v_n)$ in $\mathbb{Z}$; and, in fact,
$$|\Delta(v_1, \dots, v_n)| \geq |d_K|$$
with equality if and only if
$$\{v_1, \dots, v_n\} \text{ is an integral basis.}$$
[For this reason, an integral basis for $\mathbb{Z}_K$ is often called a *minimal basis*.]

In the case of a ring of algebraic integers, we have the following:

Proposition 84.2. *Let $K/\mathbb{Q}$ be a finite field extension of degree n and $\{\alpha_1, \dots, \alpha_n\}$ a subset of $\mathbb{Z}_K$ forming a $\mathbb{Q}$-basis for K. Then we have*
$$\mathbb{Z}[\alpha_1, \dots, \alpha_n] \subset \mathbb{Z}_K \subset \frac{1}{\Delta(\alpha_1, \dots, \alpha_n)} \mathbb{Z}[\alpha_1, \dots, \alpha_n].$$

Proof. The first inclusion is trivial. Let $\Delta = \Delta(\alpha_1, \dots, \alpha_n)$ and $\delta = \det\big((\sigma_i(\alpha_j)\big) \in \mathbb{Z}_K$, where $\sigma_1, \dots, \sigma_n : K \to \mathbb{C}$ are the $\mathbb{Q}$-embeddings of K. So $\Delta = \delta^2$ by Property 84.1(4). Suppose that $z \in \mathbb{Z}_K$. Then $z = x_1\alpha_1 + \dots + x_n\alpha_n$ for some $x_1, \dots, x_n$ in $\mathbb{Q}$. We have
$$\sigma_i(z) = x_1\sigma_i(\alpha_1) + \dots + x_n\sigma_i(\alpha_n) \text{ for } i = 1, \dots, n.$$
By Cramer's Rule, $x_i = y_i/\delta$ with y_i the determinant of the matrix $\big(\sigma_i(\alpha_j)\big)$ with the transpose $\big(\sigma_1(z), \dots, \sigma_n(z)\big)^t$, replacing its ith column for each i. Since $y_1, \dots, y_n, \delta$ are algebraic integers, $\Delta x_i = \delta^2 x_i = \delta y_i$ lies in $\mathbb{Z}_K \cap \mathbb{Q} = \mathbb{Z}$, i.e., x_i lies in $\frac{1}{\Delta}\mathbb{Z}$ for $i = 1, \dots, n$. The result follows. $\square$

Definition 84.3. Let A be a Dedekind domain with quotient field F, and K/F a finite separable field extension. The *discriminant ideal* of A_K is the ideal $\mathfrak{D}_{A_K/A}$ of A generated by the elements $\Delta(w_1, \dots, w_n)$ in A for all F-bases $\{w_1, \dots, w_n\}$ of K lying in A_K. Of course, if A is a ring of integers, then $\mathfrak{D}_{A_K/A} = (d_K)$, but in this case it is important to work with the integer d_K as it has a unique sign attached to it.

Proposition 84.4. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Suppose that $K = F(\alpha)$ with α an element of A_K. If $\mathfrak{p}$ is a nonzero prime ideal of A relatively prime to both $\mathfrak{D}_{A_K/A}$ and the conductor $\mathfrak{f}$ of $A[\alpha]$ in A_K, then $\mathfrak{p}$ is unramified in A_K. In particular, only finitely many primes in A ramify in A_K.*

PROOF. Let $^- : A_K[t] \to (A_K/\mathfrak{p}A_K)[t]$ be the canonical epimorphism and

$$\overline{m_F(\alpha)} = \overline{p_1}^{e_1} \cdots \overline{p_r}^{e_r}$$

be a factorization in $\overline{A}[t]$ into irreducible polynomials with $p_1, \dots, p_r$ monic polynomials in $A[t]$. By Theorem 82.5,

$$\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K,\ i = 1, \dots r$$

are the prime ideals lying over $\mathfrak{p}$ in A_K and $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$. Let $\alpha_1, \dots, \alpha_s$ be the roots of $m_F(\alpha)$ in a finite Galois extension of F containing K. Then $\Delta(\alpha) = \prod_{i<j}(\alpha_i - \alpha_j)^2$, so $\overline{m_F(\alpha)}$ has multiple roots if and only if $\overline{\alpha_i} = \overline{\alpha_j}$ for some $i \neq j$ if and only if $\overline{\Delta(\alpha)} = 0$. Hence $e_i = 1$ for $i = 1, \dots, r$ and each $\overline{p_i}$ has only simple roots. In particular, $A_K/\mathfrak{P}_i^{e_i} = A_K/\mathfrak{P}_i = \overline{A_K}(\overline{\alpha_i})$ is a separable extension of $\overline{A}$. □

Of course in the theorem, if $A_K = A[\alpha]$ for some α, then $\mathfrak{f}$ is the unit ideal, so the condition in this case is that $\mathfrak{p}$ is relatively prime to $\mathfrak{D}_{A_K/A}$. Unfortunately, in general A_K is not $A[\alpha]$ for any α. However, in the number field case, we can show by elementary means that if a prime ideal $p\mathbb{Z}$ in $\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ for $K/\mathbb{Q}$ finite, then p divides d_K.

Proposition 84.5. *Let K be a number field and p a prime integer. If (p) ramifies in $\mathbb{Z}_K$, then $p \mid d_K$.*

PROOF. Let L/K be a finite extension with $L/\mathbb{Q}$ a finite Galois extension. Suppose that $p \nmid d_K$ and (p) ramifies in $\mathbb{Z}_K$, say $p\mathbb{Z}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ is a factorization of $p\mathbb{Z}_K$ with $e_1 > 1$. Let $\mathfrak{A} = \mathfrak{P}_1^{e_1 - 1} \cdots \mathfrak{P}_r^{e_r} > p\mathbb{Z}_K$. We have $\mathfrak{P}_i \mid \mathfrak{A}$ for $i = 1, \dots, r$. Let $\{x_1, \dots, x_n\}$ be an integral basis for $\mathbb{Z}_K$ where $n = [K : \mathbb{Q}]$. Choose x in $\mathfrak{A} \setminus p\mathbb{Z}_K$ and write

$$x = m_1 x_1 + \cdots + m_n x_n$$

with $m_1, \dots, m_n$ in $\mathbb{Z}$. By assumption, there exists an i such that $p \nmid m_i$. We may assume that $i = 1$. We have $\{x, x_2, \dots, x_n\} \subset \mathbb{Z}_K$ is a $\mathbb{Q}$-basis for K and, using properties of determinants and Property 84.1(7), one checks that

$$\begin{aligned}\Delta(x, x_2, \dots, x_n) &= \Delta(m_1 x_1 + \cdots + m_n x_n, x_2, \dots, x_n)\\ &= m_1^2 \Delta(x_1, \dots, x_n).\end{aligned}$$

As x lies in every prime ideal in $\mathbb{Z}_K$ lying over (p), it lies in every prime ideal in $\mathbb{Z}_L$ lying over (p). Let $\mathfrak{Q}$ be a prime ideal in $\mathbb{Z}_L$ lying over (p). As $G(L/\mathbb{Q})$ acts transitively on the set of prime ideals in $\mathbb{Z}_L$ lying over (p), we have $\sigma(x)$ lies in $\mathfrak{Q}$ for every automorphism σ in $G(L/\mathbb{Q})$. By Property 84.1(4), the element $\Delta(x, x_2, \dots, x_n)$ lies in $\mathfrak{Q} \cap \mathbb{Z} = (p)$. Consequently, $p \mid \Delta(x, x_2, \dots, x_n) = m_1^2 \Delta(x_1, \dots, x_n)$. Since $p \nmid m_1$, we must have $p \mid \Delta(x_1, \dots, x_n) = d_K$, a contradiction. □

It can be shown that the converse to the last proposition is in fact true, i.e., if p is a prime integer, then (p) ramifies in a ring of algebraic integers $\mathbb{Z}_K$ if and only if $p \mid d_K$. Indeed, in Section 85, we shall prove Dedekind's Ramification Theorem that states if A is a Dedekind domain with quotient field F and K/F is a finite separable field extension, then a prime ideal $\mathfrak{p}$ ramifies in A_K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$. In the exercises for this section, we shall use localization techniques to reduce this to the case of a discrete valuation ring. In this section we shall, however, prove the following special case of it.

Proposition 84.6. *Let p be an odd prime, ζ a primitive pth root of unity in $\mathbb{C}$, and $K = \mathbb{Q}(\zeta)$. Then $\mathbb{Z}_K = \mathbb{Z}[\zeta]$. In particular, $\Delta_K = (-1)^{\frac{p-1}{2}} p^{p-2}$. Moreover, $p\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ and is the only prime ideal that ramifies in $\mathbb{Z}_K$.*

PROOF. By Example 65.7, we know that $\Delta(\zeta) = (-1)^{\frac{p-1}{2}} p^{p-2}$ and that $\mathrm{N}(\zeta) = \mathrm{N}(1-\zeta)$. Certainly, $\mathbb{Z}[\zeta] = \mathbb{Z}[1-\zeta]$. Suppose that $\mathbb{Z}[1-\zeta] = \mathbb{Z}[\zeta] < \mathbb{Z}_K$. Using Proposition 84.2, we can find an element $z \in \mathbb{Z}_K \setminus \mathbb{Z}[1-\zeta]$ satisfying

$$z = \sum_{j=i}^{p-1} \frac{m_j}{p}(1-\zeta)^j \quad \text{for some } i \geq 0 \text{ and } m_j \in \mathbb{Z} \text{ with } p \nmid m_i.$$

(why?) By Example 65.7, we know that $\prod_{j=0}^{p-1}(\zeta^j - 1) = p$, hence $p/(1-\zeta)^{p-1}$ lies in $\mathbb{Z}[\zeta]$, as $(1-\zeta) \mid (1-\zeta^j)$ for all $j > 0$. Consequently, $p/(1-\zeta)^j$ lies in $\mathbb{Z}[\zeta]$ for $j = 0, \dots, p-1$ and $zp/(1-\zeta)^j$ lies in $\mathbb{Z}_K$ for all j. It follows easily that this implies that $m_i/(1-\zeta)$ lies in $\mathbb{Z}_K$. But

$$p = \mathrm{N}_{K/\mathbb{Q}}(1-\zeta) \mid \mathrm{N}_{K/\mathbb{Q}}(m_i) = m_i^{p-1},$$

which is impossible. Therefore, $\mathbb{Z}_K = \mathbb{Z}[\zeta]$.

By Property 84.1(6), we have $d_K = \Delta(\zeta)$. Therefore, $\mathbb{Q}(\sqrt{\Delta}) = \mathbb{Q}\big(\sqrt{(-1)^{\frac{p-1}{2}} p}\big)$, the unique quadratic extension of $\mathbb{Q}$ in K by Proposition 59.14. Clearly, $p\mathbb{Z}$ ramifies in $\mathbb{Z}_{\mathbb{Q}(\sqrt{\Delta})}$, so the proposition follows. $\square$

Remark 84.7. If ζ is a primitive p^rth root of unity and ζ_p a primitive pth root of unity in $\mathbb{C}$ with p an odd prime or $p = 2$ and $r > 1$, then a modification of the proof shows that $\mathbb{Z}_{\mathbb{Q}(\zeta)} = \mathbb{Z}[\zeta] = \mathbb{Z}[1-\zeta]$ and the discriminant $d_{\mathbb{Z}_{\mathbb{Q}(\zeta)}}$ is a power of p by Lemma 65.8. Using Proposition 84.5 and the fact that $\mathbb{Q}\Big(\sqrt{(-1)^{\frac{p-1}{2}} p}\Big)$ is a subfield of $\mathbb{Q}(\zeta)$ if p is odd with $p\mathbb{Z}$ ramifying in $\mathbb{Z}_{\mathbb{Q}\left(\sqrt{(-1)^{\frac{p-1}{2}} p}\right)}$ and with 2 ramifying $\mathbb{Z}[\sqrt{-1}]$ (cf. the next section), we see that such $p\mathbb{Z}$ are precisely the prime ideals ramifying in $\mathbb{Z}_K$. In the general case, one shows that $\mathbb{Z}_L = \mathbb{Z}[\omega]$ for a primitive nth root of unity ω for arbitrary n. This is left as an exercise. The primes p that ramify in $\mathbb{Z}_L$ are precisely those odd primes dividing n and 2 if $4 \mid n$ using a similar analysis.

It can be shown that if $\mathbb{Q} < K$ is a finite extension, then $d_K > 1$ and some prime ideal in $\mathbb{Z}$ ramifies in $\mathbb{Z}_K$.

Exercises 84.8.

1. Prove Properties 84.1.

2. (Stickelberger's Criterion) Let K be a number field. Show that d_K is congruent to 0 or 1 modulo 4.
3. (Brill) Let K be a number field and r the number of embedding of K into $\mathbb{C}$ whose image does not lie in $\mathbb{R}$. Show that the sign of d_K is $(-1)^r$.
4. Let ζ be a primitive p^mth root of unity in $\mathbb{C}$ with p an odd prime or $p = 2$ and $m > 1$. Let $K = \mathbb{Q}(\zeta)$. Show that $\Delta_K \mid p^{p^{m-1}(p-1)}$ and $\mathbb{Z}_K = \mathbb{Z}[\zeta] = \mathbb{Z}[1-\zeta]$. In particular, show if p is an odd prime or $p = 2$ and $m \geq 2$, then $p\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ and is the only prime ideal ramifying in $\mathbb{Z}_K$.
5. Let $K/\mathbb{Q}$ and $L/\mathbb{Q}$ be finite and d the gcd of d_k and d_L. Show that $\mathbb{Z}_K\mathbb{Z}_L \subset \frac{1}{d}\mathbb{Z}_{KL}$ where KL is the compositum $K(L) - L(K)$ of K and L in $\mathbb{C}$. In particular, if $d = 1$, then $\mathbb{Z}_K\mathbb{Z}_L = \mathbb{Z}_{KL}$.
6. Using the last two exercises, show that if $K = \mathbb{Q}(\zeta)$ with ζ a primitive nth root of unity in $\mathbb{C}$, then $\mathbb{Z}_K = \mathbb{Z}[\zeta]$ and a prime p has $p\mathbb{Z}$ ramifies in K if and only if $p \mid n$ if p is odd or 2 if $4 \mid n$.
7. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Let $\mathfrak{p}$ be a maximal ideal in A and S a multiplicative set not containing zero. Show that $S^{-1}\mathfrak{D}_{A_K/A} = \mathfrak{D}_{S^{-1}A_K/S^{-1}A}$. In particular, if $S = A \setminus \mathfrak{p}$, then $\mathfrak{p} \mid \mathfrak{D}_{A_k/A}$ if and only if $\mathfrak{p}S^{-1}A \mid \mathfrak{D}_{S^{-1}A_K/S^{-1}A}$. (Cf. Exercise 82.7(6).) [This reduces the theorem that $\mathfrak{p}$ ramifies in A if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only if it does so locally, i.e., A is a discrete valuation ring.]
8. Let A be a discrete valuation ring with quotient field F and K/F a finite separable field extension with A-basis $\{x_1, \dots, x_n\}$. Then $\mathfrak{D}_{A_K/A} = \Delta_{K/F}(x_1, \dots, x_n)A$.

85. Dedekind's Theorem on Ramification

In this short section, we prove that a prime in a Dedekind domain ramifies in a finite separable extension of its quotient field if and only if it divides the discriminant ideal. We shall need to extend the definition of the trace of a finite separable field extension to an arbitrary finite field extension. We do this as follows: Let K/F be a finite field extension and K_s be the elements in K separable over F. The set K_s is a field (why?) called the *separable closure* of F in K. Define $\mathrm{Tr}_{K/F} : K \to F$ by $\mathrm{Tr}_{K/F} := [K : K_s]\,\mathrm{Tr}_{K_s/F}$. In particular, K/F is not separable if and only if $\mathrm{Tr}_{K/F} = 0$. (The norm $\mathrm{N}_{K/F}$ is defined by $\mathrm{N}_{K/F} := (\mathrm{N}_{K_s/F})^{[K:K_s]}$.) Exercises in previous sections that entail the use of localization reduce the proof of our desired result to the case of a discrete valuation domain, i.e., a local Dedekind domain. In particular, assuming these exercises and this extension of the trace, we are reduced to proving the following result:

Proposition 85.1. *Let A be a discrete valuation ring with quotient field F and K/F a finite separable extension. Then the maximal ideal $\mathfrak{p}$ in A ramifies in K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/F}$.*

PROOF. Let $n = [K : F]$. Since a discrete valuation ring is a PID by Remark 81.8(5), we know that A_K is a free A-module of rank n, say with integral basis $\{x_1, \dots, x_n\}$. It follows by Exercise 84.8(8) that $\mathfrak{D}_{A_K/A} = \Delta_{K/F}(x_1, \dots, x_n)A$. Hence $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only if $\Delta_{K/F}(x_1, \dots, x_n) \in \mathfrak{p}$.

Let

(a) $^{-}: A_K \to A_K/\mathfrak{p}A_K$ be the canonical epimorphism.
(b) $\overline{A} = A/\mathfrak{p}$.
(c) $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ be the factorization of $\mathfrak{p}$ in A_K into primes.
(d) $B_i = A_K/\mathfrak{P}_i^{e_i}$ for $i = 1, \dots, r$.

Using the Chinese Remainder Theorem, we identify $A_K/\mathfrak{p}A_k$ with $\bigtimes_{i=1}^r B_i$. We know that

$$n = [K : F] = \operatorname{rank}_A(A_K) = \dim_{\overline{A}} \overline{A_K}$$

by Theorem 82.2 (and its proof). Let $\mathcal{B}_i = \{u_{j_{i-1}+1}, \dots, u_{j_i}\}$ be an $\overline{A}$-basis for B_i, $i = 1, \dots r$. So $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is an $_{\overline{A}}$-basis for $\overline{A_K}$ with $|\mathcal{B}| = n$.

For each y in $\overline{A_K}$ define $\lambda_y : \overline{A_K} \to \overline{A_K}$ by $z \mapsto yz$ and $\operatorname{tr} : \overline{A_K} \to \overline{A_K}$ by $y \mapsto \operatorname{tr}(y) := \operatorname{trace}(\lambda_y)$. Let $\operatorname{tr}_i = \operatorname{tr}|_{B_i}$ for $i = 1, \dots, r$.

Note that we have

(i) $\{\overline{x_1}, \dots, \overline{x_n}\}$ is an $\overline{A}$-basis for $\overline{A_K}$.
(ii) $\operatorname{tr}(u_i u_j) = 0$ if u_i and u_j lie in different B_k's.

Claim. Let $x \in A_K$, then $\overline{\operatorname{Tr}_{K/F}(x)} = \operatorname{tr}(\overline{x})\,[= \operatorname{trace}(\lambda_{\overline{x}})]$. In particular, there exists a change of basis matrix $C \in \operatorname{GL}_n(\overline{A})$ satisfying

$$\overline{\Delta_{K/F}(x_1, \dots, x_n)} = \big(\det(C)\big)^2 \det\big(\operatorname{tr}(u_i u_j)\big),$$

hence

$$\begin{aligned} &\overline{\Delta_{K/F}(x_1, \dots, x_n)} = 0, \text{ i.e., } \Delta_{K/F}(x_1, \dots, x_n) \in \mathfrak{p}, \\ &\text{if and only if } \det\big(\operatorname{tr}(u_i u_j))\big) = 0 : \end{aligned}$$

Write $xx_i = \sum_{j=1}^n a_{ij} x_j$ with $a_{ij} \in A$ for all i and j and let α be the matrix $\big(a_{ij}\big)$. Then α is the matrix representation of λ_x relative to the basis $\{x_1, \dots, x_n\}$ and $\overline{\alpha}$ is the matrix representation of $\lambda_{\overline{x}}$ relative to the basis $\{\overline{x_1}, \dots, \overline{x_n}\}$. By Exercise 60.26(5), we have

$$\overline{\operatorname{Tr}_{K/F}(x)} = \overline{\operatorname{trace}(\lambda_x)} = \operatorname{trace}(\lambda_{\overline{x}}) = \operatorname{tr}(\overline{x}).$$

This proves the Claim. Now let $D_i := \big(\operatorname{tr}_i(u_l u_k)\big)$ for $j_{i-1} < l, k \leq j_i$, for each $i = 1, \dots, r$. Then

$$D := \Big(\operatorname{tr}_i(u_l u_k)\Big) = \begin{pmatrix} D_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_r \end{pmatrix} \quad (\text{with } j_0 = 0).$$

Suppose that $B_i = A_K/\mathfrak{P}_i^{e_i}$ is a field, i.e., $e_i = 1$. Then

$$\det D_i = \Delta_{B_i/\overline{A}}(u_{j_{i-1}+1}, \dots, u_{j_i}) = \det\big(\operatorname{Tr}_{B_i/\overline{A}}(u_l u_k)\big).$$

Since $B_i/\overline{A}$ is not separable if and only if $\operatorname{Tr}_{B_i/\overline{A}} = 0$, if follows that if B_i is a field for every i, then $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only $\det D_i = 0$ for some i if and only if there exists an i such that $B_i/\overline{A}$ is not separable.

So suppose that there exists an i with B_i not a field. We may assume that $i = 1$ and further that u_1 lies in $\mathfrak{P}_1/\mathfrak{P}_1^{e_1}$. Therefore, $(u_1 u_i)^{e_1} = 0$ for all i, hence $\lambda_{u_1 u_i}$ is nilpotent.

It follows that the first row of D consists of 0's by Example 45.7, hence $\det D = 0$ and $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$. $\square$

The generalization of the proposition to arbitrary Dedekind domains now follows by the reduction given by Exercise 84.8(7). That is, we the following theorem:

Theorem 85.2. (Dedekind Ramification Theorem) *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. The a maximal ideal $\mathfrak{p}$ of A ramifies in A_K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$.*

Immediate consequences of the theorem are:

Corollary 85.3. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Then A_K/A is unramified if and only if $\mathfrak{D}_{A_K/A} = A$.*

Remarks 85.4. Let F be a number field, so a finite extension of $\mathbb{Q}$.

1. Using Minkowski Theory, one can show that $|d_F| > 1$ (cf. Corollary 91.10), so $\mathbb{Z}_F/\mathbb{Z}$ cannot be unramified. As $\mathbb{Q}$ and finite fields are perfect, this means that there exists a prime p and a prime ideal $\mathfrak{P}$ in $\mathbb{Z}_F$ lying over $p\mathbb{Z}$ such that $e(\mathfrak{P}/p\mathbb{Z}) > 1$.
2. It can also be shown that there exist finitely many $F/\mathbb{Q}$ satisfying $d = d_F$, for some fixed integer d.
3. There can, however, be finite extensions K of the number field F that are unramified. Indeed a deep theorem of Classfield Theory says that the maximal unramified abelian extension K of F (in $\mathbb{C}$) is finite and of degree $|Cl_{\mathbb{Z}_F}|$ over F, where $CL_{\mathbb{Z}_F}$ ($= I_{\mathbb{Z}_F}/P_{\mathbb{Z}_K}$) is the class group of $\mathbb{Z}_K$ (finite for a ring of algebraic integers by Minkowski Theory – cf. Corollary 91.6). Furthermore, its Galois group $G(K/F)$ is (canonically) isomorphic to $CL_{\mathbb{Z}_F}$. [K is called the *Hilbert class field* of F.] In particular, $F < K$ if Z_F is not a PID.

Corollary 85.5. *Let $F = \mathbb{Q}(\alpha)$ with α integral over $\mathbb{Z}$ satisfying $g(\alpha) = 0$ with $g \in \mathbb{Z}[t]$ monic and p a prime in $\mathbb{Z}$. If $p \not| \, \mathrm{N}_{F/\mathbb{Q}}\big((g'(\alpha)\big)$, then $p\mathbb{Z}$ is unramified in $\mathbb{Z}_K$.*

Proof. Let $f = m_{\mathbb{Q}}(\alpha)$. By Proposition 65.6, we know that $\Delta_{F/\mathbb{Q}}(\alpha) = \pm \mathrm{N}_{F/\mathbb{Q}}\big(f'(\alpha)\big)$. Let $g = fh$ in $\mathbb{Z}[t]$, then $g' = f'h + fh'$. Consequently, $d_F \mid \Delta_{F/\mathbb{Q}}(\alpha)\,\mathrm{N}_{F/\mathbb{Q}}\big(h(\alpha)\big) = \pm \mathrm{N}_{F/\mathbb{Q}}\big(g'(\alpha)\big)$ $\square$

86. The Quadratic Case

Let d be a square-free integer. In this section, we compute the ring theory of $\mathbb{Z}_K$ with $K = \mathbb{Z}[\sqrt{d}]$ to illustrate the theory that we have developed.

Throughout this section let $\mathbb{Z}_K$ with $K = \mathbb{Z}[\sqrt{d}]$ and $G(K/F) = \{1_K, \sigma\}$, the Galois group of the Galois extension $K/\mathbb{Q}$. We first determine an integral basis and discriminant of K. We freely use Properties 84.1. We need a special case of Stickelberger's Criterion (Exercise 84.8(2)), viz.,

$$d_K \equiv 0 \mod 4 \text{ or } d_K \equiv 1 \mod 4.$$

Indeed, if $\{w_1, w_2\}$ is an integral basis for $\mathbb{Z}_K$, let $\alpha = w_1\sigma(w_2)$ and $\beta = w_2\sigma(w_1)$. Then

$$d_K = (\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta.$$

As $(\alpha-\beta)^2$, $(\alpha-\beta)^2$ and $\alpha\beta$ are all fixed by $G(K/\mathbb{Q})$, they are all integers, and the result follows. (The proof in the general case is analogous.) The above notation shall be used throughout this section. We also have:

$$\Delta(1,\sqrt{d}) = \det\begin{pmatrix}1 & \sqrt{d}\\ 1 & \sigma(\sqrt{d})\end{pmatrix}^2 = \det\begin{pmatrix}1 & \sqrt{d}\\ 1 & -\sqrt{d}\end{pmatrix}^2 = 4d.$$

As $\mathbb{Z}[\sqrt{d}] \subset \mathbb{Z}_K$, there exists an integer $a \in \mathbb{Z}$, satisfying

$$a^2 d_K = 4d, \text{ so } d_K = 4d \text{ or } d_K = d,$$

since d is square-free. (Note if $a = 2$, we must have d is odd, lest $d_K \equiv 2 \mod 4$ which contradicts Stickelberger's Criterion. So if d is even, $d_K = 4d$ and $8 \mid d_K$.) Using this and Stickelberger's Criterion, we solve our first goal.

Case 1. $d \equiv 2$ or $3 \mod 4$:
We have $d_K = 4d$ and $\mathbb{Z}_K = \mathbb{Z}[\sqrt{d}]$. In particular, $\{1, \sqrt{d}\}$ is an integral basis for $\mathbb{Z}_K$ by Property 84.1(7), and the conductor $\mathfrak{f}$ of $\mathbb{Z}[\sqrt{d}]$ in $\mathbb{Z}_K$ is $\mathbb{Z}_K$.

Case 2. $d \equiv 1 \mod 4$:
We have $(1-d)/4$ is an integer, hence $(1 \pm \sqrt{d})/2$ lies in $\mathbb{Z}_K$, as they are the roots of $t^2 - t + (1-d)/4$ in $\mathbb{Z}[t]$. It follows that

$$\mathbb{Z}[\sqrt{d}] \subset \mathbb{Z}[\frac{1+\sqrt{d}}{2}] \subset \mathbb{Z}_K.$$

We know that $\Delta\big(1, (1+\sqrt{d})/2\big) = b^2 d_K$, some integer b and computation yields $\Delta\big(1, (1+\sqrt{d})/2\big) = d$. Therefore, by Property 84.1(7), $d_K = d$, $\mathbb{Z}_K = \mathbb{Z}[(1+\sqrt{d})/2]$, and $\{1, (1+\sqrt{d})/2\}$ is an integral basis for $\mathbb{Z}_K$. In particular, the conductor $\mathfrak{f}$ of $\mathbb{Z}[\sqrt{d}]$ in $\mathbb{Z}_K$ contains the ideal $(2, 1-\sqrt{d}) = (2, 1+\sqrt{d})$.
We show $\mathfrak{f} = (2, 1-\sqrt{d})$. The element $(1+\sqrt{d})/2$ does not lie $\mathbb{Z}[\sqrt{d}]$ hence does not lie in $\mathfrak{f}$, since $\mathfrak{f} \subset \mathbb{Z}[\sqrt{d}]$. It follows if $a + b\big((1+\sqrt{d})/2\big)$ lies in $\mathfrak{f}$ with $a, b \in \mathbb{Z}$, we may assume that $a = b = 1$. As $1 + (1+\sqrt{d})/2$ does not lie in $\mathbb{Z}[\sqrt{d}]$, we have $\mathfrak{f} = (2, 1-\sqrt{d})$. If a prime ideal $\mathfrak{P} \mid (1-\sqrt{d})$, then $\mathfrak{P}\overline{\mathfrak{P}} \mid (1-d)\mathbb{Z}$. It follows that no factor of $p\mathbb{Z}_K$ with p an odd prime can be a factor of $\mathfrak{f}$.

Because of this computation, if p is a prime integer we can compute the splitting behavior of $p\mathbb{Z}_K$ in $\mathbb{Z}_K$ for all primes p by Theorem 82.5 except if $p = 2$ and $d \equiv 1 \mod 4$. In particular, it is applicable to all odd primes.

Case of an odd prime. Let p be an odd prime. We know that $2 = [K : \mathbb{Q}] = efr$ with e the ramification index of primes over (p), f the inertia index of (p), and r the number of primes in $\mathbb{Z}_K$ lying over (p). We also know that the Legendre symbol $\big(\frac{d}{p}\big)$ determines whether $t^2 - d$ splits over $\mathbb{Z}/p\mathbb{Z}[t]$ or not when d is not divisible by p, i.e., determines f for such p. Using the Kummer-Dedekind Theorem 82.5 and Proposition 84.4, we conclude, if p is an odd prime, that

1. If $p \mid d_K$, then $t^2 - d = t^2$ in $(\mathbb{Z}/p\mathbb{Z})[t]$, so

$$p\mathbb{Z}_K = (p, \sqrt{d})^2.$$

Therefore (p) ramifies in $\mathbb{Z}_K$ if and only if $p \mid d_K$.

2. The ideal $p\mathbb{Z}_K$ splits completely if and only if $\left(\frac{d}{p}\right) = 1$. If this is the case, then
$$p\mathbb{Z}_K = (p, x + \sqrt{d})(p, x - \sqrt{d})$$
a product of two distinct prime ideals with the integer x satisfying $x^2 \equiv d \mod p$.
3. The ideal $p\mathbb{Z}_K$ remains prime in $\mathbb{Z}_K$ if and only if $\left(\frac{d}{p}\right) = -1$.

Case of the even prime 2:

If $d \equiv 2$ or $3 \mod 4$, then Theorem 82.5 applies, and we see that in either case (2) ramifies (note $2 \mid d_K$), because

1. If $d \equiv 2 \mod 4$, then $2\mathbb{Z}_K = (2, \sqrt{d})^2$.
2. If $d \equiv 3 \mod 4$, then $2\mathbb{Z}_K = (2, d + \sqrt{d})^2 = (2, 1 + \sqrt{d})^2$.

So we are reduced to the case that $d \equiv 1 \mod 4$, and this reduces to the two cases of $d \equiv 1 \mod 8$ and $d \equiv 5 \mod 8$. Since Theorem 82.5 does not apply, we do this by brute force. Notice the computation shows that 2 does not ramify as expected and $(1 + \sqrt{d})/2$ lies in $\mathbb{Z}_K$.

We have

3. If $d \equiv 1 \mod 8$, then (2) splits completely in $\mathbb{Z}_K$:
We have
$$(2, \frac{1+\sqrt{d}}{2})(2, \frac{1-\sqrt{d}}{2}) = (4, 1+\sqrt{d}, 1-\sqrt{d}, \frac{1-d}{4}) \subset 2\mathbb{Z}_K$$
As $2\mathbb{Z}_K \subset (1+\sqrt{d}, 1-\sqrt{d})$, we conclude that
$$2\mathbb{Z}_K = (2, \frac{1+\sqrt{d}}{2})(2, \frac{1-\sqrt{d}}{2})$$
splits completely.
4. If $d \equiv 5 \mod 8$, then (2) remains a prime ideal in $\mathbb{Z}_K$:
We must show that $\mathbb{Z}/2\mathbb{Z} < \mathbb{Z}_K/\mathfrak{P}$ if $\mathfrak{P}|2$. If this was false, there are only two cosets of $\mathbb{Z}_K/\mathfrak{P}$, so there exists an integer a satisfying $a \equiv (1+\sqrt{d})/2 \mod \mathfrak{P}$. Since $(1+\sqrt{d})/2$ is a root of $t^2 - t + (1-d)/4$, we have $a^2 - a + (1-d)/4$ is even. As $a^2 - a$ is even, we would have $d \equiv 1 \mod 8$, a contradiction. Therefore, (2) remains a prime ideal.

As previously mentioned, a theorem of Kronecker says that any abelian extension of $\mathbb{Q}$ lies in a cyclotomic extension of $\mathbb{Q}$. We have shown that every quadratic extension of $\mathbb{Q}$ is a subfield of a cyclotomic extension of $\mathbb{Q}$ in Theorem 59.18. We use the proposition that implied this together with Hilbert Ramification Theory to interpret when an ideal (p) with p an odd prime in $\mathbb{Z}$ splits completely in $\mathbb{Z}_K$.

Proposition 86.1. *Let p and l be odd primes, $L = \mathbb{Q}(\zeta)$ with ζ a primitive l th root of unity, and $K = \mathbb{Q}(\sqrt{(\frac{-1}{l})l}\,)$. Then (p) splits completely in $\mathbb{Z}_K$ if and only if (p) splits into an even number of prime ideals in $\mathbb{Z}_L$.*

PROOF. By Proposition 59.14, we have K is the unique quadratic extension of $\mathbb{Q}$ in L. Suppose that p splits completely in $\mathbb{Z}_K$, say $p\mathbb{Z}_K = \mathfrak{P}_1\mathfrak{P}_2$ is its factorization. There exists an automorphism $\sigma \in G(L/\mathbb{Q})$ satisfying $\sigma(\mathfrak{P}_1) = \mathfrak{P}_2$ by Proposition 83.1. Then σ takes the set of prime ideals in $\mathbb{Z}_L$ over $\mathfrak{P}_1$ bijectively onto the set of prime ideals over $\mathfrak{P}_2$. Therefore the set of primes over (p) in $\mathbb{Z}_L$ is even. Conversely, suppose that the number of prime ideals in $\mathbb{Z}_L$ over (p) is even. Let $\mathfrak{P}$ be one such. Then the decomposition group $G_{\mathfrak{P}}$ of $\mathfrak{P}$ in $G(L/\mathbb{Q})$ has even index in $G(L/\mathbb{Q})$. By Galois Theory, this means the degree of the decomposition field $Z_{\mathfrak{P}}$ over $\mathbb{Q}$ is even. But $L/\mathbb{Q}$ is cyclic and even, so $K \subset Z_{\mathfrak{P}}$ using Proposition 83.3(2). $\square$

Using the fact that if ζ is a primitive nth root of unity that $\mathbb{Z}_{\mathbb{Q}(\zeta)} = \mathbb{Z}[\zeta]$, a fact that we did not prove but left as exercises (cf. Exercise 84.8(6)), one can give another proof of the Law of Quadratic Reciprocity essentially due to Kronecker. Although we do not do this, it is key to the generalization of the Law of Quadratic Reciprocity in Number Theory called Artin Reciprocity. [Cf. Ireland and Rosen, *A Classical Introduction to Modern Number Theory* [20], Chapter 13, §3 for this proof.]

We have determined the ideal structure of the quadratic number field $\mathbb{Z}_K$ and from our work how nonzero ideals in $\mathbb{Z}$ decompose in $\mathbb{Z}_K$ as $\mathbb{Z}_K$ is a Dedekind domain and ideals are products of prime ideals, unique up to order. The remaining problem is to determine the units in $\mathbb{Z}_K$. This depends on whether d is positive or negative. The case of negative d is easy, the case of positive d not so.

Case of $d < 0$. We show that

1. If $d = -1$, then $\mathbb{Z}_K^\times = \{\pm 1, \pm\sqrt{-1}\}$.
2. If $d = -3$, then $\mathbb{Z}_K^\times = \{\pm 1, \pm\zeta, \pm\zeta^2\}$, with ζ a primitive cube root of unity.
3. If $d < -3$ or $d = -2$, then $\mathbb{Z}_K^\times = \{\pm 1\}$.

We have seen this for the Gaussian integers, and the proof generalizes, i.e., we take the norm for K to $\mathbb{Q}$. Let ε be a unit in $\mathbb{Z}_K$. Then $\mathrm{N}_{K/F}(\varepsilon) = \pm 1$.

If $d \equiv 2$ or 3 modulo 4, we can write $\varepsilon = x + y\sqrt{d}$ for some integers x and y, so $x^2 + |d|y^2 = 1$. If $d = -1$, we get (1) and if $|d| > 3$ we have $\mathbb{Z}_K^\times = \{\pm 1\}$.

If $d \equiv 1 \mod 4$, then $\{1, (1+\sqrt{d}/2\}$ is an integral basis, and we see that we can write ε as $(x + y\sqrt{d})/2$ with x and y integers satisfying $x \equiv y \mod 2$. We then have $x^2 + |d|y^2 = 4$. If $d = -3$ we get (2) and if $|d| > 3$, we see that $\mathbb{Z}_K^\times = \{\pm 1\}$.

To do the case for positive d, we need two results. The first is the following:

Observation 86.2. If M is a positive real number, there exist finitely many α in $\mathbb{Z}_K$ such that $\max(|\alpha|, |\sigma(\alpha)|) < M$.

and the second a famous theorem from elementary number theory, viz.,

Proposition 86.3. *Let d be a square-free positive integer. Then Pell's equation $x^2 - dy^2 = 1$ has infinitely many solutions in integers. Further, there exists an integral solution (x_1, y_1) such that every solution is of the form (x_n, y_n) with $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$ for some integer n. In particular, the solution set in integers to Pell's equation forms an infinite cyclic group.*

that we prove in Appendix H.

Using these two results, we prove the case of positive discriminant.

Case of $d > 1$. We show that there exists a unit $u > 1$ in $\mathbb{Z}_K$ such that $(\mathbb{Z}_K)^\times = \{\pm u^m \mid m \in \mathbb{Z}\}$.

By Proposition 86.3, there exist positive integers x and y satisfying $x^2 - dy^2 = 1$. Let $\varepsilon = x + y\sqrt{d}$. Then ε is a unit in $\mathbb{Z}_K$ with $\varepsilon > 1$. Let $M > \varepsilon$ and u be a unit in $\mathbb{Z}_K$ satisfying $1 < u < M$. So $\mathrm{N}_{K/\mathbb{Q}} = u\sigma(u) = \pm 1$. If $\sigma(u) = -1/u$, then $-M < -1/u < M$ and if $\sigma(u) = 1/u$, then we also have $-M < 1/u < M$. By the observation, there exist only finitely many such u and ϵ is at least one of them. Among all of these choose the smallest unit $u > 1$. If $v > 0$ is any other unit, then there exist a unique integer n (not necessarily positive) satisfying $u^n \le v < u^{n+1}$, so $1 \le vu^{-n} < u$. By the choice of u, we must have $v = u^n$. If $v < 0$ is a unit in $\mathbb{Z}_K$, then $-v > 0$ is a unit in $\mathbb{Z}_K$ and $-v = u^n$ some integer n. The result follows.

[Note the similarity of the proof with that in Appendix H.]

The generator of the unit group in this case is called the *fundamental unit*. It can be found using continued fractions, but may not be so obvious. For example, the fundamental unit of $\mathbb{Z}[\sqrt{94}]$ is $2143295 + 221064\sqrt{94}$.

If K is a number field, then the structure of $\mathbb{Z}_K$ can be shown to be $\mu_K \times \mathbb{Z}^n$ where $n = r + s - 1$, with r the number of embeddings of K into $\mathbb{R}$ and $2s$ the number of embeddings of K into $\mathbb{C}$ not contained in $\mathbb{R}$, so the number of embeddings into $\mathbb{C}$ is $r + 2s$ (as the non-real embeddings come in complex conjugate pairs). This is proved using Minkowski Theory in Section 90.

Exercise 86.4. Prove Observation 86.2.

87. Addendum: Valuation Rings and Prüfer Domains

In this addendum, we shall generalize the concept of Dedekind domain. By Exercise 81.14(14), a Dedekind domain is a Noetherian domain A in which the localization at every nonzero prime ideal $\mathfrak{p}$ in A is a discrete valuation ring. Another characterization of a Dedekind domain was that it was a Noetherian domain A in which every fractional ideal in A was invertible. We look at these when we do not assume the domain is Noetherian.

In Exercise 81.14(13) we defined a valuation ring. We begin by recalling the definition.

Definition 87.1. A domain A with quotient field F is called a *valuation ring* of F if for every nonzero element x in F, either x lies in A or x^{-1} lies in A. In particular, if a and b are nonzero elements in A, then $a \mid b$ or $b \mid a$ in A. We call a domain a *valuation ring* if it is a valuation ring in its quotient field.

Every discrete valuation ring is a Noetherian valuation ring by Exercise 81.14(13), hence the localization of any Dedekind domain at a nonzero prime ideal is a valuation ring. Valuation rings are ubiquitous, but unfortunately a valuation ring is Noetherian only when it is a discrete valuation ring. The general concept, however, was very important in the algebraization of algebraic geometry.

Properties 87.2. Let A be a valuation ring of F. Then

(1) A is a local ring.

(2) If $a_1, \dots, a_n$ are nonzero elements in A, then there exists an i such that $a_i \mid a_j$ for $j = 1, \dots, n$.
(3) If $A \subset B \subset \mathrm{F}$ are subrings, then B is a valuation ring.
(4) A is integrally closed.
(5) Every finitely generated ideal in A is principal. A domain in which every finitely generated ideal is principal is called a *Bézout domain.*

PROOF. (1): Let $\mathfrak{A} = A \setminus A^\times$, the set of nonunits in A. As A is a valuation ring, a lies in $\mathfrak{A}$ if and only if a^{-1} does not lie in $\mathfrak{A}$.

Claim. $\mathfrak{A}$ is an ideal.

Let a, b be nonzero elements in $\mathfrak{A}$ and r a nonzero element in A. Certainly, $ra \in A$. Suppose that $ra \notin \mathfrak{A}$. Then $(ra)^{-1} \in A$, hence $a^{-1} = r(ra)^{-1}$ lies in A, a contradiction. So $\mathfrak{A}$ is closed under multiplication by elements in A. As either ab^{-1} or $a^{-1}b$ lies in A, say ab^{-1}, it follows that $a + b = b(ab^{-1} + 1)$ lies in $\mathfrak{A}$. Hence $\mathfrak{A}$ is an ideal and clearly the unique maximal ideal.

Statements (2) and (3) are immediate, (4) follows by Exercise 87.27(1), and (5) follows form (2). □

We conclude the following:

Corollary 87.3. *Let A be a Noetherian domain, not a field. Then A is a valuation ring if and only if A is a discrete valuation ring.*

PROOF. ($\Rightarrow$): By the Properties 87.2, A is a local PID so every nonzero prime ideal is maximal and integrally closed. Since it is Noetherian, it is a local Dedekind domain, hence a discrete valuation ring.

($\Leftarrow$): Is Exercise 81.14(13). A key to the proof is if $\mathfrak{m}$ is the maximal ideal of A, then $x \notin A^\times \cup \mathfrak{m}^2$ must satisfy $(x) = \mathfrak{m}$ by uniqueness of factorization of ideals into prime ideals in a Dedekind domain. It follows that A must be a PID and the result follows. □

We next look at fractional ideals in a domain a bit more closely.

Let A be a domain with quotient field F and M a fractional ideal, i.e., a nonzero A-submodule of F such that there exists a nonzero element x of A satisfying $xM \subset A$. As before, we shall let I_A denote the set of fractional ideals in A. We say a fractional ideal M is an *invertible fractional ideal* if there exists an A-submodule N of F such that $MN = A$. Let $Inv(A)$ denote the set of invertible fractional ideals. As before, we let $P_A = \{Ax \mid 0 \neq x \in F\}$, the set of *principal fractional ideals.*

Remarks 87.4. Let A be a domain with quotient field F and M an A-submodule of F.

1. Every nonzero ideal of A is a fractional ideal.
2. A nonzero finitely generated A-submodule of F is a fractional ideal — clear denominators.
3. Let x be a nonzero element of F, then $M \in Inv(A)$ if and only if $xM \in Inv(A)$. In particular, $P_A \subset Inv(A)$.
4. If A is Noetherian and $M \in I_A$, then there exists a nonzero x in F such that xM is an ideal. In particular, M is a finitely generated A-module. It follows that if A is Noetherian, I_A is the set of nonzero finitely generated A-submodules of F.

As before if $M \in I_A$, we let $M^{-1} := \{x \in F^\times \mid xM \subset A\}$. Of course, $MM^{-1} \subset A$.

Lemma 87.5. *Let A be a domain with quotient field F and M, N be A-submodules of F satisfying $MN = A$. Then $N = M^{-1}$ and M is a finitely generated A-module. In particular, $M \in I_A$.*

PROOF. By definition $N \subset M^{-1}$, so $N \subset M^{-1}MN \subset AN = N$. It follows that $N = M^{-1}$. If M is finitely generated then $M \in I_A$ by Remark 2 above. In any case, since $A = MM^{-1}$, we have an equation $1 = \sum_{i=1}^n a_i b_i$ for some a_i in M, some $b_i \in M^{-1}$, and some n. Then $M = \sum_{i=1}^n Aa_i$. Indeed, if $x \in M$, then $x = \sum_i a_i(b_i x)$ lies in $\sum_{i=1}^n Aa_i$. $\square$

We call a commutative ring a *semi-local ring* if it has only finitely many maximal ideals.

Lemma 87.6. *If A is a semi-local domain, then $Inv(A) = P_A$. In particular, A is a PID.*

PROOF. Let $\mathfrak{m}_1, \dots, \mathfrak{m}_n$ be the maximal ideals in A and suppose that $M \in Inv(A)$, so $MM^{-1} = A$. For each $i = 1, \dots, n$, choose $a_i \in M$ and $b_i \in M^{-1}$ such that $a_i b_i \notin \mathfrak{m}_i$. Since $\bigcap_{j \neq i} \mathfrak{m}_j \not\subset \mathfrak{m}_i$ for $i = 1, \dots, n$, there exist $c_i \in \left(\bigcap_{j\neq i} \mathfrak{m}_j\right) \setminus \mathfrak{m}_i$ for $i = 1, \dots n$. Let $x = \sum c_i b_i$, an element of M^{-1}. Therefore, xM is an ideal of A. If $xM = A$, then x is not zero and $M = x^{-1}A$ is a principal fractional ideal. So suppose that $xM < A$. Then $xM \subset \mathfrak{m}_i$ for some i. In particular, $xa_i \in \mathfrak{m}_i$. As $c_j b_j a_i \in \mathfrak{m}_i$ for every $j \neq i$, we have $c_i b_i a_i$ lies in $\mathfrak{m}_i$. This is impossible. $\square$

Let A be a commutative ring and $S \subset A$ a multiplicative set. If M is an A-module, then $S^{-1}M := \{\frac{m}{s} \mid s \in S,\, m \in M\}$ is an A- and an $S^{-1}A$-module in the obvious way [which is?]. If $\mathfrak{p}$ is a prime ideal in A, we denote by $M_\mathfrak{p}$ the $A_\mathfrak{p}$-module $S^{-1}M$ with $S = A \setminus \mathfrak{p}$. The following is straight-forward:

Lemma 87.7. *If A is a domain, S a multiplicative set in A not containing 0, and $M \in Inv(A)$, then $S^{-1}M \in Inv(S^{-1}A)$.*

We have the following 'local-global' principle:

Proposition 87.8. *Let A be a domain with quotient field F and M an A-submodule of F. Then the following are equivalent:*

(1) $M \in Inv(A)$.
(2) M is a finitely generated A-module and $M_\mathfrak{p} \in Inv(A)$, for all prime ideals $\mathfrak{p}$ in A.
(3) M is a finitely generated A-module and $M_\mathfrak{m} \in Inv(A)$, for all maximal ideals $\mathfrak{m}$ in A.
(4) M is a finitely generated A-module and $M_\mathfrak{m} \in P_A$, for all maximal ideals $\mathfrak{m}$ in A.

PROOF. (1) $\Rightarrow$ (2) follows from lemmas 87.7 and 87.5, (2) $\Rightarrow$ (3) is immediate, and (3) $\Rightarrow$ (4) follows from Lemma 87.6.

(4) $\Rightarrow$ (1): Suppose that $MM^{-1} < A$. Then there exist a maximal ideal $\mathfrak{m}$ in A with $MM^{-1} \subset \mathfrak{m}$. By assumption, M is finitely generated, say $M = Aa_1 + \cdots + Aa_n$, and there exists a nonzero x in M satisfying $xA_\mathfrak{m} = M_\mathfrak{m}$. By the definition of localization, there exist nonzero s_i in $A \setminus \mathfrak{m}$ satisfying $s_i a_i \in Ax$, $i = 1, \dots, n$. Set nonzero $s = s_1 \cdots s_n$.

Then $sM \subset Ax$, hence $sx^{-1}M \subset A$, i.e., $sx^{-1} \in M^{-1}$. It follows that $s = sx^{-1}x$ lies in $M^{-1}M \subset \mathfrak{m}$, a contradiction. Therefore, $MM^{-1} = A$ as desired. □

A domain is called a *Bézout domain* if every finitely generated ideal in A is principal and a *Prüfer domain* if every finitely generated ideal is invertible.

Remarks 87.9. 1. Every Bézout domain is a Prüfer domain.

2. A domain A is a Bézout domain if and only if $Inv(A) = P_A$.
3. A domain is a Prüfer domain if and only if every finitely generated fractional ideal is invertible

We leave the following generalization of the fact that a DVR is a PID as an easy exercise:

Proposition 87.10. *Let A be a local domain. Then A is a valuation ring if and only if A is a Bézout domain.*

The following shows that Prúfer domains generalize Dedekind domains when the Noetherian assumption is omitted.

Theorem 87.11. *Let A be a domain. Then the following are equivalent:*

1. *A is a Prüfer domain.*
2. *$A_\mathfrak{p}$ is a valuation ring for all prime ideals $\mathfrak{p}$ in A.*
3. *$A_\mathfrak{m}$ is a valuation ring for all maximal ideals $\mathfrak{m}$ in A.*

Proof. (1) $\Rightarrow$ (2): Let $\mathfrak{A}$ be a finitely generated ideal in $A_\mathfrak{p}$ with $\mathfrak{p}$ a prime ideal in A. Then there exist a_i in $A \setminus \mathfrak{p}$ and s_i in S, $i = 1, \ldots, n$, some n, such that $\mathfrak{A} = \sum_{i=1}^n A_\mathfrak{p}\frac{a_i}{s_i} = (\sum_{i=1}^n Aa_i)_\mathfrak{p}$. (Why?) Since $\sum Aa_i$ is invertible so is $\mathfrak{A}$. But then $\mathfrak{A}$ is principal by Proposition 87.8 as $A_\mathfrak{p}$ is local. In particular, $A_\mathfrak{p}$ is Bézout, hence a valuation ring by the previous proposition.

(2) $\Rightarrow$ (3) is immediate.

(3) $\Rightarrow$ (1): Let $\mathfrak{A}$ be a finitely generated ideal in A. If $\mathfrak{m}$ is a maximal ideal in A, then $\mathfrak{A}_\mathfrak{m}$ is principal as $A_\mathfrak{m}$ is a valuation ring. Consequently, it is invertible by Proposition 87.8. □

We know that the integral closure of a Dedekind domain in a finite separable closure of its quotient field is also a Dedekind domain (although separable is not needed by the Krull-Akizuki Theorem 96.17 to be proven below). We shall generalize this to a Prüfer domain. Indeed we shall obtain the stronger result that the field extension need only be algebraic, i.e., we do not have to assume that it is finite nor separable. To do so we introduce a concept that leads to the ubiquity of valuation rings.

Definition 87.12. Let A be a subring of B and $\mathfrak{A} < A$ an ideal. We say that $\mathfrak{A}$ *survives* in B if $\mathfrak{A}B < B$.

For example, by Claim 82.1, if A is Dedekind domain then any ideal $\mathfrak{A} < A$ survives in A_L for any field extension L of $qf(A)$. Indeed this is true if B/A is an integral extension for any commutative ring A (as we shall prove in Theorem 93.14 below).

Lemma 87.13. (Chevalley) *Let A be a subring of the commutative ring B and $\mathfrak{A} < A$ an ideal. If u is a unit in B, then $\mathfrak{A}$ survives in either $A[u]$ or in $A[u^{-1}]$.*

Proof. Suppose this is false. Then there exist elements $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ in $\mathfrak{A}$ satisfying:

$$\text{(i)} \qquad 1 = a_0 + a_1 u + \cdots + a_n u^n$$

$$\text{(ii)} \qquad 1 = b_0 + b_1 u + \cdots + b_m u^{-m}$$

in B, where we may assume that these have been chosen with $m, n \geq 0$ minimal. Since $\mathfrak{A} < A$, we may also assume that m and n are positive. We also assume that $m \leq n$. Note that $a_0 \neq 1$ and $b_0 \neq 1$ as $\mathfrak{A} < A$. Multiplying equation (ii) by $a_n u^n$ yields

$$\text{(iii)} \qquad a_n u^n (1 - b_0) = b_1 a_n u^{n-1} + \cdots + b_m a_n u^{n-m}$$

and multiplying (i) by $1 - b_0$ yields

$$\text{(iv)} \qquad 1 - b_0 = (1 - b_0) a_0 + \cdots + (1 - b_0) a_n u^n.$$

Plug (iii) into (iv) for $a_n(1 - b_0)u^n$ gives an equation of the form (i) of lesser degree in u than n, a contradiction. □

Theorem 87.14. *Let A be a domain with quotient field F with K/F a field extension, and $\mathfrak{A} < A$ an ideal. Then there exists a valuation ring B of K containing A with $\mathfrak{A}$ surviving in B.*

Proof. Let

$$\begin{aligned} \mathcal{S} = \{(A_\alpha, \mathfrak{A}_\alpha) \mid A \subset A_\alpha \subset K \ &\text{ subrings}, \\ &\mathfrak{A} \subset \mathfrak{A}_\alpha < A_\alpha \ \text{ with } \mathfrak{A}_\alpha \text{ an ideal in } A_\alpha\} \end{aligned}$$

Partially order $\mathcal{S}$ by $\subseteq$ of pairs. Since $(A, \mathfrak{A}) \in \mathcal{S}$, a Zorn Lemma argument produces a maximal element $(B, \mathfrak{B}) \in \mathcal{S}$. Suppose that B is not a valuation ring of K. Then there exists a nonzero element x in K with neither x nor x^{-1} in B. As $\mathfrak{B}$ survives in $B[x]$ or $B[x^{-1}]$, say $B[x]$, we would have $(B, \mathfrak{B}) < (B[x], B[x]\mathfrak{B})$. Since $B[x] \subset B[x, x^{-1}] \subset K$, the pair $(B[x], B[x]\mathfrak{B})$ lies in $\mathcal{S}$, a contradiction. □

Definition 87.15. Let A and B be local rings with maximal ideals $\mathfrak{m}, \mathfrak{n}$, respectively. We say that B *dominates* A if A is a subring of B and $\mathfrak{m} = A \cap \mathfrak{n}$.

Corollary 87.16. *Let A be a local domain with its quotient field F lying in a field K. Set*

$$\mathcal{D}_K = \{B \text{ a local ring} \mid B \subset K\}$$

partially ordered by domination. Then

(1) *There exists a valuation ring V in $\mathcal{D}_K$ dominating A.*
(2) *A is a valuation ring in $\mathcal{D}_F$ if and only if A is maximal in $\mathcal{D}_F$.*

PROOF. (1) follows from Theorem 87.14 with $\mathfrak{A}$ the maximal ideal of A.
(2): The proof of Theorem 87.14 implies the if statement. Conversely, suppose that $A \subset B \subset F$ are valuation rings of F with maximal ideals $\mathfrak{m}_A$ and $\mathfrak{m}_B$, respectively, with B dominating A, we must show that $A = B$. If $A < B$, then there exists a nonzero element x in $B \setminus A \subset F$. Since A is a valuation ring, x^{-1} lies in $A \subset B$, so $x \in B^{\times} \setminus A$. It follows that x^{-1} lies in $\mathfrak{m}_A = A \setminus A^{\times}$. Thus $\mathfrak{m}_B \cap A = \mathfrak{m}_A = A \setminus A^{\times}$, a contradiction. □

This corollary has an application to Prüfer domains.

Corollary 87.17. *Let R be a Prüfer domain with quotient field F. Suppose that A is a valuation ring in F containing R. Then there exists a prime ideal $\mathfrak{p}$ in R such that $A = R_{\mathfrak{p}}$.*

PROOF. Let $\mathfrak{m}$ be the maximal ideal of A and $\mathfrak{p} = \mathfrak{m} \cap R$, a prime ideal in R. Let $s \in R \setminus \mathfrak{p}$. Then s^{-1} lies in A, lest $s \in \mathfrak{m} \cap R$. Therefore, $R_{\mathfrak{p}} \subset A$. By Theorem 87.11 $R_{\mathfrak{p}}$ is a valuation ring, hence by Corollary 87.16, we have $R_{\mathfrak{p}} = A$. □

This corollary gives a starting point for the theory of algebraic functions in one variable (geometrically curve theory). The algebraic formulation is the following:

Corollary 87.18. *Let F be a field and A a valuation ring satisfying $F \subset A \subset F(t)$. Then there exists an irreducible polynomial f in $F[t]$ satisfying $A = F[t]_{(f)}$ or $A = F[t^{-1}]_{(t^{-1})}$*

PROOF. If t lies in A, then $F[t] \subset A$. As $F[t]$ is a PID, it is Prüfer, so the result follows from Corollary 87.17. So we may assume that $t \notin A$. As A is a valuation ring t^{-1} lies in A. By Corollary 87.17, $A = F[t^{-1]}]_{\mathfrak{p}}$ for some prime ideal in $F[t^{-1}]$. Since $t \notin A$, we have $t^{-1} \in \mathfrak{p}$, so $(t^{-1}) = \mathfrak{p}$. □

We now wish to generalize Theorem 81.5 to the more general theorem about Prüfer domains previously mentioned. We need the following lemma.

Lemma 87.19. *Let A be an integrally closed local domain with quotient field F and $a \in F$. Suppose that there exists a polynomial f in $A[t]$ with $f(a) = 0$ and at least one coefficient of f is a unit in A. Then either a or a^{-1} lies in A.*

PROOF. Suppose that $ba^n + ca^{n-1} + g(a) = f(a) = 0$ with $b, c \in A$ and $g \in A[t]$ satisfying $\deg g < n - 1$. If b is a unit in A, then we are done as A is integrally closed, so we may assume that $b \notin A^{\times}$. Then

$$(*) \qquad 0 = (ba)^n + c(ba)^{n-1} + b^{n-1}g(a) = (ba)^n + c(ba)^{n-1} + g_1(ba)$$

for some $g_1 \in A[t]$ satisfying $\deg g_1 < n - 1$. Since A is integrally closed, ba lies in A, as ba is integral over A by (*). If ba is a unit in A, then so is $b^{-1}a^{-1}$. This implies that $a^{-1} \in A$. Therefore, we ay assume that ba lies in the maximal ideal $\mathfrak{m}$ of A. We know that $(ba + c)a^{n-1} + g(a) = 0$. If $c \in A^{\times}$, then $ba + c \in A^{\times}$, as A is local and $ba \in \mathfrak{m}$. But this means that a is integral over A, so $a \in A$ as A is integrally closed. Therefore, we may assume that $ba + c$ lies in $\mathfrak{m}$ and one of the coefficients of g is a unit. As $\deg\big((ba + c)t^{n-1} + g\big) < n$, we are done by induction on n. □

Theorem 87.20. *Let A be a Prüfer domain with quotient field F. Let L/F be an algebraic (possibly infinite) algebraic field extension. Then A_L is Prüfer.*

Proof. Let $\mathfrak{m}$ be a maximal ideal in A_L and $\mathfrak{p} = A \cap \mathfrak{m}$. We need to show that $B_{\mathfrak{m}}$ is a valuation ring. Since A is a Prüfer domain, $A_{\mathfrak{p}}$ is a valuation ring by Theorem 87.11. Let a be an element in L and f a nonzero polynomial in $F[t]$ with $f(a) = 0$. Multiplying by a suitable element in A, we may assume that $f \in A[t]$. Since $A_{\mathfrak{p}}$ is a valuation ring there exists a coefficient r of f dividing all the other coefficients of f in A. Thus $\frac{1}{r} f \in A_{\mathfrak{p}}[t]$ and has a coefficient that is a unit in $A_{\mathfrak{p}}$, hence in $(A_L)_{\mathfrak{m}}$. By the lemma, either a or a^{-1} lies in $(A_L)_{\mathfrak{m}}$ as needed. □

Corollary 87.21. *Let A be a Dedekind domain with quotient field F. Let L/F be an algebraic (possibly infinite) algebraic field extension. Then A_L is a Prüfer domain. In particular, if $\widetilde{\mathbb{Q}}$ is an algebraic closure of $\mathbb{Q}$, then $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is a Prüfer domain.*

Remarks 87.22. 1. If $\widetilde{\mathbb{Q}}$ is an algebraic closure of $\mathbb{Z}$, then $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is not Noetherian so not a Dedekind domain.

2. The ring of entire functions on an open Riemann surface is a Prúfer domain. In fact, it is a Bézout domain.
3. If in the proposition, Bézout is substituted for Prúfer, then it can be shown that A_L is also Bézout.
4. Bergman showed that if $A = \mathbb{Z}\big[\sqrt{-5}\,\big] + t\mathbb{Q}\big[\sqrt{-5}\,\big][t]$, i.e., A is the ring of polynomials in $\mathbb{Q}\big[\sqrt{-5}\,\big][t]$ having constant term in $\mathbb{Z}\big[\sqrt{-5}\,\big]$, then A is not Noetherian, not Bézout, but is Prüfer.
5. Every ideal in a Dedekind domain can be generated by two elements by Exercise 81.14(11). The same is false for finitely generated ideals in Prüfer domains. Schulting gave the first counterexample. It was subsequently generalized by Swan.

We give two further applications of valuation rings that are useful in commutative algebra. The first has as a consequence Zariski's Lemma 41.10 that we saw was the key to the Hilbert Nullstellensatz in Section 41.

Theorem 87.23. *Let F be an algebraically closed field and A a domain with $R \subset A$ a subring. Suppose that A is a finitely generated R-algebra and $a \in A$ nonzero. Then there exists an element $0 \neq r \in R$ satisfying the following condition: whenever $\varphi : A \to F$ is a ring homomorphism with $0 \neq \varphi(r)$, there exists a ring homomorphism $\psi : A \to F$ satisfying $\psi|_R = \varphi$ and $0 \neq \psi(a)$.*

Proof. By induction it suffices to assume that $A = F[u]$.

Case 1. The element u is transcendental over R.

Let $a = \sum_{i=0}^{n} a_i u^i$ in $F[u]$ with a_0 nonzero. We show $r = a_0$ works. So suppose that $\varphi : R \to F$ is a ring homomorphism with $\varphi(a_0) \neq 0$. Since F is infinite, there exists an element $x \in F$ such that $\sum_{i=0}^{n} \varphi(a_i)x^i \neq 0$. Let $\psi : A \to F$ extend φ by setting $\psi(u) = x$. This works.

Case 2. The element u is algebraic over R.

As R is a domain and a nonzero, a^{-1} is also algebraic over the quotient field of R, so there exist equations:

$$a_0u^n + a_1u^{n-1} + \cdots + a_n = 0$$
$$b_0a^{-m} + b_1a^{1-m} + \cdots + b_m = 0$$

for appropriate a_i and b_j in A with a_0 and b_0 nonzero. It follows that u and a^{-1} are integral over $R[(a_0b_0)^{-1}]$. Set $r = a_0b_o$ and suppose that $\varphi : R \to F$ is a ring homomorphism satisfying $\varphi(r) \neq 0$. Set $\mathfrak{p} = \ker\varphi$. Then $\mathfrak{p}$ survives in $R[r^{-1}]$ by choice, so there exists a valuation ring B in the quotient field of $R[r^{-1}]$ such that $\mathfrak{p}R[r^{-1}]$ survives in B. If $b \in B$, we can write $b = r_1/r_2$ with $r_1, r_2 \in A$ and $r_2 \notin \mathfrak{p}$ as $qf(A) = qf\big(R[r^{-1}]\big)$. Hence we have a ring homomorphism $\widetilde{\varphi} : B \to F$ given by $r_1/r_2 \mapsto \varphi(r_1)/\varphi(r_2)$. We show that $A \subset B$ and $\psi : A \to F$ with $\psi = \widetilde{\varphi}|_A$ work. Since the integral closure of $R[r^{-1}]$ lies in B, we have a^{-1} and u lie in B. Thus $A \subset B$ and $a \in B^\times$. It follows that a does not lie in the maximal ideal of B, so $\widetilde{\varphi}(a) \neq 0$. □

Corollary 87.24. (Zariski's Lemma) *Let K/F be an extension of fields. Suppose that K is a finitely generated F-algebra. Then K/F is a finite field extension.*

Proof. Let $\widetilde{F}$ be an algebraic closure of F. Then apply the theorem with $R = F$, $A = \widetilde{F}$, and $a = 1$. By the Theorem, if x in $\widetilde{F}$ is transcendental over F, the homomorphism $\widetilde{F} \to \widetilde{F}$ sending $x \to 0$ cannot occur, which is false. □

Since we have seen that Zariski's Theorem implies the Hilbert Nullstellensatz, this gives an alternate proof of it. As a second application, we wish to show if A is an integrally closed domain, then so is $A[t]$. To do this we first need the following:

Theorem 87.25. (Krull) *Let A be a domain with quotient field F and*

$$\mathcal{S} = \{B \mid A \textit{ is a subring of } B \textit{ and } B \textit{ is a valuation ring}\}.$$

The A is integrally closed if and only if $A = \bigcap_{\mathcal{S}} B$.

Proof. ($\Leftarrow$): As valuation rings are integrally closed,this follows by Exercise 87.27(5). ($\Rightarrow$): Certainly, $A = \bigcap_{\mathcal{S}} B$, so assume that there exists an element u in $(\bigcap_{\mathcal{S}} B) \setminus A$. Then u does not lie in $A_F = A$, hence $u \notin A[u^{-1}]$ by Exercise 87.27(1). Let $v = u^{-1}$, then $vA[v] < A$, as $v \notin A[v]^\times$. Hence there exists a valuation ring $B_0 \subset F$ such that $vA[v]$ survives in B_0. Since u lies in B for every $u \in \mathcal{S}$, we must have $v \in B_0^\times$. But this means that $vA[v]$ cannot survive in B_0, a contradiction □

Theorem 87.26. *Let A be an integrally closed domain. Then $A[t]$ is integrally closed.*

Proof. We have

$$(*) \qquad \Big(\bigcap_{\substack{A\subset B\subset F \\ B \text{ a valuation ring}}} B\Big)[t] = \Big(\bigcap_{\substack{A\subset B\subset F \\ B \text{ a valuation ring}}} B[t]\Big).$$

(Why?) and by the previous theorem,

$$A = \Big(\bigcap_{\substack{A\subset B\subset F \\ B \text{ a valuation ring}}} B\Big).$$

As valuation rings are integrally closed by Exercise 87.27(5), we may assume that A is a valuation ring.

Let f be a nonzero element of $F(t)$ (the quotient field of $A[t]$) be integral over $A[t]$. We must show that f lies in $A[t]$. Since $F[t]$ is a UFD hence integrally closed, f lies in $F[t]$. Write $f = \frac{1}{a}g$ with $g \in A[t]$ and $a \in A$ nonzero. If $a \in A^\times$, we are done, so assume not. As A is a valuation ring, there exists a nonzero coefficient b of g dividing all the coefficients of g, so we may write $f = \frac{b}{a}h$ with $h \in A[t]$ and h has some coefficient 1. If $a \mid b$ we are done, so assume not. As A is a valuation ring, we must have $b \mid a$ in A, so $f = \frac{1}{c}h$ for some nonzero element c in A. As c is not a unit in A and A is local it lies in the maximal ideal $\mathfrak{m}$ of A. As f is integral over $A[t]$, there exists polynomials, $\alpha_1(t), \dots, \alpha_n(t)$ in $A[t]$ for some n satisfying $f^n + \alpha_1(t)f^{n-1} + \cdots + \alpha_n(t) = 0$ in $F[t]$. Hence $h^n + c\alpha_1(t)h^{n-1} + \cdots + c^n\alpha_n(t) = 0$ in $A[t]$. Consequently, $h^n \equiv 0 \mod A[t]\mathfrak{m}$. But $A[t]/\mathfrak{m}A[t] = (A/\mathfrak{m})[t]$ (why?) is a domain, i.e., $A[t]\mathfrak{m}$ is a prime ideal in $A[t]$, so $h \in A[t]\mathfrak{m}$ $(= \mathfrak{m}[t])$. This implies that every coefficient of h lies in $\mathfrak{m}$, a contradiction.

□

Exercises 87.27.

1. Let A be a commutative ring and u a unit in A. If R is a subring of A, show that u^{-1} is integral over R if and only if u^{-1} lies in $R[u]$.
2. Let A be a commutative ring. Then A is a domain if and only if $A_\mathfrak{m}$ is a domain for all maximal ideals $\mathfrak{m}$ of A
3. Show 87.7.
4. Establish Proposition 87.10
5. Prove that the intersection of integrally closed domains each of which lies in a common field is integrally closed.

CHAPTER XVI

Algebraic Number Fields

In the previous chapter we studied the algebra of Dedekind domains. Since the ring of integral elements (i.e., the ring of algebraic integers) in an (algebraic) number field (a finite field extensions of the rational numbers) is a Dedekind domain, we also developed some of the properties of a ring of algebraic integers in that chapter. For example, we investigated Hilbert ramification theory of a ring of algebraic numbers including a discussion of its discriminant. We also fully computed the theory for quadratic number rings. In the case of a quadratic number ring, the units were easy to compute (especially in the non-real case). In this chapter, we wish to classify the units in any ring of algebraic numbers. We also wish to investigate the discriminant of a ring of algebraic numbers more thoroughly. Two of the main advantages in the special case of number theory is that its ideals are finitely generated free $\mathbb{Z}$-modules and its quotient fields by its nonzero prime ideals are finite fields.

88. Ideal and Counting Norms

In this section, we need a preliminary discussion of norms, that we could have done previously. We know that if A is a integrally closed domain with quotient field F and K is a finite separable extension, that $N_{K/F} : K \to F$ maps A_K to A. If we also assume that A is a Dedekind domain, then we have unique factorization of fractional ideals into products of prime ideals in the Dedekind domain A_K. This will allow us to define the norm of a fractional ideal of A_K to be a fractional ideal in A. If $F = \mathbb{Z}$, i.e., $A_K = \mathbb{Z}_K$, then the norm of an element in $\mathbb{Z}_K$ is an integer. Moreover, if $\mathfrak{p}$ is a nonzero prime ideal in $\mathbb{Z}_K$, then $\mathbb{Z}_K/\mathfrak{p}$ is a finite (separable) extension of the finite field $\mathbb{Z}/\mathfrak{p}\cap\mathbb{Z}$ of degree $f(\mathfrak{p}/\mathfrak{p}\cap\mathbb{Z})$. This allows us to define another norm. We will then show how these norms are interrelated.

Since we have investigated the case for an arbitrary Dedekind domain, we shall begin by studying the case of norms of ideals in this more general setting.

We shall need the following that we leave as an exercise.

Lemma 88.1. *Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F. If $\mathfrak{A} \in I_A$, then $\mathfrak{A}_K \cap A = \mathfrak{A}$ and $I_A \to I_{A_K}$ by $\mathfrak{A} \to \mathfrak{A}A_K$ is a group monomorphism.*

Definition 88.2. Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F. If $\mathfrak{P}$ is a nonzero prime ideal in A_K, we know that $f(\mathfrak{P}/\mathfrak{P}\cap A) := \dim_{A/(\mathfrak{P}\cap A)}(A_K/\mathfrak{P})$ is a finite integer by Theorem 82.2. By the unique factorization of fractional ideals in a Dedekind domain, we can define $\mathrm{N}_{K/F} : I_{A_K} \to I_A$ to be the group homomorphism induced by $\mathfrak{P} \mapsto (\mathfrak{P}\cap A)^{f(\mathfrak{P}/\mathfrak{P}\cap A)}$ called the *ideal norm map.*

By the multiplicativity the ideal norm, we have:

Remark 88.3. Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F. If $\mathfrak{A} \in I_{A_K}$, then $N_{K/F}(\mathfrak{A}) = \sum_{a\in\mathfrak{A}} \mathrm{N}(a)A$, i.e., the ideal generated by norms of elements.

We now show that ideal norms have similar properties as norms of elements and how the two are interrelated. We need two lemmas.

Lemma 88.4. *Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F. If $\mathfrak{A} \in I_A$, then $\mathrm{N}_{K/F}(\mathfrak{A}A_K) = \mathfrak{A}^{[K:F]}$.*

Proof. As the ideal norm is a group homomorphism, we may assume that $\mathfrak{A} = \mathfrak{p}$ is a nonzero prime ideal in A. Let $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \dots \mathfrak{P}_r^{e_r}$ be a prime factorization of $\mathfrak{p}A_K$ in A_K and $f_i = f(\mathfrak{P}_i/\mathfrak{p})$ for $i = 1, \dots, r$. By definition, $\mathrm{N}_{K/F}(\mathfrak{p}A_K) = \mathfrak{p}^{e_1 f_1} \cdots \mathfrak{p}^{e_r f_r}$, hence $\mathrm{N}_{K/F}(\mathfrak{p}A_K) = \mathfrak{p}^{[K:F]}$ by Theorem 82.2. $\square$

Lemma 88.5. *Let L/F be a finite Galois extension and A a Dedekind domain with quotient field F, $\mathfrak{p}A_L = \mathfrak{P}_1^{e_1} \dots \mathfrak{P}_r^{e_r}$ a factorization, with $\mathfrak{p}$ a nonzero prime in A. Then $e = e_i$ and $f_i = f(\mathfrak{P}_i/\mathfrak{p})$ for $i = 1, \dots, r$ and $\mathrm{N}_{L/F}(\mathfrak{P}_1)A_L = (\mathfrak{P}_1 \cdots \mathfrak{P}_r)^{ef} = \prod_{G(L/F)} \sigma(\mathfrak{P}_1^e)$.*

Proof. By Proposition 83.2 (and Proposition 83.1), we have $e = e_i$ and $f = f_i$ for $i = 1, \dots, r$ and $G(L/F)$ acts transitively on $\mathfrak{P}_1, \dots, \mathfrak{P}_r$. The result follows by Lemma 88.4. $\square$

Proposition 88.6. *Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F and $x \in K^\times$. Then*

$$\mathrm{N}_{K/F}(x)A = \mathrm{N}_{K/F}(xA_K),$$

where the norm on the right-hand side is the ideal norm and the norm on the left-hand side is the field norm.

Proof. Let L/F be the normal closure of K/F and $x \in K^\times$, then

$$\mathrm{N}_{L/F}(x)A = \mathrm{N}_{K/F}(x^{[L:K]})A = \big(\mathrm{N}_{K/F}(x)A\big)^{[L:K]}$$
$$\mathrm{N}_{L/F}(xA_L) = \mathrm{N}_{K/F}\left((xA_K)^{[L:K]}\right) = \left(\mathrm{N}_{K/F}(xA_K)^{[L:K]}\right)$$

by Lemma 88.4 and the multiplicativity of f on towers by Exercise 82.7(2). By uniqueness of factorization of ideals, we are reduced to the case that $L = K$, i.e., K/F is Galois. We then have

$$\begin{aligned} \mathrm{N}_{K/F}(x)A_K &= \prod_{G(K/F)} \sigma(x)A_K = \prod_{G(K/F)} \sigma(xA_K) \\ &= \prod_{G(K/F)} \prod_{\mathfrak{P}} \sigma\big(\mathfrak{P}^{v_\mathfrak{P}(x)}\big) = \prod_{G(K/F)} \prod_{\mathfrak{P}} \sigma(\mathfrak{P})^{v_\mathfrak{P}(x)} \\ &= \mathrm{N}(xA_K)A_K \end{aligned}$$

by Lemma 88.5 and the multiplicativity of N. By Lemma 88.1, $\mathfrak{A}A_K \cap A = \mathfrak{A}$, so the result follows. $\square$

We also need another norm in the case of a number field.

Definition 88.7. Let K be a number field. If $\mathfrak{A}$ is a nonzero ideal in $\mathbb{Z}_K$, define the *counting norm* of $\mathfrak{A}$ to be the integer $\mathfrak{N}(\mathfrak{A}) := |\mathbb{Z}_K/\mathfrak{A}|$.

We want to extend the counting norm $\mathfrak{N}$ to a group homomorphism $\mathfrak{N} : I_{\mathbb{Z}_K} \to \mathbb{Q}^\times$. To do so, we now show that the counting norm is multiplicative. We also show how it is related to the ideal norm.

Proposition 88.8. *Let K be a number field and $\mathfrak{A}, \mathfrak{B} \subset \mathbb{Z}_K$ nonzero ideals. Then*

(1) $\mathfrak{N}(\mathfrak{A}) = \prod_{\operatorname{Max}(\mathbb{Z}_K)} \mathfrak{N}(\mathfrak{p})^{v_\mathfrak{p}(\mathfrak{A})}$.

(2) $\mathfrak{N}(\mathfrak{A}\mathfrak{B}) = \mathfrak{N}(\mathfrak{A})\,\mathfrak{N}(\mathfrak{B})$.

(3) If E/K is a finite extension, then $\mathfrak{N}(\mathfrak{A}\mathbb{Z}_E) = \mathfrak{N}(\mathfrak{A}\mathbb{Z}_K)^{[K:F]}$.

(4) If $a \in \mathbb{Z}_K$ is nonzero, then $\mathfrak{N}_{K/\mathbb{Q}}(a\mathbb{Z}_K) = |\operatorname{N}_{K/\mathbb{Q}}(a)|$.

(5) $\mathfrak{N}(\mathfrak{A}) = |\operatorname{N}_{K/\mathbb{Q}}(\mathfrak{A})|$, *where we mean a positive generator for the ideal $\operatorname{N}_{K/\mathbb{Q}}(\mathfrak{A})$ on the right-hand side.*

Proof. (2): If $\mathfrak{A}$ and $\mathfrak{B}$ are relatively prime, we have $\mathbb{Z}_K = \mathfrak{A} + \mathfrak{B}$. In this case, we have $\mathfrak{N}(\mathfrak{A}\mathfrak{B}) = \mathfrak{N}(\mathfrak{A})\mathfrak{N}(\mathfrak{B})$ by the Chinese Remainder Theorem. So we need only show:

Claim Let $\mathfrak{p}$ be a nonzero prime ideal in $\mathbb{Z}_K$. Then $\mathfrak{N}(\mathfrak{p}^r) = \mathfrak{N}(\mathfrak{p})^r$ for all positive integers r:

Fix r and choose $\alpha \in \mathfrak{p}^r \setminus \mathfrak{p}^{r+1}$. Then we have (cf. Exercise 79.14(1))

(i) $\mathfrak{p}^r = \alpha\mathbb{Z}_K + \mathfrak{p}^{r+1}$, the greatest common divisor of $\alpha\mathbb{Z}_K$ and $\mathfrak{p}^{r+1}$.

(ii) $\alpha\mathfrak{p}^r = \alpha\mathbb{Z}_K \cap \mathfrak{p}^{r+1}$, the least common multiple of $\alpha\mathbb{Z}_K$ and $\mathfrak{p}^{r+1}$.

Since $\alpha\mathfrak{p}^r \subset \mathfrak{p}^{r+1}$, it follows that for each r, we have a $\mathbb{Z}_K$-isomorphism $\varphi : \mathbb{Z}_K/\mathfrak{p} \to \mathfrak{p}^r/\mathfrak{p}^{r+1}$ given by $\beta \mapsto \beta\alpha$, which is also an isomorphism of one dimensional $\mathbb{Z}_K/\mathfrak{p}$-vector spaces. If follows (say by the Jordan-Hölder Theorem) that

$$\mathfrak{N}(\mathfrak{p}^s) = |\mathbb{Z}_K/\mathfrak{p}^s| = \prod_{i=0}^{s-1} |\mathfrak{p}^i/\mathfrak{p}^{i+1}| = \mathfrak{N}(\mathfrak{p})^s$$

where $\mathfrak{p}^0 = \mathbb{Z}_K$. This proves the Claim and establishes (2).

(1) follows from (2).

(3): By (2), we may assume that $\mathfrak{A} = \mathfrak{p}$ is a prime ideal in $\mathbb{Z}_K$. Let $\mathfrak{p}\mathbb{Z}_E = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ be a factorization in $\mathbb{Z}_E$ and $f_i = f(\mathfrak{P}_i/\mathfrak{p})$ for $i = 1, \dots, r$. By (2), we have $\mathfrak{N}(\mathfrak{P}_i^{e_i}) = \mathfrak{N}(\mathfrak{P}_i)^{e_i} = \mathfrak{N}(\mathfrak{p})^{e_i f_i}$, hence

$$\mathfrak{N}(\mathfrak{p}\mathbb{Z}_E) = \mathfrak{N}(\mathfrak{p})^{\sum_{i=1}^r e_i f_i} = \mathfrak{N}(\mathfrak{p})^{[K:F]}$$

by Theorem 82.2 as needed.

(5): As above, we may assume that $\mathfrak{A} = \mathfrak{p}$ is a prime ideal in $\mathbb{Z}_K$. Let $p\mathbb{Z} = \mathfrak{p} \cap \mathbb{Z}$. Then $\mathfrak{N}(\mathfrak{p}) = p^f$ with $f = [\mathbb{Z}_K/\mathfrak{p} : \mathbb{Z}/p\mathbb{Z}]$. By definition $\operatorname{N}_{K/\mathbb{Q}}(\mathfrak{p}) = p^f\mathbb{Z}$. Statement (5) now follows.

(4) follows from (5). $\square$

Corollary 88.9. *Let K be a number field and $\mathfrak{A}, \mathfrak{B} \subset \mathbb{Z}_K$ nonzero ideals. Then $\mathfrak{N}$ extends to a group homomorphism $\mathfrak{N} : I_{\mathbb{Z}_K} \to \mathbb{Q}^\times$. Moreover, the counting norm satisfies all the properties in Proposition 88.8 for all $\mathfrak{A}, \mathfrak{B} \in I_A$.*

The key property that we shall need, besides the multiplicativity of the counting norm, is that we can replace the ideal norm of a principal fractional ideal by the absolute value of the field norm of a positive generator of it. In particular, if u is a unit in $\mathfrak{A}_K$ and $\alpha \in \mathbb{Z}_K$, then $\mathfrak{N}(a\mathbb{Z}_K) = |\operatorname{N}_{K/\mathbb{Q}}(au)|$.

Exercises 88.10.
Let K/F be a finite separable field extension and A a Dedekind domain with quotient field F. Suppose that $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in A. Prove the following:

1. If $\mathfrak{A}A_K \mid \mathfrak{B}A_K$ in A_K, then $\mathfrak{A} \mid \mathfrak{B}$ in A.
2. Lemma 88.1.
3. Remark 88.3

89. Lattices in Number Fields

In this section, we show how to view a ring of algebraic integers as a lattice in an appropriate Euclidean space, i.e., a finite dimensional real inner product space in an appropriate way and thereby establish that it is a discrete subgroup there. Our goal in this section is to prove the Minkowski Lattice Point Theorem, a most useful result in number theory. We shall assume the necessary topology and analysis in this section.

We start with the setup that we shall need throughout the rest of this chapter. Let K be a number field of degree n over $\mathbb{Q}$. Then there exist n embeddings of K into $\mathbb{C}$ by Proposition 56.11 and the comment following (cf. Proposition 58.3).

We write these embeddings as

$$\sigma_1, \ldots, \sigma_{r_1} : K \to \mathbb{R}$$

for all the *real embeddings* (i.e., those with image in $\mathbb{R}$) and

$$\sigma_{r_1+1}, \ldots, \sigma_{r_1+r_2}, \overline{\sigma}_{r_1+1}, \ldots, \overline{\sigma}_{r_1+r_2} : K \to \mathbb{C}$$

for the *complex embeddings* (i.e., those with image not contained in $\mathbb{R}$), where $^- : \mathbb{C} \to \mathbb{C}$ is complex conjugation and $\overline{\sigma}_{r_1+i}$ is the composition $^- \circ \sigma_{r_1+i}$ for $i = 1, \ldots, r_2$. In particular, we have $[K : \mathbb{Q}] = n = r_1 + 2r_2$. If $x \in K$, we shall use the following notation:

$$x^{(i)} = \sigma_i(x),\ i = 1, \ldots, n, \text{ where } \sigma_{r_1+r_2+j} = \overline{\sigma}_{r_1+j} \text{ for } j = 1, \ldots r_2.$$

We view $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2} =$

$$\{(x_1, \ldots, x_{r_1}, \ldots, x_{r_1+r_2}) \mid x_i \in \mathbb{R},\ 1 \le i \le r_1, \\ \text{and } x_j \in \mathbb{C},\ r_1 + 1 \le j \le r_2\}$$

as a real vector space (with $r_1 = 0$ or $r_2 = 0$ allowed and with each $\mathbb{C}$ having basis $\{1, \sqrt{-1}\}$).
Therefore, if $V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$, we have an $\mathbb{Z}$-embedding

$$i_{\mathbb{C}} : K \to \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \text{ via } x \mapsto \big(\sigma_1(x), \ldots, \sigma_{r_1+r_2}(x)\big) = (x^{(1)}, \ldots, x^{(r_1+r_2)})$$

where the map will usually be suppressed. If $V = \mathbb{R}^n$, $n = r_1 + 2r_2$, we shall also need another additive embedding given by

$$i_{\mathbb{R}} : K \to \mathbb{R}^n \text{ where}$$
$$x \mapsto \big(x^{(1)}, \dots, Re(x^{(r_1+1)}), Im(x^{(r_1+1)}), \dots, Re(x^{(r_1+r_2)}, Im(x^{(r_1+r_2)})\big).$$

Definition 89.1. Suppose that V is an n-dimensional real inner product space, e.g., $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ above, and $\{v_1, \dots, v_r\}$ a linearly independent set in V. Let $\mathcal{L} := \coprod_{i=1}^{r} \mathbb{Z}v_i$. Then $\mathcal{L}$ is a free $\mathbb{Z}$-module of rank r called an *r-dimensional lattice* in V with *lattice basis* $\{v_1, \dots, v_r\}$. We often write $\mathcal{L}$ as $\mathcal{L}(v_1, \dots, v_r)$ for clarity, when we wish to highlight the lattice basis. If $r = n = \dim V$, we call $\mathcal{L}$ a *full lattice* in V, i.e., if $\{v_1, \dots, v_r\}$ is a basis for V. If $\mathcal{L}$ is a full lattice in V, we let

$$D_{\mathcal{L}} = D_{\mathcal{L}(v_1,\dots,v_n)} := \{\sum_{i=1}^{n} a_i v_i \mid 0 \le a_i < 1,\, a_i \in \mathbb{R},\, 1 \le i \le n\}$$

called the *fundamental domain* (or *parallelopiped*) of $\mathcal{L}$.

Note that the fundamental domain $D_{\mathcal{L}}$ of $\mathcal{L}$ depends on $\{v_i, \dots, v_n\}$ even thought $\mathcal{L}$ does not.

Remarks 89.2. Let V be a finite dimensional real inner product space.

1. The determinant of the change of basis matrix of any two lattices bases of a lattice $\mathcal{L}$ in V is ± 1.
2. If $\mathcal{L}$ is a full lattice in V, then $V/\mathcal{L}$ is a compact space.
3. Suppose that $\mathcal{L} = \mathcal{L}(v_1, \dots, v_n)$ is a full lattice in V and $\mathcal{B} = \{e_1, \dots, e_n\}$ is an orthonormal basis for V. If $v_i = \sum_{i=1}^{n} a_{ij} e_j$, then
$$\mathrm{Vol}(D_{\mathcal{L}}) = \mathrm{Vol}(V/\mathcal{L}) = |\det(a_{ij})|$$
and is independent of lattice basis by (1).

Example 89.3. Let $V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ or $\mathbb{R}^n$, K a number field of degree n. We know that $\mathbb{Z}_K$ is a free abelian group of rank n, so is a full lattice when viewed in the image of K. Let $0 < \mathfrak{A} < \mathbb{Z}_K$ be an ideal, hence also $\mathbb{Z}$-free. By Corollary 81.6, there exists a nonzero $m \in \mathfrak{A} \cap \mathbb{Z}$, so $0 < m\mathbb{Z}_K \subset \mathfrak{A} < \mathbb{Z}_K$. In particular, $\mathbb{Z}_K/m\mathbb{Z}_k$ is a finite additive torsion group. Since $\mathbb{Z}_K/\mathfrak{A}$ is a finite group, $\mathfrak{A}$ is also a full lattice in V. More generally, if $\mathfrak{A}$ is a fractional ideal, there exists a nonzero $c \in \mathbb{Z}_K$ such that $c\mathfrak{A}$ is an ideal in $\mathbb{Z}_K$. Therefore, $\mathrm{N}_{K/\mathbb{Q}}(c)$ is an integer so $\mathrm{N}_{K/\mathbb{Q}}(c)\mathfrak{A} \subset \mathbb{Z}_K$ also. As $\mathrm{N}_{K/\mathbb{Q}}(c) \in \mathbb{Z}$, we also have $[\frac{1}{\mathrm{N}_{K/\mathbb{Q}}(c)}\mathbb{Z}_K : \mathfrak{A}]$ is finite and $\mathfrak{A}$ is a full lattice in V. Note under $i_{\mathbb{C}}$ and $i_{\mathbb{R}}$, these induce lattices in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ and $\mathbb{R}^{r_1+2r_2}$, respectively, with fundamental domains having the same volume. In particular, this will also be true for the lattices induced by any $\mathfrak{A} \in I_{\mathbb{Z}_K}$

Lemma 89.4. *Let V be a finite dimensional real inner product space and $\mathcal{L} = \mathcal{L}(v_1, \dots, v_n)$ be a full lattice in V. Then $\{\lambda + D_{\mathcal{L}} \mid \lambda \in \mathcal{L}\}$ partitions V.*

Proof. Let $v \in V$, then $V = \sum_{i=1}^{n}(q_i + r_i)v_i$ with $0 \le r_i < 1$ and $q_i \in \mathbb{Z}$ for $i = 1, \dots, n$. Thus the sets $\lambda + D_{\mathcal{L}}$ cover V. If we have $(\lambda + D_{\mathcal{L}}) \cap (\lambda' + D_{\mathcal{L}}) \neq \emptyset$, then $D_{\mathcal{L}} \cap \big((\lambda - \lambda') + D_{\mathcal{L}}\big) \neq \emptyset$. It follows that $\lambda = \lambda'$. □

If V is a finite dimensional real inner product space and $\mathcal{L}$ an additive subgroup of V, we call $\mathcal{L}$ a *discrete subgroup* of V if $\mathcal{L}$ meets every compact subset of V in a finite set.

Proposition 89.5. *Let V be a finite dimensional real inner product space and $\mathcal{L}$ an additive subgroup of V. Then $\mathcal{L}$ is a lattice if and only if $\mathcal{L}$ is discrete. In particular, if $\mathcal{L}$ is discrete, then it is a finitely generated additive group.*

PROOF. ($\Rightarrow$): Let $C \subset V$ be a compact set, $\mathcal{L} = \mathcal{L}(w_1, \dots, w_r)$. If $\sum_{i=1}^{r} a_i w_i$ lies in $C \cap \mathcal{L}$, then the integers $|a_i|$'s must be uniformly bounded.
($\Leftarrow$): Let $\{v_1, \dots, v_n\}$ be a basis for V. We may assume that $\mathcal{L} \neq 0$, and we induct on $n = \dim V$. If $n = 1$, choose $a > 0$ minimal such that $v = av_1$ lies in $\mathcal{L}$. (Such an a exists as we can intersect $\mathcal{L}$ with a large enough ball centered at the origin.) Let $bv_1 \in \mathcal{L}$ with $b \neq 0$. Write $b = da + c$ with $0 \leq c < a$ and $d \in \mathbb{Z}$. By hypothesis, we must have $c = 0$, so $bv_1 = dav_1$, hence $\mathcal{L} = \mathbb{Z}v$.
Now suppose that $n > 1$. We may assume that $\mathcal{L}$ lies in no proper subspace of V by induction. Set $V_0 = \coprod_{i=1}^{n-1} \mathbb{R}v_i$ and $\mathcal{L}_0 = \mathcal{L} \cap V_0 < \mathcal{L}$. So $\mathcal{L}_0$ cannot lie in any proper subspace of V_0, hence by induction $\mathcal{L}_0$ is a full lattice of rank $n-1$ in V_0. Let $\mathcal{L}_0 = \mathcal{L}(u_1, \dots, u_{n-1})$. Then $\{u_1, \dots, u_{n-1}, v_n\}$ is a basis for V. Since $\mathcal{L}_0 < \mathcal{L}$, we see that there exists an element $\lambda \in \mathcal{L}$ satisfying $\lambda = \sum_{i=1}^{n-1} r_i u_i + r_n v_n$ with $0 \leq r_i < 1$, $i = 1, \dots, n-1$, and $r_n > 0$. Let

$$C = \{\sum_{i=1}^{n-1} x_i u_i + x_n v_n \mid 0 \leq x_i \leq 1,\, i = 1, \dots, n-1,\, |x_n| \leq r_n\},$$

a compact subset of V. Hence $\mathcal{L} \cap C$ is a finite set. Choose $u_n \in \mathcal{L} \setminus \mathcal{L}_0$ to satisfy

$$u_n = \sum_{i=1}^{n-1} x_i u_i + x_n v_n,\ \ 0 \leq x_i < 1,\, i = 1, \dots, n-1, \text{ and } \ |x_n| > 0 \ \text{ minimal.}$$

Claim. $\mathcal{L} = \mathcal{L}(u_1, \dots, u_n)$.
If not then there exists a $w \in \mathcal{L} \setminus \mathcal{L}(u_1, \dots, u_n)$. Since $\mathcal{L}(u_1, \dots, u_n) = \mathcal{L}(u_1, \dots, u_{n-1}) \oplus \mathbb{Z}u_n$, it is a full lattice in V. Subtracting a suitable nonzero integral multiple of u_n and then by a suitable element of $\mathcal{L}_0$, we see, using $\mathbb{Z}u_n \cap \mathcal{L}_0 = 0$, that there exist $0 \leq b_i < 1$, $i = 1, \dots, n-1$, and $0 < a_n|x_n| < |x_n|$ satisfying $w_0 = \sum_{i=1}^{n-1} b_i u_i + a_n x_n v_n$ lies in $\mathcal{L}$. This contradicts the minimality of $|x_n|$. (Cf. the $n = 1$ case.) □

For the rest of this chapter, we shall let μ denote the natural (Lebesgue) measure on a finite dimensional real inner product space.

Lemma 89.6. *Let V be a finite dimensional real inner product space and $\mathcal{L}_1 \subset \mathcal{L}_2$ be full lattices in V. Then $\mathcal{L}_2/\mathcal{L}_1$ is finite and*

$$\mu(D_{\mathcal{L}_1}) = \mathrm{Vol}(V/\mathcal{L}_1) = [\mathcal{L}_2 : \mathcal{L}_1] \,\mathrm{Vol}(V/\mathcal{L}_2) = [\mathcal{L}_2 : \mathcal{L}_1] \mu(D_{\mathcal{L}_2}).$$

PROOF. Since $\mathcal{L}_2/\mathcal{L}_1$ is a finitely generated group of rank zero, it is torsion hence finite. Let $\mathcal{L}_2 = \mathcal{L}(v_1, \dots, v_n)$. By linear algebra, there exists a $\mathbb{Z}$-basis $\{w_1, \dots, w_n\}$ for $\mathcal{L}_1$ satisfying

$$(i) \qquad w_i = c_{ii} v_i + c_{i,i+1} v_{i+1} + \cdots + c_{in} v_n,\ \ c_{ij} \in \mathbb{Z} \ \text{ and } \ c_{ii} > 0 \ \text{ for all } i, j.$$

Moreover,

$$\alpha := |\det(c_{ij})| = c_{11}\cdots c_{nn} \text{ is independent of } \{w_1, \dots, w_n\}$$

(as change of bases matrices for lattices lie in $\mathrm{GL}_n(\mathbb{Z})$). So $\mathrm{Vol}(V/\mathcal{L}_1) = \alpha\,\mathrm{Vol}(V/\mathcal{L}_2)$. We finish by proving:

Claim. $\alpha = [\mathcal{L}_2 : \mathcal{L}_1]$:

Let $v \in \mathcal{L}_2$. Using (i), we see that there exist $x_i \in \mathbb{Z}$, $0 \le x_i < c_{ii}$, satisfying

$$(ii) \qquad v \equiv x_1v_1 + \cdots + x_nv_n \mod \mathcal{L}_1.$$

It follows that $\alpha \ge [\mathcal{L}_2 : \mathcal{L}_1]$. (as the number of possible x_i is c_{ii} all i.)

To show that $\alpha = [\mathcal{L}_2 : \mathcal{L}_1]$, it suffices to show that all the v's in (ii) give different cosets, i.e., given

$$v = \sum_{i=1}^n x_iv_i, \quad v' = \sum_{i=1}^n x_i'v_i \text{ in } \mathcal{L}_2 \text{ and } v \equiv v' \mod \mathcal{L}_1$$
$$\text{with } x_i, x_i' \in \mathbb{Z} \text{ satisfying } 0 \le x_i, x_i' < c_{ii},$$

we must show $v = v'$. But if this is the case, $v - v' = \sum_{j=1}^n b_jw_j$ for some $b_j \in \mathbb{Z}$, hence

$$(iii) \qquad v - v' = \sum_{i=1}^n (x_i - x_i')v_i = \sum_{j=1}^n b_jw_j = \sum_{j=1}^n b_j\Big(\sum_{i=1}^n c_{ji}v_i\Big)$$

by (i). If $v \ne v'$, then there exists a minimal s satisfying $x_s \ne x_s'$. Then by (iii), we have $b_j = 0$ for all $j < s$ and $c_{ss}b_s = x_s - x_s' \ne 0$. However, $0 < |x_s - x_s'| < c_{ss} \le |c_{ss}b_s|$, a contradiction. So $v = v'$ and the claim, hence the lemma, is proven. □

The lemma allows us to establish the following important computation.

Proposition 89.7. *Let K be a number field of degree n, $V = \mathbb{R}^n$ (or $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$), and $\mathfrak{A} \in I_{\mathbb{Z}_K}$. Then*

$$\mu(D_{\mathfrak{A}}) = \mathrm{Vol}(V/\mathfrak{A}) = \mathfrak{N}(\mathfrak{A})2^{-r_2}\sqrt{|d_K|}.$$

Proof. Let $\{w_1, \dots, w_n\}$ be a basis for K and $V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$. Write $w_i^{(r_1+j)} = x_i^{(r_1+j)} + \sqrt{-1}y_i^{(r_1+j)}$, $x_i^{(r_1+j)}, y_i^{(r_1+j)} \in \mathbb{R}$ for $j = 1, \dots, r_2$. Set $\Delta = \Delta_{K/\mathbb{Q}}(w_1, \dots, w_n)$ and

$\mathcal{L} = \mathcal{L}(w_1, \ldots, w_n)$. Then we have

$$\begin{aligned}
\Delta &= \det \begin{pmatrix} w_1^{(1)} & \cdots & w_n^{(1)} \\ \vdots & & \vdots \\ w_1^{(r_1)} & & w_n^{(r_1)} \\ x_1^{(r_1+1)} + \sqrt{-1}y_1^{(r_1+1)} & \cdots & x_n^{(r_1+1)} + \sqrt{-1}y_n^{(r_1+1)} \\ \vdots & & \vdots \\ x_1^{(r_1+1)} - \sqrt{-1}y_1^{(r_1+1)} & \cdots & x_n^{(r_1+1)} - \sqrt{-1}y_n^{(r_1+1)} \\ \vdots & & \vdots \end{pmatrix}^2 \\
&= (-2\sqrt{-1})^{2r_2} \det \begin{pmatrix} w_1^{(1)} & \cdots & w_n^{(1)} \\ \vdots & & \vdots \\ x_1^{(r_1+1)} & \cdots & x_n^{(r_1+1)} \\ \vdots & & \vdots \\ y_1^{(r_1+1)} & \cdots & y_n^{(r_1+1)} \\ \vdots & & \vdots \end{pmatrix}^2 \\
&= (-1)^{r_2} 2^{2r_2} \operatorname{Vol}(V/\mathcal{L})^2
\end{aligned}$$

Therefore, (computing $\operatorname{Vol}(V/\mathcal{L})$ using the standard basis), we have $\operatorname{Vol}(V/\mathcal{L}) = 2^{-r_2}\sqrt{|\Delta|}$. If $\{w_1, \ldots, w_n\}$ is an integral basis for $\mathbb{Z}_K$, then $\operatorname{Vol}(V/\mathbb{Z}_K) = 2^{-r_2}\sqrt{|d_K|}$. If $0 < \mathfrak{A} \subset \mathbb{Z}_K$ is an ideal, by Lemma 89.6,

$$\operatorname{Vol}(V/\mathfrak{A}) = [\mathbb{Z}_K : \mathfrak{A}]\, 2^{r_2} \sqrt{|d_K|} = \mathfrak{N}(\mathfrak{A})\, 2^{r_2} \sqrt{|d_K|}.$$

Finally, suppose that $\mathfrak{A} \in I_{\mathbb{Z}_K}$. By Remark 79.14(1), there exists a nonzero integer m satisfying $m\mathfrak{A} \subset \mathfrak{A}$. Then

$$m^n \operatorname{Vol}(V/\mathfrak{A}) = \operatorname{Vol}(V/m\mathfrak{A}) = \mathfrak{N}(m\mathfrak{A})2^{r_2}\sqrt{|d_K|} = m^n\mathfrak{N}(\mathfrak{A})2^{r_2}\sqrt{|d_K|}.$$

The result follows. □

In order to prove our goal, we shall need the following:

Lemma 89.8. *Let V be a finite dimensional real inner product space and $S \subset V$ be measurable. If $\mathcal{L}$ is a full lattice in V with $\mu(S) > \mu(D_{\mathcal{L}})$, then there exist $s_1, s_2 \in S$ satisfying $0 \neq s_1 - s_2 \in \mathcal{L}$.*

Proof. The set S is an infinite set, since $\mu(S) > 0$. Suppose that for all $s_1 \neq s_2$ in S, we have $s_1 - s_2 \notin \mathcal{L}$. It follows that the the sets

$$(*) \qquad S - \lambda,\ \lambda \in \mathcal{L} \text{ are all disjoint.}$$

Since the measure μ is translation invariant, for all $\lambda \in \mathcal{L}$, we have $\big((S-\lambda) \cap D_{\mathcal{L}}\big) + \lambda = (D_{\mathcal{L}} + \lambda) \cap S$. (Check.) It therefore follows that $\mu\big((S-\lambda)\cap D_{\mathcal{L}}\big) = \mu\big((D_{\mathcal{L}}+\lambda)\cap S\big)$. As $V = \bigvee_{\mathcal{L}}(D_{\mathcal{L}} + \lambda)$, we have $S = \bigvee_{\mathcal{L}} S \cap (D_{\mathcal{L}} + \lambda)$. Hence

$$\mu(S) = \sum_{\mathcal{L}} \mu\big(S \cap (D_{\mathcal{L}} + \lambda)\big) = \sum_{\mathcal{L}} \big((S-\lambda) \cap D_{\mathcal{L}}\big) \leq \mu(D_{\mathcal{L}})$$

by (*), a contradiction. □

Definition 89.9. Let V be a finite dimensional real inner product space and $X \subset V$. We say:

(i) X is *symmetric* if $x \in X$, then $-x \in X$.

(ii) X is *convex* if $x, y \in X$, then $tx + (1-t)y \in X$ for $0 \leq t \leq 1$, (i.e, the line segment from x to y lies in X).

We now establish the goal of this section.

Theorem 89.10. (Minkowski Lattice Point Theorem) *Let V be an n-dimensional real inner product space and $\mathcal{L}$ be a full lattice in V. Suppose that $X \subset V$ is a symmetric, convex, measurable set (respectively, and also compact) and satisfies $\mu(X) > 2^n\mu(D_{\mathcal{L}})$ (respectively, and $\mu(X) \geq 2^n\mu(D_{\mathcal{L}})$). Then $X \cap \mathcal{L} \neq \{0\}$.*

Proof. Suppose that $\mu(X) > 2^n\mu(D_{\mathcal{L}})$. Set

$$S := \frac{X}{2} = \{\frac{x}{2} \mid x \in X\}, \text{ so } \mu(S) = \frac{1}{2^n}\mu(X) > \mu(D_{\mathcal{L}}).$$

By Lemma 89.8, there exist $s_1, s_2 \in S$ such that $0 \neq s_1 - s_2$ lies in $\mathcal{L}$. Let $s_i = \frac{x_i}{2}$ with $\frac{x_i}{2} \in X$ for $i = 1, 2$. So $s_1 - s_2 = \frac{x_1 - x_2}{2}$. By symmetry, $-x_2 \in X$ and by convexity, $s_1 - s_2$, the midpoint of the line segment connecting x_1 and $-x_2$, lies in X. We are, therefore, done unless X is compact and $\mu(x) = 2^n\mu(D_{\mathcal{L}})$. Suppose this is the case. Then for each $\epsilon > 0$, the set $(1+\epsilon)X$ is compact, symmetric, convex, and measurable. Therefore, by what we have seen, there exists a point $0 \neq \lambda_\epsilon \in (1+\epsilon)X \cap \mathcal{L}$. As $(1+\epsilon)X \cap \mathcal{L}$ is a finite set for all $\epsilon > 0$ by Proposition 89.5, the set $X = \bigcap_{\epsilon > 0}(1+\epsilon)X$ must contain some point $0 \neq \lambda \in X \cap \mathcal{L}$. □

We give a few applications of the Minkowski Lattice Point Theorem.

Application 89.11. Let $f_i = \sum_{j=1}^n a_{ij}t_j \in \mathbb{R}[t_1, \dots, t_n]$, $i = 1, \dots, n$, be linear forms (i.e., homogeneous of degree one), $\alpha = \det\big((a_{ij})\big) \neq 0$, and $c_1, \dots, c_n > 0$ in $\mathbb{R}$. Suppose that $c_1 \cdots c_n \geq |\alpha|$. Then there exists a nonzero $x \in \mathbb{Z}^n$ satisfying $|f_i(x)| \leq c_i$ for $i = 1, \dots, n$. (E.g., let $c_i = |\alpha|^{\frac{1}{n}}$.)

Proof. Let $\mathcal{L} = \mathbb{Z}^n$ and $X = \{x \in \mathbb{R}^n \mid |f_i(x)| \leq c_i\}$ for $i = 1, \dots, n$. Then X is symmetric, convex, compact, and

$$\begin{aligned}
\mu(X) &= \int_{|f_1| \leq c_1} \cdots \int_{|f_n| \leq c_n} dx_1 \cdots dx_n \\
&= \int_{|f_1| \leq c_1} \cdots \int_{|f_n| \leq c_n} \left|\frac{\partial(x_1, \dots, x_n)}{\partial(f_1, \dots, f_n)}\right| df_1 \cdots df_n \\
&= \frac{1}{|\alpha|} \int_{|f_1| \leq c_1} \cdots \int_{|f_n| \leq c_n} df_1 \cdots df_n \\
&= \frac{2^n}{|\alpha|} c_1 \cdots c_n \geq 2^n = 2^n\mu(D_{\mathcal{L}}).
\end{aligned}$$

The result follows from the Minkowski Lattice Point Theorem. □

Recall that the *gamma function* $\Gamma(s)$ for $\mathrm{Re}(s) > 0$ is defined by $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$. We have $\Gamma(s+1) = s\Gamma(s)$, $\Gamma(n+1) = n!$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and $\dfrac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$ is the volume of the n-ball for any integer $n > 0$.

Application 89.12. Let $f_i = \sum_{j=1}^n a_{ij}t_j \in \mathbb{R}[t_1, \dots, t_n]$, $i = 1, \dots, n$, be linear forms and $\alpha = \det\big((a_{ij})\big) \neq 0$. Then there exists a nonzero $x \in \mathbb{Z}^n$ satisfying $f_1^2(x) + \cdots + f_n^2(x) \leq \frac{4}{\pi}\left(|\alpha|\,\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}\right)$.

Proof. Let $X = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n f_i^2(x) \leq c^2\}$, a compact, symmetric, convex ellipsoid. Let $\mathcal{L} = \mathbb{Z}^n$ and $c = \big(2^n\Gamma(\frac{n}{2}+1)|\alpha|/\pi^{\frac{n}{2}}\big)^{\frac{1}{n}}$, then

$$
\begin{aligned}
\mu(X) &= \underset{\sum_{i=1}^n f_i^2 \leq c^2}{\int \cdots \int} dx_1 \cdots dx_n = \underset{\sum_{i=1}^n f_i^2 \leq c^2}{\int \cdots \int} \Big|\frac{\partial(x_1, \dots, x_n)}{\partial(f_1, \dots, f_n)}\Big| df_1 \cdots df_n \\
&= \frac{1}{|\alpha|} \underset{\sum_{i=1}^n f_i^2 \leq c^2}{\int \cdots \int} df_1 \cdots df_n = \frac{c^n}{|\alpha|} \underset{\sum_{i=1}^n y_i^2 \leq 1}{\int \cdots \int} dy_1 \cdots dy_n \\
&= \frac{c^n}{|\alpha|}\Big(\frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\Big) = 2^n.
\end{aligned}
$$

The result follows from the Minkowski Lattice Point Theorem. $\square$

Remark 89.13. Let $Q(t_1, \dots, t_n) = \sum_{i,j=1}^n b_{ij}t_it_j \in \mathbb{R}[t_1, \dots, t_n]$ be a *quadratic form* written with $b_{ij} = b_{ji}$ for all i, j (which can always be arranged as $2 \neq 0$). We say Q is *positive definite* if $Q(x) > 0$ for all nonzero $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. By the Principal Axis Theorem in linear algebra, the matrix $B = (b_{ij}) \in \mathrm{GL}_n(\mathbb{R})$ is diagonalizable with positive eigenvalues (which are squares). In particular, there exists a matrix $A = (a_{ij})$ in $\mathrm{GL}_n(\mathbb{R})$ such that $B = A^tA$. Let $f_i = \sum_{i=1}^n a_{ij}t_j$ in $\mathbb{R}[t_1, \dots, t_n]$, linear forms, and $T^t = (t_1 \cdots t_n)$. Then we have $Q = T^tBT = f_1^2 + \cdots + f_n^2$. The real number $\det B = (\det A)^2$ is called the *discriminant* of Q and written $\mathrm{disc}\, Q$.

So we have:

Application 89.14. Let $Q(t_1, \dots, t_n) = \sum_{i,j=1}^n b_{ij}t_it_j \in \mathbb{R}[t_1, \dots, t_n]$ be a positive definite quadratic form. Then there exists a nonzero $x \in \mathbb{Z}^n$ satisfying

$$
Q(x) \leq \frac{4}{\pi}\big(\sqrt{\mathrm{disc}\, Q}\ \Gamma(\frac{n}{2}+1)\big)^{\frac{2}{n}}.
$$

We use this to prove a classical theorem of Eisenstein-Hermite. Let $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a symmetric bilinear form, $\mathcal{L} = \mathcal{L}(v_1, \dots, v_n)$ a full lattice in $\mathbb{R}^n$ satisfying $b|_{\mathcal{L}\times\mathcal{L}} : \mathcal{L}\times\mathcal{L} \to \mathbb{Z}$ with $B_b := \big((b(v_i, v_j)\big)$ in $\mathrm{GL}_n(\mathbb{R})$, i.e., $\det B_b \neq 0$. We then say that $(\mathcal{L}, b)$ is a $\mathbb{Z}$-*bilinear space*. We call such a bilinear space *positive definite* if $Q_b = T^tB_bT$ is positive definite. This is equivalent to $q_b : V \to \mathbb{R}$ defined by $q_b(x) = b(x, x) > 0$ for all $0 \neq x \in \mathbb{R}^n$. We call n the *rank* of $\mathcal{L}$. and $\det B_b$ the *discriminant* of $(\mathcal{L}, b)$. It is unique up to a $\mathbb{Z}$-*isometry* (i.e., a $\mathbb{Z}$-isomorphism preserving b) as a change of basis matrix has determinant ± 1).

Note that if $(\mathcal{L}, b)$ is a positive definite $\mathbb{Z}$-bilinear space, there exists a nonzero $x \in \mathbb{Z}^n$ satisfying Application 89.14.

Theorem 89.15. (Eisenstein-Hermite) *For each pair of integers (n, d) with $n > 0$, there exist finitely many positive define $\mathbb{Z}$-bilinear spaces of rank n and discriminant d up to isometry.*

PROOF. Let $(\mathcal{L}, b)$ be a positive definite $\mathbb{Z}$-bilinear space of rank n and discriminant $d = \operatorname{disc}\mathcal{L}$. Set $c(n, d) = \frac{4}{\pi}\big(\Gamma(\frac{n}{2}+1)^{\frac{2}{n}}\big)\sqrt{d}$. By Application 89.14, there exists a nonzero $x \in \mathcal{L}$ that satisfies $0 < q(x) := b(x, x) \le c(n, d)$. Let

$$\mathcal{L}_0 := \{y \in \mathcal{L} \mid b(x, y) \equiv 0 \mod q(x)\} \subset \mathcal{L}.$$

Clearly, we have $[\mathcal{L} : \mathcal{L}_0] \le q(x)$.

Claim. $\mathcal{L}_0 = \mathbb{Z}x \oplus (\mathbb{Z}x)^\perp$ over $\mathbb{Z}$ (with the obvious definition):

As $(\mathcal{L}, b)$ is positive definite, we have $\mathbb{Z}x \cap (\mathbb{Z}x)^\perp = 0$. Let $y \in \mathcal{L}_0$. By definition, $b(x, y) = aq(x)$ for some $a \in \mathbb{Z}$ and $b(x, y - ax) = 0$. Consequently, $y = (y - ax) + ax$ lies in $(\mathbb{Z}x)^\perp + \mathbb{Z}x$ proving the claim.

In the notation of Remark 89.13, we had $B_b = A^t A$. This translates into

$$\operatorname{disc}\mathcal{L}_0 = \mu(D_{\mathcal{L}_0})^2 = [\mathcal{L} : \mathcal{L}_0]^2 \mu(D_{\mathcal{L}})^2 = [\mathcal{L} : \mathcal{L}_0]^2 \operatorname{disc}\mathcal{L}$$

using Lemma 89.6. It follows that

$$\operatorname{disc}\mathbb{Z}x \ \operatorname{disc}(\mathbb{Z}x)^\perp = \operatorname{disc}\mathcal{L}_0 \le q(x)^2 d \le c(n, d)^2 d.$$

By induction on $\operatorname{rank}\mathcal{L}$, there exist finitely many positive definite $\mathbb{Z}$-bilinear spaces $(\mathbb{Z}x)^\perp$, $\mathbb{Z}x$ (up to isometry) of discriminant all bounded by $c(n, d)^2 d$. Hence there exist finitely many positive definite $\mathbb{Z}$-bilinear spaces $\mathcal{L}_0$ (up to isometry). Since $\mathcal{L}/\mathcal{L}_0$ is a finite group and b is determined by its values on a basis for $\mathcal{L}$, there exist only finitely many possible $(\mathcal{L}, b)$ (up to isometry). $\square$

We shall see in Corollary 91.14 below that there exist finitely many number fields K with fixed discriminant d.

Exercises 89.16. 1. Use the Minkowski Lattice Point Theorem to prove Fermat's Theorem that positive primes congruent to one modulo four are sums of two squares.

2. Use the Minkowski Lattice Point Theorem to prove Lagrange's Theorem that every positive integer is a sum of four squares.
3. Prove an analogous result as in the Eisenstein-Hermite Theorem for indefinite $\mathbb{Z}$-bilinear spaces, i.e., those $\mathbb{Z}$-bilinear spaces $(\mathcal{L}, b)$ in which there exist $x, y \in \mathcal{L}$ satisfying $b(x, x) < 0$ and $b(y, y) > 0$.
4. Let $f_i = \sum_{j=1}^{n} a_{ij} t_j \in \mathbb{R}[t_1, \dots, t_n]$, $i = 1, \dots, n$, be linear forms, $\alpha = \det\big((a_{ij})\big) \ne 0$, and $r \in \mathbb{R}$. Show there exists a nonzero $x \in \mathbb{Z}^n$ satisfying $\sum_{i=1}^{n} |f_i(x)| \le r$ if $r \ge (n!|\alpha|)^{\frac{1}{n}}$.

90. Units in a Ring of Algebraic Numbers

In Section 86, we computed the units in the ring of algebraic numbers in a quadratic number field $\mathbb{Q}(\sqrt{m})$, m a square-free integer. In this section, we generalize this result by proving the Dirichlet Unit Theorem that determines the units in $\mathbb{Z}_K$ for any algebraic number field K. We continue to use the notation of Section 89.

We begin with a computation.

Lemma 90.1. (Minkowski) *Let $A = (a_{ij}) \in \mathbb{M}_n(\mathbb{R})$ satisfy $a_{ij} < 0$ for all $i, j = 1, \dots, n$ and $i \neq j$. Suppose that $\sum_{j=1}^n a_{ij} > 0$ for each $i = 1, \dots, n$. Then $A \in \mathrm{GL}_n(\mathbb{R})$.*

Proof. (Artin). Suppose that $\det A = 0$. Let $v_1, \dots, v_n$ denote the columns of A. Then $v_1, \dots, v_n$ are linearly dependent, so there exist $x_1, \dots, x_n \in \mathbb{R}$ not all zero such that $\sum_{i=1}^n x_i v_i = 0$, i.e., $\sum_{i=1}^n x_i a_{li} = 0$ for $l = 1, \dots, n$. Choose l such that $|x_l|$ is maximal. Multiplying by -1 if necessary, we may assume that $x_l > 0$. Since $x_l \geq x_i$ and $a_{li} < 0$ for all $i \neq l$, $i = 1, \dots, n$, we have

$$0 = x_l a_{ll} + \sum_{\substack{i=1 \\ i \neq l}}^n x_i a_{li} \geq x_l a_{ll} + \sum_{\substack{i=1 \\ i \neq l}}^n x_l a_{li} = x_l \sum_{i=1}^n a_{li} > 0,$$

a contradiction. □

Construction/Definition 90.2. Let K be a number field of degree n. As we have n embeddings of K with $\sigma_{r_1+r_2+j} = \overline{\sigma}_{r_1+j}$ for $j = 1, \dots, r_2$, we set

$$e_i := \begin{cases} 1 & \text{if } 1 \leq i \leq r_1 \\ 2 & \text{if } 1 \leq r_1 + 1 \leq i \leq r_1 + r_2 \end{cases}$$

(with r_1 or $r_2 = 0$ possible), as we shall be interested in norms from K.
If $x \in K$, define

$$\ell_i(x) = e_i \log |x^{(i)}|,\ 1 \leq i \leq r_1 + r_2$$

and the *logarithmic map*

$$\ell : K^\times \to \mathbb{R}^{r_1+r_2} \text{ by } x \mapsto \big(\ell_1(x), \dots, \ell_{r_1+r_2}(x)\big).$$

In particular, setting

$$U_K := \mathbb{Z}_K^\times, \text{ we have if } \varepsilon \in U_K, \text{ then } \mathrm{N}_{K/\mathbb{Q}}(\varepsilon) = \prod_{i=1}^n \varepsilon^{(i)} = \pm 1.$$

Taking $\log(\prod_{i=1}^n |\varepsilon^{(i)}|)$, we get

$$0 = \sum_{i=1}^{r_1+r_2} e_i \log |\varepsilon^{(i)}|. \tag{90.3}$$

It follows that $\{\log |\varepsilon^{(1)}|, \dots, \log |\varepsilon^{(r_1+r_2)}|\}$ are linearly dependent. We eliminate $\log |\varepsilon^{(r_1+r_2)}|$ from this linearly dependent set. i.e., $\varepsilon^{(r_1+r_2)}$ and set $r = r_1 + r_2 - 1$. So $\ell(U_K) \subset \mathbb{R}^r$ lying on the hyperspace in $\mathbb{R}^{r_1+r_2}$ defined by $0 = \sum_{i=1}^{r_1+r_2} t_i$ in $\mathbb{R}^{r+1}$.

Lemma 90.4. *Let K be a number field of degree n and $r = r_1 + r_2 - 1$. Suppose that $C \subset \mathbb{R}^{r_1+r_2}$ is a compact set. Then $|\ell^{-1}(C \cap \ell(U_K))| < \infty$. In particular, $\ell(U_K) \subset \mathbb{R}^r$ is a lattice (although not a priori full).*

Proof. We have C is compact and equation (90.3) holds. It follows that if $\ell(\varepsilon) \in C$ for $\varepsilon \in U_K$ then all the coordinates of such an $\ell(\varepsilon)$ are uniformly bounded, i.e., there exists an $M_0 > 0$ such that $\ell_i(\varepsilon) \leq M_0$ for $i = 1, \dots, r_1 + r_2$ for any $\varepsilon \in U_K$ satisfying

$\ell(\varepsilon) \in C$. In particular, $|\varepsilon^{(i)}|^{e_i} < e^{M_0}$, for $i = 1, \dots, r_1 + r_2$. As $M_1 = e^{M_0} > 1$, we have $|\varepsilon^{(i)}| < M_1$, for $i = 1, \dots, n$. Let $Y = \{x \in \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \mid |x_i| \leq M_1\}$, a compact subset of $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ containing $i_{\mathbb{C}}\Big(\ell^{-1}\big(C \cap \ell(U_K)\big)\Big)$. Since $i_{\mathbb{C}}(\mathbb{Z}_K)$ is a lattice in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$, we have $\Big|i_{\mathbb{C}}\Big(\ell^{-1}\big(C \cap \ell(U_K)\big)\Big)\Big| < \infty$ by Proposition 89.5. Hence $\ell(U_K) \subset \mathbb{R}^r$ is a lattice by Proposition 89.5. $\square$

Corollary 90.5. *Let K be a number field. Then $\ker \ell$ is a finite cyclic group.*

PROOF. Any finite subgroup of $K^\times$ is cyclic. $\square$

Notation 90.6. Let K be a number field. Let $W_K := \ker \ell$, and (as usual) let μ_K denote the roots of unity in K.

We can now classify the units in a ring of algebraic integers.

Theorem 90.7. (Dirichlet Unit Theorem) *Let K be a number field of degree $n = r_1 + 2r_2$ and $r = r_1 + r_2 - 1$. Let $\ell : U_K \to \mathbb{R}^r$ be the logarithmic map given by $\varepsilon \mapsto \big(\ell_1(\varepsilon), \dots, \ell_r(\varepsilon)\big)$. Then*

(1) $W_K = \mu_K$ is a finite cyclic group.
(2) $\ell(U_K) \subset \mathbb{R}^r$ is a full lattice.
(3) $U_K \cong W_K \times \mathbb{Z}^r$.

PROOF. (1): If $w \in W_K$, then $w^{|W_K|} = 1$ by Corollary 90.5, so $W_K \subset \mu_K$. Conversely if $x \in U_K$ satisfies $x^m = 1$ for some $m > 0$, then $\ell_i(x) = 0$ for $i = 1, \dots, n$, hence $x \in W_K$.
(2): We know that $\ell(U_K) \subset \mathbb{R}^r$ is a lattice by Lemma 90.4, so we need to show that it is a full lattice. To do this it suffices to find $\varepsilon_1, \dots, \varepsilon_r$ in U_K such that $\mathcal{L}\big(\ell(\varepsilon_1), \dots, \ell(\varepsilon_r)\big)$ is a full lattice in $\mathbb{R}^r$.
Let $V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ and define $\mathrm{N} : V \to \mathbb{R}$ by

$$\mathrm{N}(x) = x_1 \cdots x_{r_1} x_{r_1+1} \cdots x_{r_1+1} \cdots x_{r_1+r_2} \overline{x}_{r_1+1} \cdots , \overline{x}_{r_1+r_2}.$$

Viewing $i_{\mathbb{C}} : K \hookrightarrow V$ as an inclusion, we have $\mathrm{N}|_K = \mathrm{N}_{K/Q}$. Let $\{w_1, \dots, w_n\}$ be an integral basis for $\mathbb{Z}_K$. Therefore, $\mathbb{Z}_K \hookrightarrow V$ is a full lattice in V viewing $w_j = (w_j^{(1)}, \dots, w_j^{(r_1+r_2)})$. In particular, V is a $\mathbb{Z}_K$-module via $\alpha(v_1, \dots, v_{r_1+r_2}) = (\alpha^{(1)} v_1, \dots, \alpha^{(r_1+r_2)} v_{r_1+r_2})$ for all $\alpha \in \mathbb{Z}_K$. It follows if $x \in V$ satisfies $\mathrm{N}(x) \neq 0$, then $\mathcal{L}_x := \mathbb{Z}_K x = \mathbb{Z} w_1 x \oplus \cdots \oplus \mathbb{Z} w_n x$ is a full lattice in V. Using an analogous calculation that we did to prove Proposition 89.7, we see that if $\mathrm{N}(x) \neq 0$, then

$$\mathrm{Vol}(V/\mathcal{L}_x) = |\,\mathrm{N}(x)|\, 2^{-r_2} \sqrt{|d_K|}.$$

In particular, if $Nx = 1$, then

$$\mathrm{Vol}(V/\mathcal{L}_x) = 2^{-r_2} \sqrt{|d_K|}$$

is independent of $x \in V$ satisfying $\mathrm{N}(x) = 1$.

Claim. Let X be a compact, convex, symmetric, measurable subset of V with $\mu(X) > 2^{r_2} \sqrt{d_K}$. Then

(1) There exists $M_0 = M_0(X) > 0$ such that if $\beta \in X$, $|\beta_i| < M_0$, $i = 1, \dots, n$.
(2) If $x \in V$ satisfying $\mathrm{N}(x) = 1$, then there exists $0 \neq \alpha \in \mathbb{Z}_K$ such that $\alpha x \in \mathcal{L}_x \cap X$ and an $M = M(X) > 0$ independent of x satisfying $|\,\mathrm{N}_{K/\mathbb{Q}}(\alpha)| < M$:

The first statement is immediate as X is compact. For the second, fix $x \in V$ that satisfies $\mathrm{N}(x) = 1$. By the Minkowski Lattice Point Theorem, there exists $0 \neq \alpha \in \mathbb{Z}_K$ satisfying $\alpha x \in \mathcal{L}_x \cap X$. As X is compact, by (1) all the coordinates of αx are uniformly bounded by a constant $M_0 = M_0(X) > 0$. Since $\mathrm{N}(\alpha x) = \mathrm{N}_{K/\mathbb{Q}}(\alpha)\,\mathrm{N}(x) = \mathrm{N}_{K/\mathbb{Q}}(\alpha)$, $M = M_0^n$ works and the claim is proven.

By Proposition 88.8. we have

$$\mathrm{N}_{K/Q}(\alpha)\mathbb{Z} = \mathrm{N}_{K/Q}(\alpha\mathbb{Z}_K) = \prod_{\mathfrak{p}} N_{K/Q}(\mathfrak{p})^{v_{\mathfrak{p}}(\alpha)} = \prod_{\mathfrak{p}} \mathfrak{N}(\mathfrak{p})^{v_{\mathfrak{p}}(\alpha)}\mathbb{Z}.$$

Consequently, if $0 \neq \alpha \in \mathbb{Z}_K$ satisfies $|\,\mathrm{N}_{K/Q}(\alpha)| < M$, then $\mathfrak{N}(\mathfrak{p}) < M$ for all prime ideals $\mathfrak{p}$ such that $v_{\mathfrak{p}}(\alpha) > 0$ and the allowable values $v_{\mathfrak{p}}(\alpha) > 0$ are also bounded. It follows that there exist finitely many principal integral ideals $\alpha\mathbb{Z}_K$ (so $\alpha \in \mathbb{Z}_K$) satisfying $0 < |\,\mathrm{N}_{K/\mathbb{Q}}(\alpha)| < M$.

Let

$$\alpha_1\mathbb{Z}_K, \dots, \alpha_s\mathbb{Z}_K$$

be all the finitely many nonzero principal integral ideals that satisfy $|\,\mathrm{N}_{K/Q}(\alpha_i)| < M$ for $i = 1, \dots, s$.

[Note that if $|U_K|$ is infinite, then $|\,\mathrm{N}_{K/\mathbb{Q}}(\alpha u)| = |\,\mathrm{N}_{K/\mathbb{Q}}(\alpha)|$ for any algebraic integer α and any $u \in U_K$. In particular, in the above, it is crucial that we deal with principal ideals rather than elements.]

Let α be any element that satisfies the claim. Then there exists k, $1 \leq k \leq s$, such that $\alpha\mathbb{Z}_K = \alpha_k\mathbb{Z}_K$ and an $\varepsilon \in U_K$ with $\alpha = \varepsilon\alpha_k$.

Therefore, for each x satisfying $\mathrm{N}(x) = 1$, we have:

(90.8)
$$\begin{gathered}\text{There exists an } \varepsilon \in U_K \text{ and a } k,\ 1 \leq k \leq s, \text{ with } \varepsilon\alpha_k x \in X.\\ \text{In particular, } |\varepsilon^{(i)}\alpha_k^{(i)}x_i| < M_0,\ i = 1, \cdots, r_1 + r_2.\end{gathered}$$

Fix j for $1 \leq j \leq r_1 + r_2$ and choose

$$x \in V \text{ with } \mathrm{N}(x) = 1 \text{ and } |x_i| >> 0 \text{ for all } i \neq j,\ i = 1, \dots, r_1 + r_2.$$

Let ε_j be the corresponding unit in (90.8).

Since there exist finitely many $\alpha_k^{(i)}$, in (90.8), we may assume that we have chosen x with the $|x_i|$, $i \neq j$ so large, $1 \leq i \leq r_1 + r_2$, that 90.8 also implies

(*)
$$|\varepsilon_j^{(i)}| < 1 \text{ i.e., } \ell_i(\varepsilon_j) < 0 \text{ for all } 1 \leq i \leq r_1 + r_2.$$

For each $j \leq r$, choose such $x \in V$ and corresponding ε_j satisfying (90.8) and (*). Set

$$\beta = \begin{pmatrix} \ell_1(\varepsilon_1) & \cdots & \ell_r(\varepsilon_1) \\ \vdots & & \vdots \\ \ell_1(\varepsilon_r) & \cdots & \ell_r(\varepsilon_r) \end{pmatrix}.$$

As $\ell_i(\varepsilon_j) < 0$ for $i \neq j$ and

$$0 = \sum_{i=1}^{r_1+r_2} \ell_i(\varepsilon_j) = \sum_{i=1}^{r} \ell_i(\varepsilon_j) + \ell_{r+1}(\varepsilon_j),$$

we have for all $j \leq r$,

$$\sum_{i=1}^{r} \ell_i(\varepsilon_j) = -\ell_{r+1}(\varepsilon_j).$$

By Minkowski's Lemma 90.1, $\det \beta \neq 0$. Therefore, $\ell(\varepsilon_1), \ldots, \ell(\varepsilon_r)$ are linearly independent as needed.

(3): Since $\ell(U_K)$ is $\mathbb{Z}$-free, (3) follows from the split exact sequence

$$1 \to W_K \to U_K \to \ell(U_K) \to 1. \qquad \square$$

Remark 90.9. Let K be a number field. The proof of the Dirichlet Unit Theorem shows how to get r linearly independent units generating a subgroup of finite index in U_K. To get all units, one must use some type of descent. This is usually not easy.

Definition 90.10. Let K be a number field of degree n over $\mathbb{Q}$ and $\{\varepsilon_1, \ldots, \varepsilon_r\}$ a basis for the free part of U_K. We call this basis a set of *base units*. Therefore, if $\upsilon \in U_K$, there exist unique $v_i \geq 0$, $i = 1, \ldots, r$, and a unit $w \in W_K$ such that $\upsilon = w\varepsilon^{v_1} \cdots \varepsilon^{v_r}$. We let

$$R_K := |\det\big(\ell_i(\varepsilon_j)\big)|, \text{ for } 1 \leq i, j \leq r$$

called the *regulator* of K.

Remark 90.11. Let K be a number field of degree n over $\mathbb{Q}$.

1. R_K is independent of the choice of the last embedding that we drop.

   PROOF. Let $A =$

$$\begin{pmatrix} \ell_1(\varepsilon_1) & \cdots & \ell_r(\varepsilon_1) & \ell_{r+1}(\varepsilon_1) \\ \vdots & & \vdots & \\ \ell_1(\varepsilon_r) & \cdots & \ell_r(\varepsilon_r) & \ell_{r+1}(\varepsilon_r) \\ e_1 & & e_r & 0 \end{pmatrix},$$

   with

$$e_i := \begin{cases} 1 & \text{if } 1 \leq i \leq r_1 \\ 2 & \text{if } 1 \leq r_1 < i \leq r_1 + r_2 \end{cases}$$

   as usual. Then A and the matrices

$$\begin{pmatrix} \ell_1(\varepsilon_1) & \cdots & \ell_r(\varepsilon_1) & \sum_{i=1}^{r+1} \ell_i(\varepsilon_1) \\ \vdots & & \vdots & \\ \ell_1(\varepsilon_r) & \cdots & \ell_r(\varepsilon_r) & \sum_{i=1}^{r+1} \ell_i(\varepsilon_r) \\ e_1 & & e_r & \sum_{i=1}^{r+1} e_i \end{pmatrix} \text{ and } \begin{pmatrix} \ell_1(\varepsilon_1) & \cdots & \ell_r(\varepsilon_1) & 0 \\ \vdots & & \vdots & \\ \ell_1(\varepsilon_r) & \cdots & \ell_r(\varepsilon_r) & 0 \\ e_1 & & e_r & n \end{pmatrix}.$$

   all have the same determinant. Hence $\det A = (-1)^{n-1} n R_K$. It follows that R_K is independent of the coordinate that is dropped. $\square$

2. R_K is independent of the choice of base units as a change of basis matrix has determinant equal to ± 1.
3. R_K is very important but an elusive invariant.

We next indicate how the Dirichlet Unit Theorem generalizes. We leave the proofs as exercises.

Definition 90.12. Let K be a number field of degree n over $\mathbb{Q}$. We call the set of nonzero prime ideals, $\mathrm{Max}(\mathbb{Z}_K)$, in $\mathbb{Z}_K$, the set of *finite primes* in K and the set $S_\infty := \{\sigma_1, \dots, \sigma_n\}$ of embeddings $K \to \mathbb{C}$ the set of *infinite primes* in K with $\{\sigma_1, \dots, \sigma_{r_1}\}$ the set of *real infinite primes* and $\{\sigma_{r_1+1}, \dots, \sigma_n\}$ the set of *complex infinite primes*. We let $\mathcal{S}_K = \mathrm{Max}(\mathbb{Z}_K) \cup S_\infty(K)$ called the set of *extended primes* of $\mathbb{Z}_K$. [These represent all the necessary completions of K – the finite primes $\mathfrak{p}$ giving rise to the $\mathfrak{p}$-adic completions of K.] If $S \subset \mathcal{S}(K)$ is a finite set, we set

$$K^S := \{a \in K^\times \mid v_{\mathfrak{p}}(a) = 0 \ \text{ for all } \ \mathfrak{p} \in \mathrm{Max}(\mathbb{Z}_K) \setminus S\},$$

called the *group of S-units* of K. Let I_K^S be the group generated by $\mathrm{Max}(\mathbb{Z}_K) \setminus S$ and $I_K(S)$ the group generated by $\mathfrak{p} \in S$. So $U_K = \ker \iota$, where $\iota : K^S \to I_K(S)$ is given by $a \mapsto a\mathbb{Z}_K$. For example, $U_K = K^{S_\infty(K)}$.

In algebraic number theory, one often studies finite extensions E of K with finite sets S containing $S_\infty(K) \cup \{\mathfrak{p} \in \mathrm{Max}\,\mathbb{Z}_K) \mid \mathfrak{p}\,\text{ramifies in}\, E\}$.

Theorem 90.13. (Dirichlet-Chevalley-Hasse Unit Theorem) *Let K be a number field and $S \subset \mathcal{S}(K)$ be a finite set containing $S_\infty(K)$. Then $K^S \cong W_K \times \mathbb{Z}^{|S|-1}$*

Corollary 90.14. *Let K be a number field, ζ_N be a primitive mth root of unity in K, and $S \subset \mathcal{S}(K)$ a finite set containing $S_\infty(K)$. Then $[K^S : (K^S)^N] = N^{|S|}$.*

Exercise 90.15. 1. Prove Theorem 90.13.
2. Prove Corollary 90.14.

91. Minkowski Bound

In this section, we use the Minkowski Lattice Theorem in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ to show that the class number $h_K := |I_{\mathbb{Z}_K}/P_{\mathbb{Z}_K}|$ of a ring of algebraic integers in the number field K is finite. We shall also show that there exist only finitely many algebraic number fields with a given discriminant. We continue to use the notation of Section 89.

We investigate the following subset of $V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$: For any $x \in V$, we shall write $x = (x_1, \dots x_{r_1}, z_{r_1+1}, \dots, z_{r_1+r_2})$. Set

$$\begin{aligned} X(t) = X_{r_1,r_2}(t) &:= \{x \in V \mid \sum_{i=1}^{r_1} |x_i| + 2\sum_{i=r_1+1}^{r_1+r_2} |z_i| \leq t\} \\ &= \{x \in V \mid \sum_{i=1}^{r_1} |x_i| + 2\sum_{i=1}^{r_2} \sqrt{x_{r_1+i}^2 + y_{r_1+i}^2} \leq t\}. \end{aligned}$$

[Note this is the same as

$$X(t) := \{x \in V \mid \sum_{i=1}^{n} |x_i| \leq t,\ x_{r_1+r_2+j} = \overline{x}_{r_1+j} \ \text{ for } \ j = 1, \dots, r_2\}].$$

Lemma 91.1. *$X(t)$ is a compact, convex, symmetric set with measure*

$$\mu\big(X(t)\big) = \frac{2^{r_1}}{n!}\Big(\frac{\pi}{2}\Big)^{r_2} t^n.$$

PROOF. Clearly, $X(t)$ is compact and symmetric. Let $0 \le a \le 1$, $b = 1 - a$ and $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ in $X(t)$. Then $\sum_{i=1}^n |ax_i + by_i| \le \sum_{i=1}^n (a|x_i| + b|y_i|) \le at + bt = t$. Therefore, $X(t)$ is convex. It remains to determine $\mu\big(X(t)\big)$.

Set $X_{0,0}(t) = 1$. Changing variables shows that

$$\mu\big(X_{r_1,r_2}(t)\big) = t^n \mu\big(X_{r_1,r_2}(1)\big), \tag{*}$$

so we need only compute $\mu\big(X_{r_1,r_2}(1)\big)$. Suppose that $r_1 > 0$. Then by (*) and induction, we have

$$\begin{aligned}
\mu\big(X_{r_1,r_2}(1)\big) &= 2\int_0^1 \mu\big(X_{r_1-1,r_2}(1-t)\big)dt \\
&= 2\mu\big(X_{r_1-1,r_2}(1)\big)\int_0^1 (1-t)^{r_1-1+2r_2}dt \\
&= \frac{2}{n}\mu\big(X_{r_1-1,r_2}(1)\big) = \frac{2^{r_1}}{(r_1+2r_2)\cdots(2r_2+1)}\mu\big(X_{0,r_2}(1)\big).
\end{aligned}$$

Therefore, it suffices to compute $\mu\big(X_{0,r_2}(1)\big)$. We have

$$\mu\big(X_{0,r_2}(1)\big) = \iint\limits_{x^2+y^2\le\frac{1}{4}} \mu\big(X_{0,r_2-1}(1-2\sqrt{x^2+y^2})\big)dxdy.$$

Let $\rho = \sqrt{x^2+y^2}$, $x = \rho\cos\theta$ be polar coordinates. Then by (*) and letting $u = 1 - 2\rho$, we have

$$\begin{aligned}
\mu\big(X_{0,r_2}(1)\big) &= \int_{\theta=0}^{\theta=2\pi}\int_{\rho=0}^{\rho=\frac{1}{2}} \mu\big(X_{0,r_2-1}(1-2\rho)\big)\rho d\rho \\
&= 2\pi\mu\big(X_{0,r_2-1}(1)\big)\int_{\rho=0}^{\rho=\frac{1}{2}}(1-2\rho)^{2(r_2-1)}\rho d\rho \\
&= \frac{\pi}{2}\mu\big(X_{0,r_2-1}(1)\big)\int_{u=1}^{u=0} u^{2(r_2-1)}(1-u)du \\
&= \frac{\pi}{2}\mu\big(X_{0,r_2-1}(1)\big)\frac{1}{2r_2(2r_2-1)} = \big(\frac{\pi}{2}\big)^{r_2}\frac{1}{(2r_2)!},
\end{aligned}$$

and we are done. □

Definition 91.2. Define the *Minkowski Bound* to be

$$B_{n,r_2} := \big(\frac{4}{\pi}\big)^{r_2}\frac{n!}{n^n}.$$

If K is a number field of degree n_K, we let $B_K = B_{n_K,r_2^K}$ where $n_K = r_1^K + 2r_2^K$.

Note that $B_K \to 0$ as $n_K \to \infty$.

The key result in this section is the following crucial inequality.

Theorem 91.3. *Let K be a number field and $\mathfrak{A} \in I_{\mathbb{Z}_K}$. Then there exists a nonzero $a \in \mathfrak{A}$ satisfying:*

$$|\operatorname{N}_{K/\mathbb{Q}}(a)| \le B_K\,\mathfrak{N}(\mathfrak{A})\sqrt{|d_K|}.$$

Proof. Let $X(t) = X_{r_1,r_2}(t) \subset V = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ be as above. Choose t to satisfy $\mu\big(X(t)\big) = 2^n\mu(V/\mathfrak{A})$. By Proposition 89.7, we then have

$$\begin{aligned}\frac{2^{r_1}}{n!}\big(\frac{\pi}{2}\big)^{r_2}t^n &= \mu\big(X(t)\big) = 2^n\mu(V/\mathfrak{A})\\ &= 2^n\,\mathfrak{N}(\mathfrak{A})\,2^{-r_2}\sqrt{|d_K|} = 2^{r_1+r_2}\,\mathfrak{N}(\mathfrak{A})\sqrt{|d_K|}.\end{aligned}$$

Therefore,

$$t^n = \big(\frac{4}{\pi}\big)^{r_2}\,n!\,\mathfrak{N}(\mathfrak{A})\sqrt{|d_K|}.$$

By the Minkowski Lattice Point Theorem, there exists a nonzero β in $\mathfrak{A}\cap X(t)$. Therefore, $\sum_{i=1}^n |\beta^{(i)}| \leq t$. Since the geometric mean is bounded by the arithmetic mean, we have $\big(\prod_{i=1}^n |\beta^{(i)}|\big)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n |\beta^{(i)}|$. Hence $|\,\mathrm{N}_{K/\mathbb{Q}}(\beta)| \leq \dfrac{t^n}{n^n} \leq B_K\mathfrak{N}(\mathfrak{A})\sqrt{|d_K|}$ as needed. □

This theorem has important applications.

Definition 91.4. Let K be a number field. Recall the *class group* of $\mathbb{Z}_K$ is $Cl_{\mathbb{Z}_K} := I_{\mathbb{Z}_K}/P_{\mathbb{Z}_K}$. We let $h_K := |Cl_{\mathbb{Z}_K}|$, called the *class number of* K (or $\mathbb{Z}_K$).

Corollary 91.5. *Let K be a number field and $\mathcal{A} \in Cl_{\mathbb{Z}_K}$. Then there exists a nonzero integral ideal $\mathfrak{A} \subset \mathbb{Z}_K$ in $\mathcal{A}$ satisfying $\mathfrak{N}(\mathfrak{A}) \leq B_K\sqrt{|d_K|}$.*

Proof. Let $\mathfrak{B} \in \mathcal{A}$. Then by the theorem, there exists a nonzero $a \in \mathfrak{B}^{-1}$ satisfying $|\,\mathrm{N}_{K/\mathbb{Q}}(a)| \leq B_K\,\mathfrak{N}(\mathfrak{B}^{-1})\sqrt{|d_K|}$. Since $\mathfrak{N}(a\mathbb{Z}_K) = |\,\mathrm{N}_{K/\mathbb{Q}}(a)|$ by Corollary 88.9, we have $\mathfrak{N}(a\mathfrak{B}) \leq B_K\sqrt{|d_K|}$. As $a \in \mathfrak{B}^{-1}$, the fractional ideal $a\mathfrak{B} \subset \mathbb{Z}_K$ is an ideal. It follows that $\mathfrak{A} = a\mathfrak{B} \in \mathcal{A}$ works. □

Corollary 91.6. *Let K be a number field. Then the class number h_K is finite.*

Proof. The inequality $\mathfrak{N}(\mathfrak{A}) \leq B_K\sqrt{|d_K|}$ is satisfied by finitely many ideals $\mathfrak{A} \subset \mathbb{Z}_K$, since only finitely many ideals can contract to any fixed $m\mathbb{Z}$ by the unique factorization of ideals in a Dedekind domain. □

Remark 91.7. The class group of a Dedekind domain may not be finite. For example, the Dedekind domain $E = \mathbb{C}[t, t_1]/\big(t_1^2 - t(t-1(t+1)\big)$, the integral closure of $\mathbb{C}[t]$ in the quotient field of E, corresponding to the points on the elliptic curve $y^2 = x(x-1)(x+1)$ has an infinite class group.

Corollary 91.8. *Let K be a number field. Suppose that for every positive prime integer p satisfying $p \leq B_K\sqrt{|d_K|}$, any prime $\mathfrak{p}$ in $\mathbb{Z}_K$ is principal if it satisfies $p\mathbb{Z} = \mathfrak{p}\cap\mathbb{Z}$. Then $h_K = 1$. In particular, $\mathbb{Z}_K$ is a UFD.*

Proof. Let $\mathcal{A} \in Cl_{\mathbb{Z}_K}$. Choose $\mathfrak{A} \in \mathcal{A}$ an integral ideal satisfying $\mathfrak{N}(\mathfrak{A}) \leq B_K\sqrt{|d_K|}$. Let $\mathfrak{N}(\mathfrak{A}) = p_1 \cdots p_n$ be a factorization into primes in $\mathbb{Z}$. As $\mathfrak{N}$ is multiplicative, the hypothesis implies that $\mathfrak{A}$ is principal. So $Cl_{\mathbb{Z}_K} = 1$. □

Example 91.9. Let $K = \mathbb{Q}(\sqrt{m})$ with m a square-free integer. Then $h_K = 1$ if $m = -1, -2, -3, -7, -11, -19, -43, -67, -163$.

Baker and Stark showed that the fields in the example are the only imaginary quadratic fields K with $h_K = 1$.

Corollary 91.10. *Let K be a number field of degree greater than one. Then $|d_K| > 1$. In particular, $\mathbb{Z}_K$ is not unramified over $\mathbb{Z}$.*

Proof. We know that $\left(\frac{\pi}{4}\right)^{\frac{n}{2}} \frac{n^n}{n!} > 1$ if $n > 1$ (as the left hand side is increasing). Therefore, there exists a nonzero integral ideal $\mathfrak{A} \subset \mathbb{Z}_K$ satisfying

$$\sqrt{|d_K|} \geq \frac{\mathfrak{N}(\mathfrak{A})}{B_K} \geq \left(\frac{\pi}{4}\right)^{r_2} \frac{n^n}{n!} \geq \left(\frac{\pi}{4}\right)^{n} \frac{n^n}{n!} > 1. \tag{91.11}$$

Therefore, K ramifies over $\mathbb{Z}$ by the Dedekind Ramification Theorem 85.2. □

This result is special to $\mathbb{Q}$. If we look at relative number theory, i.e., $L/K/\mathbb{Q}$ is finite with $K \neq \mathbb{Q}$, it can happen that $\mathbb{Z}_L$ is unramified over $\mathbb{Z}_K$ with $K < L$. For any such field K, there exists a unique number field L with L/K the maximal abelian extension of K that satisfies $\mathbb{Z}_L$ is unramified over $\mathbb{Z}_K$. It is called the *Hilbert class field* of K. It has the properties that its degree over K is h_K and $G(L/K) \cong I_{Z_K}$ (canonically). So Corollary 91.10 says that $\mathbb{Q}$ is Hilbert class field of $\mathbb{Q}$.

For the next consequence, we need Stirling's Formula (that we do not prove).

Facts 91.12. (Stirling's Formula) *Let $N >> 0$ be an integer. Then*

$$\sqrt{2\pi N}\, N^N\, e^{-N} < N! < \sqrt{2\pi N}\, N^N\, e^{-N}\left(1 + \frac{1}{12N-1}\right).$$

Corollary 91.13. *Let K be a number field of degree $n_K > 1$. Then $\dfrac{n_K}{\log(|d_K|)}$ is bounded.*

Proof. By equation (91.11) (and squaring), we have, with $n = n_K$,

$$\frac{n}{\log(|d_K|)} \leq \frac{n}{2\log\left(\frac{n^n}{n!}\left(\frac{\pi}{4}\right)^{\frac{n}{2}}\right)}.$$

Therefore, by Stirling Formula, we have

$$\begin{aligned}\frac{n}{\log(|d_K|)} &\leq \frac{n}{2\log\left(\frac{e^n}{\sqrt{2\pi n}}\left(\frac{\pi}{4}\right)^{\frac{n}{2}}\frac{1}{1+\frac{1}{12n-1}}\right)}\\ &\leq \frac{n}{2n - \log(2\pi n) + n\log(\frac{\pi}{4}) + \log\left(1 + \frac{1}{12n-1}\right)}\end{aligned}$$

which is bounded (say by L'Hôpital's Rule). □

Corollary 91.14. *There exists finitely many number fields K with discriminant d.*

Proof. Fix a field K and set $n = n_K$. We use the embedding $i_{\mathbb{C}} : K \to \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$.

Case 1. $r_2 > 0$:

Define $X \subset \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ to be the set of those $x = (x_1, \ldots, x_{r_1+r_2})$ in $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ that satisfy:

(i) $|x_{r_1+1} - \overline{x}_{r_1+1}| = 2|Im(x_{r_1+1})| \leq C\sqrt{|d_K|}$

(ii) $|x_{r_1+1} + \overline{x}_{r_1+1}| = 2|Re(x_{r_1+1})| \leq \frac{1}{2}$

(iii) $|x_i| < \frac{1}{2}$ if $i \neq r_1 + 1$, where $C = C(n)$ is chosen to satisfy $\mu(X) > 2^n\mu(D_{\mathbb{Z}_K}) = 2^{n-r_2}\sqrt{|d_K|}$. The set X is measurable, symmetric, and convex, so there exists a nonzero $\alpha \in \mathbb{Z}_K \cap X$ by the Minkowski Lattice Theorem. Since $|\operatorname{N}_{K/\mathbb{Q}}(\alpha)|$ is a positive integer, $(i), (ii), (iii)$ imply that $|Im(\alpha^{(r_1+1)}| > 1$. Thus $\alpha^{(r_1+1)} \neq \alpha^{(r_1+r_2+1)} = \overline{\alpha}^{(r_1+1)}$ and $|\alpha^{(r_1+1)}| > 1$. It follows by $(i), (ii), (iii)$ that $\alpha^{(r_1+1)} \neq \alpha^{(i)}$ for any $i \neq r_1 + 1$. Let $L/\mathbb{Q}$ be the normal closure of $K/\mathbb{Q}$ and extend each $\sigma_i : K \to \mathbb{C}$, $i = 1, \dots, n$ to L. Then $\sigma_{r_1+1}(\alpha) \neq \sigma_i(\alpha)$ for any $i \neq r_1 + 1$. Therefore, $\sigma_{r_1+1}(\alpha) = \sigma_{r_1+1}\sigma_j^{-1}\sigma_i(\alpha)$ if and only if $i = j$. It follows that α has n distinct conjugates. Since the degree of the minimal polynomial $m_{\mathbb{Q}}(\alpha)$ satisfies $\deg m_{\mathbb{Q}}(\alpha) = n = [K : \mathbb{Q}]$, we must have $K = \mathbb{Q}(\alpha)$.

We have, therefore, shown the following:

Conclusion. K is determined by $m_{\mathbb{Q}}(\alpha)$ in $\mathbb{Z}[t]$ (so monic) whose coefficients are elementary symmetric functions in the $\alpha^{(i)}$.

As $\alpha \in X$, the coefficients in the conclusion are integers and bounded by a function only depending on n and $|d_K|$.

For any fixed n_K, d_K, there exist only finitely many such polynomials, hence only finitely many K with fixed n_K, d_K for K having a complex embedding. By Corollary 91.13, if $d = d_K$ is fixed, then $n = n_K$ is bounded. So we are done in this case.

Case 2. $r_2 = 0$:

Define $X \subset \mathbb{R}^n$ by those $x \in \mathbb{R}^n$ satisfying

(i) $|x_1| \leq C\sqrt{|d_K|}$.

(ii) $|x_i| \leq \frac{1}{2}$ for $j > i$

where $C = C(n)$ is defined as before. Now argue as in Case 1. (This is easier.) □

We end by giving a few computations of class numbers using Corollary 91.5.

Examples 91.15. 1. Write $\mathfrak{A} \sim \mathfrak{B}$ if $\mathfrak{A}$ and $\mathfrak{B}$ in $I_{\mathbb{Z}_K}$ are in the same ideal class in $Cl_{\mathbb{Z}_K}$. Computation involves not only the Minkowski bound, but also knowledge of the way primes split. As as example, using the computations in Section 86, we illustrate this for $K = \mathbb{Q}(\sqrt{-47})$. We have $\{1, \frac{1+\sqrt{-47}}{2}\}$ is an integral basis for $\mathbb{Z}_K$, $d_K = -47$, $B_K = 2\sqrt{47}/\pi = 14/2 < 5$ and $\operatorname{N}_{K/\mathbb{Q}}(x + y\frac{1+\sqrt{-47}}{2}) = x^2 + xy + 12y^2$. We know that it suffices to look at prime ideals in $\mathbb{Z}_K$ lying over $2, 3$ respectively. As 2 and 3 split in $\mathbb{Z}_K$, we have prime factorizations into primes, $2\mathbb{Z}_K = \mathfrak{p}_2\overline{\mathfrak{p}}_2$ and $3\mathbb{Z}_K = \mathfrak{p}_3\overline{\mathfrak{p}}_3$ using the computations in Section 86. As 2 is not a norm from $\mathbb{Z}_K$, we know that $\mathfrak{p}_2 \not\sim 1$. We also know that $2 \not| \frac{1+\sqrt{-47}}{2}$ in $\mathbb{Z}_K$. Changing notation if necessary, this means that $v_{\mathfrak{p}_2}(\frac{1+\sqrt{-47}}{2}) = 2$ and $v_{\overline{\mathfrak{p}}_2}(\frac{1+\sqrt{-47}}{2}) = 0$, $v_{\mathfrak{p}_3}(\frac{1+\sqrt{-47}}{2}) = 1$ and $v_{\overline{\mathfrak{p}}_3}(\frac{1+\sqrt{-47}}{2}) = 0$, hence $(\frac{1+\sqrt{-47}}{2})\mathbb{Z}_K \sim \mathfrak{p}_3\mathfrak{p}_2^2$. It follows that $\mathfrak{p}_3 \sim \mathfrak{p}_2^2$ and the class of $\mathfrak{p}_2$ generates $Cl_{\mathbb{Z}_K}$. Since $\mathfrak{p}_2$ is not principal, $h_K > 1$. Let $x = 4 + \frac{1+\sqrt{-47}}{2}$, then $2 \not| x$ in $\mathbb{Z}_K$ and $\operatorname{N}_{K/\mathbb{Q}}(x) = 32$. Arguing as above, we see that $\mathfrak{p}_2^5$ is principal and $Cl_{\mathbb{Z}_K}$ is cyclic of order 5.

2. We give a real quadratic example: $K = \mathbb{Q}(\sqrt{31})$, and again use Section 86. [A fundamental unit is $\varepsilon = 1520+273\sqrt{31}$.] We have $\mathbb{Z}_K$ has integral basis $\{1, \sqrt{31}\}$, $d_K = 4\cdot 31$, $B_K = \sqrt{31} < 6$, and $\operatorname{N}_{K/\mathbb{Q}}(x+y\sqrt{31}) = x^2 - 31y^2$. So we need only look at primes lying above $2, 3, 5$. We know that 3 and 5 split in $\mathbb{Z}_K$, say $3\mathbb{Z}_K = \mathfrak{p}_3\mathfrak{p}_3'$ and $5\mathbb{Z}_K = \mathfrak{p}_5\mathfrak{p}_5'$. We

also know that $2\mathbb{Z}_K = \mathfrak{p}_2^2$ ramifies. Since $\mathrm{N}_{K/\mathbb{Q}}(39+7\sqrt{31}) = 2$, we see that $\mathfrak{p}_2 \sim 1$. We also have $5 = 6^2 - 31 \cdot 1^2$ is a norm so $\mathfrak{p}_5 \sim 1$. Moreover, $-2\cdot 3\cdot 5 = -30 = 1^2 - 31\cdot 1^2$ is a norm. It follows that $\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5 \sim 1$. Therefore, $\mathfrak{p}_3 \sim 1$ also. Hence $h_K = 1$.

3. We give the following example that Lang says was Artin's favorite. Let $g = t^5 - t + 1 \in \mathbb{Z}[t]$, α a root of g in $\mathbb{C}$ and $K = \mathbb{Q}(\alpha)$. Let $L/\mathbb{Q}$ be the normal closure of $K/\mathbb{Q}$. Then g is irreducible (as it is irreducible in $\mathbb{Z}/5\mathbb{Z}$), and one checks that $G(L/\mathbb{Q}) \cong S_5$, $r_1 = 3$, $r_2 = 1$, and $\Delta(\alpha) = 2869 = 19 \cdot 151$ [as $\Delta(t^5 + at + b) = 5^5 b^4 + 2^8 a^5$]. So $d_K = \Delta(a)$ (and has only two primes and is square-free) and $\mathbb{Z}_K = \mathbb{Z}[\alpha]$. We have
$$B_K\sqrt{|d_K|} = \Big(\frac{4}{\pi}\Big)\Big(\frac{5!}{5^5}\Big)\sqrt{2869} < 4.$$
so every ideal class contains an integral ideal $\mathfrak{A}$ with $\mathfrak{N}(\mathfrak{A}) = 1, 2$, or 3. Suppose that $\mathfrak{N}(\mathfrak{A}) = 2, 3$, respectively. Then $\mathfrak{A}$ is a prime ideal $\mathfrak{p}$ with $\mathfrak{N}(\mathfrak{A}) = p := 2, 3$, respectively. It follows that $f(\mathfrak{p}/p\mathbb{Z}) = 1$ and $g \mod p$ has a root in $\mathbb{Z}/p\mathbb{Z}$, which it does not. Therefore, $\mathfrak{N}(\mathfrak{A}) = 1$ and $\mathbb{Z}_K$ is a PID.

Exercises 91.16. 1. Prove Example 91.9.

2. Let m is a square-free negative number and suppose that $h_{\mathbb{Z}_{\mathbb{Q}(\sqrt{m})}} = 1$. Show all of the following:
   (*i*) We must have $m \equiv 5 \mod 8$ except when $m = -1$, -2, or -7.
   (*ii*) If p is an odd prime satisfying $m < -4p$, then m is a nonsquare modulo p.
   (*iii*) If $m < -19$, then m is congruent to one of the following modulo $840 : -43, -67, -163, -403, -547,$
   (*iv*) If $-2000 < m < 0$, then $\mathbb{Z}_{\mathbb{Q}(\sqrt{m})}$ is one of the quadratic number fields given by Example 91.9.
3. Show that $h_{\mathbb{Z}_{\mathbb{Q}(\sqrt{-5})}} = 2$.
4. Show that $h_{\mathbb{Z}_{\mathbb{Q}(\sqrt{-23})}} = 3$.
5. Show that $Cl_{\mathbb{Z}_{\mathbb{Q}(\sqrt{-21})}}$ is the Klein 4-group.
6. Show that $h_{\mathbb{Z}_{\mathbb{Q}(\sqrt{2},\sqrt{-3})}} = 1$. [An integral basis for $\mathbb{Z}_{\mathbb{Q}(\sqrt{2},\sqrt{-3})}$ is $\{1, \sqrt{2}, (1+\sqrt{-3})/2, (1+\sqrt{2}+\sqrt{-6})/2\}$.]
7. Fill in the details to Example 91.15(3).
8. Let K be a number field and $\mathfrak{A} \subset \mathbb{Z}_K$ be an ideal such that $(\mathfrak{A})^m = a\mathbb{Z}_K$ for some $a \in \mathbb{Z}_K$. Show that $\mathfrak{A}\mathbb{Z}_{K(\sqrt[m]{a})}$ is principal.
9. Let K be a number field. Show there exists a finite extension L/K such that every nonzero ideal in $\mathbb{Z}_K$ becomes principal in $\mathbb{Z}_L$.

CHAPTER XVII

Introduction to Commutative Algebra

In this chapter, all rings will be commutative. We shall be particularly interested in developing an algebraic dimension theory that coincides with the intuitive notion of the number of variables needed to describe a geometric object. In addition, we shall establish a generalization of unique factorization into prime ideals as happens in Dedekind domains to arbitrary Noetherian rings.

92. Zariski Topology

Throughout this section, R will denote commutative ring.

For a nonzero commutative ring R, we introduced in §41, the Spectrum of R,

$$\mathrm{Spec}(R) := \{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal}\},$$

and defined the Zariski topology on it as follows:
If T is a subset of R, define

$$V_R(T) := \{\mathfrak{p} \mid \mathfrak{p} \in \mathrm{Spec}(R) \text{ with } T \subset \mathfrak{p}\}$$

called the *(abstract) variety* of T. We have

Lemma 92.1. *Let R be a commutative ring. Then*

(1) *If T is a subset of R, then $V_R(T) = V_R(\langle T\rangle)$.*
(2) *If $T_1 \subset T_2$ are subsets of R, then $V_R(T_1) \supset V_R(T_2)$.*
(3) *$V_R(\emptyset) = \mathrm{Spec}(R)$.*
(4) *$V_R(R) = \emptyset$.*
(5) *If $T_i, i \in I$, are subsets of R, then $V_R(\bigcup_I T_i) = \bigcap_I V_R(T_i)$.*
(6) *If $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in R, then*

$$V_R(\mathfrak{A}\mathfrak{B}) = V_R(\mathfrak{A} \cap \mathfrak{B}) = V_R(\mathfrak{A}) \cup V_R(\mathfrak{B}).$$

In particular, the collection

$$\mathcal{C} := \{V_R(T) \mid T \subset R\}$$

forms a system of closed sets for the Zariski topology on $\mathrm{Spec}(R)$.

Proof. This essentially was Exercise 41.15(6).
As an example, we show (6).
Since $\mathfrak{A}\mathfrak{B} \subset \mathfrak{A} \cap \mathfrak{B}$, we have $V(\mathfrak{A}\mathfrak{B}) \supset V(\mathfrak{A} \cap \mathfrak{B})$ by (2).
If $\mathfrak{p} \in V(\mathfrak{A}\mathfrak{B})$, then $\mathfrak{A}\mathfrak{B} \subset \mathfrak{p}$, hence $\mathfrak{A} \subset \mathfrak{p}$ or $\mathfrak{B} \subset \mathfrak{p}$, equivalently, $\mathfrak{p} \in V(\mathfrak{A})$ or $\mathfrak{p} \in V(\mathfrak{B})$, i.e., $V(\mathfrak{A}\mathfrak{B}) \subset V_R(\mathfrak{A}) \cup V_R(\mathfrak{B})$.
Finally, if $\mathfrak{p} \in V(\mathfrak{A}) \cup V(\mathfrak{B})$, then $\mathfrak{A} \cap \mathfrak{B} \subset \mathfrak{p}$, so $V(\mathfrak{A}) \cup V(\mathfrak{B}) \subset V(\mathfrak{A} \cap \mathfrak{B})$. □

We shall always assume that $\mathrm{Spec}(R)$ is given the Zariski topology. If R is clear, we shall abbreviate $V_R(T)$ by $V(T)$ and if $\mathfrak{A} = (a_1, \ldots, a_n)$ is a finitely generated ideal in R, we shall write $V(a_1, \ldots, a_n)$ for $V(\mathfrak{A})$.

The Zariski topology is very coarse. In general, as we shall see, it is not a Hausdorff space, i.e., we cannot necessarily find disjoint open neighborhoods of distinct points. Moreover, we shall see that the set of *closed points* in $\mathrm{Spec}(R)$, i.e., those points $\mathfrak{p}$ in $\mathrm{Spec}(R)$ such that $\{\mathfrak{p}\}$ is a closed set, is precisely the set of maximal ideals, so points are usually not closed.

Lemma 92.2. *Let X be a subset of $\mathrm{Spec}(R)$ and $\overline{X}$ its closure in $\mathrm{Spec}(R)$. If $\mathfrak{A} = \bigcap_{\mathfrak{p}\in X} \mathfrak{p}$, then $\overline{X} = V(\mathfrak{A})$.*

PROOF. ($\subset$:) As $\mathfrak{A} = \bigcap_X \mathfrak{p} \subset \mathfrak{P}$ for all $\mathfrak{P} \in X$, we have $X \subset V(\mathfrak{A})$, hence $\overline{X} \subset V(\mathfrak{A})$. ($\supset$:) Let $\overline{X} = V(\mathfrak{B})$, with $\mathfrak{B}$ an ideal in R. If $\mathfrak{p} \in X \subset \overline{X} = V(\mathfrak{B})$, then $\mathfrak{B} \subset \mathfrak{p}$, hence $\mathfrak{B} \subset \bigcap_X \mathfrak{p} = \mathfrak{A}$. Therefore, $\overline{X} = V(\mathfrak{B}) \supset V(\mathfrak{A})$ by Lemma 92.1(2). □

Corollary 92.3. *Let R be a commutative ring. Then*

(1) $\mathrm{Max}(R) \subset \mathrm{Spec}(R)$ is the set of closed points in $\mathrm{Spec}(R)$.
(2) Let $\mathfrak{p}$ and $\mathfrak{P}$ be prime ideals. Then $\mathfrak{P} \in \overline{\{\mathfrak{p}\}}$, the closure of $\{\mathfrak{p}\}$ in $\mathrm{Spec}(R)$, if and only if $\mathfrak{p} \subset \mathfrak{P}$.
(3) $\overline{\{\mathfrak{p}\}} = V(\mathfrak{p})$ for all prime ideals in R.
(4) $\mathrm{Spec}(R)$ is a T_0-topological space, i.e., if $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are two distinct prime ideals in R, then there exists an open set in $\mathrm{Spec}(R)$ containing one of $\mathfrak{p}_1$, $\mathfrak{p}_2$, but not the other.

PROOF. (3) follows from the lemma and (2) follows from (3).
(1): Let $\mathfrak{p} \in \mathrm{Spec}(R)$. Then we have $\mathfrak{p}$ is a closed point if and only if $\overline{\mathfrak{p}} = \{\mathfrak{q} \in \mathrm{Spec}(R) \mid \mathfrak{p} \subset \mathfrak{q}\} = \{\mathfrak{p}\}$ if and only if $\mathfrak{p} \in \mathrm{Max}(R)$.
(4): By (2), if $\mathfrak{p}_1 \not\subset \mathfrak{p}_2$ are prime ideals, then $\mathfrak{p}_2$ lies in the open set $\mathrm{Spec}(R) \setminus V(\mathfrak{p}_1)$. □

Definition 92.4. An ideal $\mathfrak{A} \subset R$ is called a *radical ideal* if $\mathfrak{A} = R$ or $\mathfrak{A} < R$ and
$$\mathfrak{A} = \sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ some } n \in \mathbb{Z}^+\} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{\mathfrak{A}\subset\mathfrak{p}\in\mathrm{Spec}(R)} \mathfrak{p}.$$
Equivalently, if a in R satisfies $a^n \in \mathfrak{A}$, then $a \in \mathfrak{A}$.

Examples 92.5. The following are radical ideals:

(1) Prime ideals.
(2) The nilradical, $\mathrm{nil}(R)$, of R as $\mathrm{nil}(R) = \bigcap_{V(0)} \mathfrak{p} = \sqrt{(0)}$.

Corollary 92.6. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Then $V(\mathfrak{A}) = V(\sqrt{\mathfrak{A}})$ and*
$$V(\mathfrak{A}) \subset V(\mathfrak{B}) \text{ if and only if } \sqrt{\mathfrak{A}} \supset \sqrt{\mathfrak{B}}.$$
Let
$$\mathcal{R} := \{\mathfrak{A} \subset R \mid \mathfrak{A} \text{ is a radical ideal}\}$$
$$\mathcal{V} := \{V(\mathfrak{A}) \mid \mathfrak{A} \text{ is an ideal of } R\}.$$

Then $V : \mathcal{R} \to \mathcal{V}$ given by $\mathfrak{A} \mapsto V(\mathfrak{A})$ is an order-reversing bijection with inverse $\mathcal{I} : \mathcal{V} \to \mathcal{R}$ given by $V \mapsto \bigcap_V \mathfrak{p}$.

Proof. Clearly, $V(\mathfrak{A}) = V(\sqrt{\mathfrak{A}})$ and $V(\mathfrak{A}) \subset V(\mathfrak{B})$ implies $\sqrt{\mathfrak{B}} \subset \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \sqrt{\mathfrak{A}}$. □

We leave the following as an exercise:

Proposition 92.7. *Let r be an element of R and*

$$D(r) := \operatorname{Spec}(R) \setminus V(r) = \{\mathfrak{p} \in \operatorname{Spec}(R) \mid r \notin \mathfrak{p}\}.$$

Then

1. *A finite intersection of basic open sets is a basic open set.*
2. *$\{D(r) \mid r \in R\}$ is a base for the topology of $\operatorname{Spec}(R)$, i.e., every open set in $\operatorname{Spec}(R)$ is a union of such $D(r)$.*
3. *$\operatorname{Spec}(R)$ is quasi-compact, i.e., every open cover of $\operatorname{Spec}(R)$ has a finite subcover.*
4. *$\operatorname{Spec}(R)$ is a Hausdorff space if and only if $\operatorname{Spec}(R) = \operatorname{Max}(R)$.*

The set $D(r)$ above is called a *basic open set* in $\operatorname{Spec}(R)$.

Definition 92.8. A nonempty topological space is called *irreducible* if whenever U, V are nonempty open subsets of X, then $U \cap V \neq \emptyset$. A subspace Y of a topological space X is called *irreducible* if it is irreducible in the induced topology.

Examples 92.9. 1. Points are irreducible.
2. A nonempty Hausdorff space is irreducible if and only if it consists of a single point.

Proposition 92.10. *Let X be a nonempty topological space. Then the following are equivalent:*

1. *X is irreducible.*
2. *If $W_1, W_2 < X$ are closed, then $W_1 \cup W_2 < X$.*
3. *If $\{W_1, \dots, W_n\}$ is a finite closed cover of X, then $X = W_i$ for some i.*
4. *If U is a nonempty open subset of X, then U is dense in X.*
5. *Every open set in X is connected.*

We leave the proof as an exercise. For varieties, the concept of irreducibility replaces connectivity.

Notation 92.11. We know if R is a nonzero commutative ring, then there exist minimal elements in $\operatorname{Spec}(R)$ by Zorn's Lemma that we called *minimal prime ideals* of R (cf. Remark 28.18). We let

$$\operatorname{Min}(R) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid \text{there exists no prime ideal } \mathfrak{p}_0 < \mathfrak{p}\}$$

denote the nonempty set of *minimal primes ideals* in R.

Lemma 92.12. *$\operatorname{Spec}(R)$ is irreducible if and only if $\operatorname{nil}(R)$ is a prime ideal if and only if $|\operatorname{Min}(R)| = 1$. In particular, if R is a domain, then $\operatorname{Spec}(R)$ is irreducible.*

Proof. We know that $\mathrm{nil}(R) = \bigcap_{\mathrm{Spec}(R)} \mathfrak{p} = \bigcap_{\mathrm{Min}(R)} \mathfrak{p} \subset \mathfrak{P}$ for every prime ideal $\mathfrak{P}$ and $V(0) = V(\sqrt{0}) = V\big(\mathrm{nil}(R)\big) = \mathrm{Spec}(R)$.

Check 92.13. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Then

(i) $\mathfrak{A} \subset \mathrm{nil}(R)$ if and only if $\sqrt{\mathfrak{A}} \subset \mathrm{nil}(R)$.

(ii) $\sqrt{\mathfrak{A}\mathfrak{B}} \supset \sqrt{\mathfrak{A}}\sqrt{\mathfrak{B}}$.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Then, using the check, we have $|\,\mathrm{Min}(R)| = 1$ if and only if $\mathrm{nil}(R)$ is a prime ideal if and only if $\mathfrak{A}\mathfrak{B} \subset \mathrm{nil}(R)$ implies $\mathfrak{A} \subset \mathrm{nil}(R)$ or $\mathfrak{B} \subset \mathrm{nil}(R)$ if and only if $\sqrt{\mathfrak{A}}\sqrt{\mathfrak{B}} \subset \mathrm{nil}(R)$ implies $\sqrt{\mathfrak{A}} \subset \mathrm{nil}(R)$ or $\sqrt{\mathfrak{B}} \subset \mathrm{nil}(R)$ if and only if $V(\mathfrak{A}\mathfrak{B}) = V(\mathfrak{A}) \cup V(\mathfrak{B}) = \mathrm{Spec}(R)$ implies $V(\mathfrak{A}) = \mathrm{Spec}(R)$ or $V(\mathfrak{B}) = \mathrm{Spec}(R)$ if and only if $\mathrm{Spec}(R)$ is irreducible. $\square$

Since $\mathrm{Spec}(R)$ is a topological space based on an algebraic structure, via the ring R, topological maps of spectra should arise from homomorphisms of the underlying algebraic structures. Indeed, we have:

Definition 92.14. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. Define the *associated map* to be the map arising from φ by

$$^{a}\varphi : \mathrm{Spec}(B) \to \mathrm{Spec}(A) \ \text{ via } \ ^{a}\varphi(\mathfrak{P}) = \varphi^{-1}(\mathfrak{P}).$$

That this map makes sense follows by:

Lemma 92.15. *Let $\varphi : A \to B$ be a ring homomorphism. Then $^{a}\varphi(\mathfrak{B})$ is a prime ideal of A for every prime ideal $\mathfrak{P}$ in B, i.e.,*

$$^{a}\varphi : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$$

is defined. Moreover, if $T \subset A$ is a subset, then

$$(^{a}\varphi)^{-1}\big(V(T)\big) = V\big(\varphi(T)\big).$$

In particular, $^{a}\varphi : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$ is a continuous map.

Proof. If $\mathfrak{P} < B$ is a prime ideal, then the ideal $\varphi^{-1}(\mathfrak{P}) < A$ is an ideal and φ induces a monomorphism $A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$ with $B/\mathfrak{B}$ a domain. Consequently, $A/\varphi^{-1}(\mathfrak{P})$ is also a domain, i.e., $\varphi^{-1}(\mathfrak{P})$ is a prime ideal.

We also have

$$\begin{aligned}(^{a}\varphi)^{-1}\big(V(T)\big) &= \{\mathfrak{P} \in \mathrm{Spec}(B \mid \varphi^{-1}(\mathfrak{P}) = {}^{a}\varphi(\mathfrak{P}) \supset T\} \\ &= \{\mathfrak{P} \in \mathrm{Spec}(B) \mid \mathfrak{P} \supset \varphi(T)\} = V\big(\varphi(T)\big).\end{aligned}$$ $\square$

It is easy to check:

Remark 92.16. If $\varphi : A \to B$ and $\psi : B \to C$ are ring homomorphisms of commutative rings, then $^{a}(\psi \circ \varphi) = {}^{a}\varphi \circ {}^{a}\psi$.

Definition 92.17. A map $f : X \to Y$ or topological spaces is called *dominant* if $\mathrm{im}\, f$ is dense in Y.

The relationship between φ and $^{a}\varphi$ is illustrated by:

Lemma 92.18. *Let $\varphi : A \to B$ be a ring homomorphism.*

(1) *If φ is an epimorphism, then ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is an injective closed map with image $V(\ker\varphi)$. In particular, ${}^a\varphi$ induces a homeomorphism $V(\ker\varphi) \cong \operatorname{Spec}(B) \cong \operatorname{Spec}(A/\ker\varphi)$.*
(2) *If $\mathfrak{A}$ is an ideal, then $V(\mathfrak{A})$ is homeomorphic to $\operatorname{Spec}(R/\mathfrak{A})$.*
(3) *If φ is a monomorphism, then ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a dominant map. [It may not be onto.]*

PROOF. (1) follows from the Correspondence Principle and (1) implies (2).
(3): Suppose that $a \in A$ satisfies $D(a) := \operatorname{Spec}(A) \setminus V(a)$ is nonempty. We must show that $D(a) \cap \operatorname{im}{}^a\varphi$ is also nonempty. Suppose that this is false, then $\operatorname{im}{}^a\varphi \subset V(a)$. So

$$a \in \bigcap_{\operatorname{Spec}(B)} {}^a\varphi(\mathfrak{P}) = \bigcap_{\operatorname{Spec}(B)} \varphi^{-1}(\mathfrak{P}),$$

hence $\varphi(a) \in \bigcap_{\operatorname{Spec}(B)} \mathfrak{P} = \operatorname{nil}(B)$, i.e., $\varphi(a)$ is nilpotent in B. In particular, there exists a positive integer n with $0 = \varphi(a)^n = \varphi(a^n)$. As φ is monic, $a^n = 0$ in A, so $a \in \operatorname{nil}(A) \subset \mathfrak{p}$ for every prime ideal $\mathfrak{p}$. It follows that $D(a) = \emptyset$, a contradiction. $\square$

Note: The inclusion $i : \mathbb{Z} \to \mathbb{Z}_{(p)}$, with $\mathbb{Z}_{(p)} = \{\frac{a}{b} \mid a, b \in \mathbb{Z},\ p \nmid b\}$, the localization of $\mathbb{Z}$ at the prime ideal (p), is injective, but ${}^a i$ is not surjective as $\operatorname{Spec}(\mathbb{Z}_{(p)}) = \{0, p\mathbb{Z}_{(p)}\}$ (but does have dense image as shown).

Localization will play a decisive role in this chapter as well as being a primary tool in commutative algebra. It is, therefore, useful to coalesce some of the properties of localization, many of which appeared in §29 (some as exercises) and subsequent sections, together with some new observations, leaving all the proofs as useful exercises.

Remark 92.19. Let R be a nonzero commutative ring and S a multiplicative set in R. We have seen that we have a canonical ring homomorphism $\varphi_R : R \to S^{-1}R$ given by $r \mapsto r/1$. This makes $S^{-1}R$ into a commutative R-algebra. If $\mathfrak{A}$ in R is an ideal, then

$$S^{-1}\mathfrak{A} := \{\frac{a}{s} \mid a \in A,\ s \in S\} \subset S^{-1}R$$

is an ideal. Let $\mathfrak{B}$ be another ideal in R. The following are true:

(*i*) $S^{-1}\mathfrak{A} = S^{-1}R$ if and only if $\mathfrak{A} \cap S \neq \emptyset$.
(*ii*) $\mathfrak{A} \subset \varphi_R^{-1}(S^{-1}\mathfrak{A}) \subset R$ is an ideal.
(*iii*) If $\mathfrak{D}$ is an ideal in $S^{-1}R$, then there exists an ideal $\mathfrak{C}$ in R satisfying $S^{-1}\mathfrak{C} = \mathfrak{D}$. Moreover, $\mathfrak{D} = S^{-1}\big((\varphi_R)^{-1}(\mathfrak{D})\big)$.
(*iv*) If $\mathfrak{A} < \mathfrak{B}$ and $\mathfrak{A} \cap S = \emptyset$, then $S^{-1}\mathfrak{A} < S^{-1}\mathfrak{B}$.
(*v*) $S^{-1}(\mathfrak{A} + \mathfrak{B}) = S^{-1}\mathfrak{A} + S^{-1}\mathfrak{B}$.
(*vi*) $S^{-1}(\mathfrak{A} \cap \mathfrak{B}) = S^{-1}\mathfrak{A} \cap S^{-1}\mathfrak{B}$.
(*vii*) $S^{-1}(\mathfrak{A}\mathfrak{B}) = S^{-1}\mathfrak{A}\, S^{-1}\mathfrak{B}$.
(*viii*) The map

$$\{\mathfrak{p} \in \operatorname{Spec} R \mid \mathfrak{p} \cap S = \emptyset\} \to \operatorname{Spec}(S^{-1}R) \text{ given by } \mathfrak{p} \mapsto S^{-1}\mathfrak{p}$$

is a bijection and, in fact, a homeomorphism.

As mentioned before, the most important examples of localization of a commutative ring R are:

(i) $R_{\mathfrak{p}}$ the localization of R at the multiplicative set $R \setminus \mathfrak{p}$ where $\mathfrak{p}$ is a prime ideal of R, called the localization at $\mathfrak{p}$.

(ii) R_a, the localization of R at the multiplicative set $S = \{a^n \mid n \geq 0\}$ called the localization at a, i.e., at "the open neighborhood $D(a)$ of a". (Note that $D(a)$ is homeomorphic to $\mathrm{Spec}(S^{-1}R)$.)

We return to the topology that interests us. We have seen that irreducible topological spaces are applicable to our theory. However, connected subspaces are not useful, so we need a substitute for the usual decomposition of a space into connected components into a decomposition of another type. This will be a decomposition into irreducible components.

Lemma 92.20. *Let X be a topological space and Y a subset of X.*

(1) Y is irreducible if and only if $\overline{Y}$ in X is irreducible.
(2) Every irreducible subspace of X is contained in some maximal irreducible subspace. Any such is closed.

Proof. (1): Let $U \subset X$ be an open subset, then $Y \subset \overline{U}$ if and only if $\overline{Y} \subset \overline{U}$. By definition, the closure of $U \cap \overline{Y}$ in $\overline{Y}$ is $\overline{U} \cap \overline{Y}$, so $U \cap Y$ is dense in Y if and only if $U \cap \overline{Y}$ is dense in $\overline{Y}$.

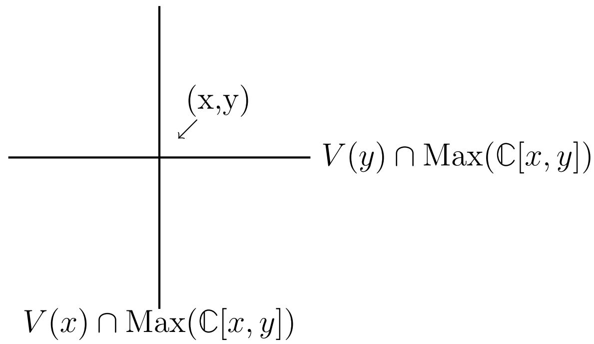
(2) follows from Zorn's Lemma. □

A maximal irreducible subspace of a nonempty topological space X is called an *irreducible component* of X. If X is a Hausdorff space, then $\{x\}$, $x \in X$, are the irreducible components of X, so this is not a useful concept in that case.

Definition 92.21. Let X be a topological space. Then a point $x \in X$ is called a *generic point* of X if $X = \overline{\{x\}}$.

Corollary 92.22. $\{V(\mathfrak{p}) \mid \mathfrak{p} \in \mathrm{Min}(R)\}$ *is the set of irreducible components of* $\mathrm{Spec}(R)$ *and* $\mathfrak{p} \in \mathrm{Spec}(R)$ *is a generic point of* $V(\mathfrak{p})$.

Remark 92.23. Irreducible components may intersect nontrivially. For example, let x, y be indeterminants, then $V(xy)$ in $\mathrm{Spec}(\mathbb{C}[x,y])$ has irreducible components $V(x)$ and $V(y)$ and they intersect in (x,y), a maximal ideal in $\mathbb{C}[x,y]$. Topologically, $V(xy) \cong \mathrm{Spec}\left(\mathbb{C}[x,y]/(xy)\right)$. Identifying $\mathrm{Max}(\mathbb{C}[x,y])$ with $\mathbb{C}^2$ by $(x-a, y-b) \mapsto (x,y)$ using the Hilbert Nullstellensatz, we can view the y-axis as the closed points in $V(x)$ and the x-axis as the closed points in $V(y)$. The origin corresponds to the intersection of $V(x)$ and $V(y)$. The following picture illustrates, this where we have identified $V(x) \cap \mathrm{Max}(\mathbb{C}[x,y])$ and the y-axis, $V(y) \cap \mathrm{Max}(\mathbb{C}[x,y])$ and the x-axis) and (x,y) with the origin, using the Hilbert Nullstellensatz.



Definition 92.24. A topological space X is called *Noetherian* if the collection of open sets in X satisfies the ascending chain condition. Equivalently, the collection of closed sets in X satisfies the descending chain condition.

Example 92.25. If R is a commutative Noetherian ring, then $\mathrm{Spec}(R)$ is a Noetherian topological space, but the converse is false. For example, let F be a field, $A = F[t_1, \dots, t_n, \dots]$, and $\mathfrak{A} = (t_1, t_2^2, \dots, t_n^n, \dots)$. Then the ring $B = A/\mathfrak{A}$ is not Noetherian, as $\mathrm{nil}(B)$ is not nilpotent, but $\mathrm{Spec}(B)$ is a Noetherian space. In fact, $|\,\mathrm{Spec}(B)| = 1$.

Proposition 92.26. *Let X be a Noetherian space. Then there exist finitely many irreducible components of X and X is their union.*

PROOF. This proof is a typical proof when we have a Noetherian condition. Let

$$\begin{aligned} \mathcal{F}(X) &:= \{V \mid V \subset X \text{ closed}\} \\ \mathcal{C}(X) &:= \{C \mid C \subset \mathcal{F}(X), \\ &\qquad C \text{ is a finite union of closed irreducible sets in } X\}. \end{aligned}$$

Claim. $\mathcal{C}(X) = \mathcal{F}(X)$:

If this is false, then by the Minimal Principle (equivalent to the descending chain condition), there exists an element $Y \in \mathcal{F}(X) \setminus \mathcal{C}(X)$ that is minimal. As, $Y \notin \mathcal{C}(X)$, it is not irreducible, $Y = Y_1 \cup Y_2$ with $Y_i < Y$ and $Y_i \in \mathcal{F}(X)$, for $i = 1, 2$. By minimality, $Y_i \in \mathcal{C}(X)$, hence $Y = Y_1 \cup Y_2$ in $\mathcal{C}(X)$. This contradiction establishes the claim.

By the claim, we have $X = \bigcup_{i=1}^n C_i$ for some $C_i \subset X$ closed irreducible. Clearly, we may assume that C_j is not a subset of C_i for $i \neq j$. If C is an irreducible component of X, then we have $C \subset \bigcup_{i=1}^n (C \cap C_i)$ is irreducible. It follows that $C \subset C_i$ for some i, hence $C = C_i$ for some i as C is a maximal closed irreducible subset of X. The result follows. $\square$

Corollary 92.27. *Let R be a commutative Noetherian ring. Then* $\mathrm{Min}(R)$ *is a finite set and* $\mathrm{Spec}(R) = \bigcup_{\mathrm{Min}(R)} V(\mathfrak{p})$.

Our primary interest in this chapter will be the Krull dimension of a ring that we shall just call the *dimension of a ring*. For the convenience of the reader, we define it again, in an equivalent formulation, that will be of considerable interest for us.

Definition 92.28. Let R be a nonzero commutative ring and $\mathfrak{P}$ a prime ideal in R. We define the *height* of $\mathfrak{P}$ by

$$\mathrm{ht}\,\mathfrak{P} := \sup\{n \mid \mathfrak{p}_0 < \cdots < \mathfrak{p}_n = \mathfrak{P}, \text{ with } \mathfrak{p}_i \in \mathrm{Spec}(R),\ i = 1, \dots, n\}$$

if finite and $= \infty$ if not, i.e., the maximal number of (proper) links of chains of prime ideal in $\mathfrak{P}$. If $\mathfrak{A} < R$ is an ideal, we define the *height* of $\mathfrak{A}$ to be

$$\operatorname{ht}\mathfrak{A} := \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A})\} = \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}\}.$$

Then the *(Krull) dimension* of R is defined to be

$$\dim R := \sup\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Spec}(R)\} = \sup\{\operatorname{ht}\mathfrak{m} \mid \mathfrak{m} \in \operatorname{Max}(R)\}$$

if finite and $= \infty$ otherwise.

If $\mathfrak{A}$ is an ideal, we define the *dimension* of $V(\mathfrak{A})$ to be

$$\dim V(\mathfrak{A}) := \dim R/\mathfrak{A}.$$

More generally, if X is a topological space, the *combinatorial dimension* of X is defined to be

$$\dim X := \sup\{n \mid Y_0 < \cdots < Y_n \text{ each } Y_i \subset X \text{ closed and irreducible}\}$$

if finite and $= \infty$ if not. In particular, if $\mathfrak{A} \subset R$ is an ideal, then

$$\dim R = \dim \operatorname{Spec}(R) \text{ and } \dim V(\mathfrak{A}) = \dim \operatorname{Spec}(R/\mathfrak{A}).$$

Note that this is consistent, since $V(\mathfrak{A})$ is homeomorphic to $\operatorname{Spec}(R/\mathfrak{A})$.

Remarks 92.29. Let R be a nonzero commutative ring and F a field.

1. If R is a domain, then $\dim R = 0$ if and only if R is a field.
2. $\dim R = 0$ if and only if $\operatorname{Spec}(R) = \operatorname{Max}(R)$.
3. Let R be a PID but not a field. Then $\operatorname{Spec}(R) = \{(0)\} \cup \operatorname{Max}(R)$ with (0) not maximal, so $\dim R = 1$, e.g., $R = \mathbb{Z}$ or $F[t]$. More generally, any Dedekind domain is of dimension one.
4. $\dim F[t_1, \ldots, t_n] \geq n$, since
$$0 < (t_1) < (t_1, t_2) < \cdots < (t_1, \ldots, t_n)$$
is a (proper) chain of prime ideals. In fact, $\dim F[t_1, \ldots, t_n] = n$, a result that we shall show later.
5. $\dim F[t_1, \ldots, t_n, \ldots] = \infty$, but there do exist Noetherian rings of infinite dimension.
6. Recall a *local ring* is a commutative ring with a unique maximal ideal. [Such a ring can still have $\operatorname{Spec}(R)$ infinite, and in fact, this is the case if $\dim R > 1$ and R is Noetherian.] If R is a local ring with maximal ideal $\mathfrak{m}$, we say $(R, \mathfrak{m})$ is a local ring. We shall see that a local Noetherian ring has finite dimension.
7. Maximal ideals need not have the same height. For example, let $\mathfrak{p} = (3)$ in $\operatorname{Spec}(\mathbb{Z})$ and $\mathbb{Z}_{\mathfrak{p}}$ the localization of $\mathbb{Z}$ at $\mathfrak{p}$. Then $\mathbb{Z}_{\mathfrak{p}}[t]$ is a Noetherian UFD of dimension two containing maximal ideals $(3, t)$ and $(3t - 1)$, maximal ideals of height one and two, respectively.
8. Suppose that $\mathfrak{A} < R$ is an ideal. Then we have:
$$\operatorname{ht}\mathfrak{A} + \dim R/\mathfrak{A} \leq \dim R,$$
i.e.,
$$\operatorname{ht}\mathfrak{A} + \dim V(\mathfrak{A}) \leq \dim \operatorname{Spec}(R).$$

In general, inequality is possible. If we have equality, then we should view $\operatorname{ht}\mathfrak{A}$ as $\operatorname{codim} V(\mathfrak{A}) = \operatorname{codim}_{\operatorname{Spec}(R)} V(\mathfrak{A})$. One of our primary goals of this chapter is to establish this in the special case when R is an affine F-algebra, i.e., a finitely generated commutative F-algebra, and $\mathfrak{A}$ is a prime ideal.

9. If $\mathfrak{p}$ is a prime ideal in R, let $\mathfrak{p}R_{\mathfrak{p}} = \mathfrak{p}_{\mathfrak{p}}$, the localization of $\mathfrak{p}$ at $R \setminus \mathfrak{p}$. Then $(R_{\mathfrak{p}}, \mathfrak{p}_{\mathfrak{p}})$ is a local ring and satisfies $\dim R_{\mathfrak{p}} = \operatorname{ht} \mathfrak{p}R_{\mathfrak{p}}$.

Remark 92.30. We can also generalize the notions above to modules. Let M and N be R-modules. Then

$$(N : M) := \{x \in R \mid xM \subset N\}$$

is a ideal in R. E.g., $(0 : M) = \operatorname{ann}_R M$. We define the *dimension* of M by

$$\dim M := \dim V\big((0 : M)\big) = \dim V\big(\operatorname{ann}_R(M)\big)$$

and the *support* of M by

$$\operatorname{Supp}(M) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid M_{\mathfrak{p}} \neq 0\},$$

where $M_{\mathfrak{p}} = \{\frac{m}{s} \mid m \in M, s \in R \setminus \mathfrak{p}\}$, the localization of M at $\mathfrak{p}$, an $R_{\mathfrak{p}}$-module. [We leave the proof that we can localize an R-module to an $S^{-1}R$-module with S a multiplicative set in R as an exercise.]

Exercises 92.31.

1. Prove Proposition 92.7.
2. Prove Proposition 92.10.
3. Show Check 92.13 is true.
4. Check Remark 92.16.
5. Prove the assertions in Remark 92.19.
6. Let R be a nonzero commutative ring, S a multiplicative set in R, and M an R-module. Show that $S^{-1}M := \{\frac{r}{s} \mid r \in R,\ s \in S\}$ with the obvious definition is an $S^{-1}R$-module called the *localization* of M at S. If

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$

is an exact sequence of R-modules, prove that the sequence

$$0 \to S^{-1}M' \xrightarrow{S^{-1}f} S^{-1}M \xrightarrow{S^{-1}g} S^{-1}M'' \to 0$$

of $S^{-1}R$-modules is exact where, e.g., $S^{-1}f : S^{-1}M' \to S^{-1}M$, is defined by $S^{-1}f(\frac{r}{s}m) = \frac{r}{s}f(m)$, i.e., the $S^{-1}R$-module homomorphism induced by f. Moreover, if the original exact sequence is split (cf. Exercise 38.18(11)), so is the localized one.

7. In the previous exercise, if $S = R \setminus \mathfrak{p}$ with $\mathfrak{p}$ a prime ideal in R, and $f : M \to N$ is an R-homomorphism, let $f_{\mathfrak{p}}$ denote $S^{-1}f$. Then show $\ker(f_{\mathfrak{p}}) = (\ker f)_{\mathfrak{p}}$ and $\operatorname{coker}(f_{\mathfrak{p}}) = (\operatorname{coker} f)_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$ in R. Moreover, show that the following are equivalent:
   - (*i*) $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is an exact sequence of R-modules.
   - (*ii*) $0 \to M'_{\mathfrak{p}} \xrightarrow{f_{\mathfrak{p}}} M_{\mathfrak{p}} \xrightarrow{g_{\mathfrak{p}}} M''_{\mathfrak{p}} \to 0$ is an exact sequence of $R_{\mathfrak{p}}$-modules for all prime ideals $\mathfrak{p}$.

(*iii*) $0 \to M'_{\mathfrak{m}} \xrightarrow{f_{\mathfrak{m}}} M_{\mathfrak{m}} \xrightarrow{g_{\mathfrak{m}}} M''_{\mathfrak{m}} \to 0$ is an exact sequence of $R_{\mathfrak{m}}$-modules for all maximal ideals $\mathfrak{m}$.

In particular, f is an R-monomorphism (respectively, R-epimorphism) if and only if $f_{\mathfrak{p}}$ is an $R_{\mathfrak{p}}$-monomorphism (respectively $R_{\mathfrak{p}}$-epimorphism) for all prime ideals $\mathfrak{p}$ in R if and only if $f_{\mathfrak{p}}$ is an $R_{\mathfrak{m}}$-monomorphism (respectively $R_{\mathfrak{m}}$-epimorphism) for all maximal ideals $\mathfrak{m}$ in R. It follows that $M = 0$ if and only if $M_{\mathfrak{p}} = 0$ for all prime ideals $\mathfrak{p}$ in R if and only if $M_{\mathfrak{m}} = 0$ for all maximal ideals $\mathfrak{m}$ in R.

8. Let R be a nonzero commutative ring and $a_1, \dots, a_n \in R$. Show $D(a_1 \cdots a_n) = \cap_{i=1}^n D(a_i)$ and every open set in $\mathrm{Spec}(R)$ is a union of basic open sets.
9. Show the assertions in Example 92.25 are true.
10. Let X be a topological space. Show that X is Noetherian if and only if every open subset of X is quasi-compact.
11. For $f \in R$, let $X = \mathrm{Spec}(R)$ and $X_f = \mathrm{Spec}(R_f)$, where R_f the localization of R at the basic open set $D(f)$. Write $R(X_f)$ for the localization R_f.
For $f, g, h \in R$ and $U = X_f,\ U' = X_g,\ U'' = X_h,$ show the following:
   (a) $R(U)$ depends only on U and not on f.
   (b) If $U' \subset U$, then there exists a positive integer n and an element $x \in R$ such that $g^n = xf$. Using this, defines a homomorphism $\rho_{UU'} : R(U) \to R(U')$ via $a/f^m \mapsto ax^m/g^{mn}$. This map depends only on U and U' and is called the *restriction homomorphism.*
   (c) If $U = U'$, then ρ_{UU} is the identity map.
   (d) If $U'' \subset U' \subset U$, then the diagram

$$\begin{tikzcd} R(U) \arrow[rr, "\rho_{UU''}"] \arrow[dr, "\rho_{UU'}"'] & & R(U'') \\ & R(U') \arrow[ur, "\rho_{U'U''}"'] & \end{tikzcd}$$

   commutes. This implies that $X := (X, R)$ is a *presheaf of rings.*
   (e) Let $\{U_i \,|\, i \in I\}$ be a finite covering of a basic open set U in X by basic open sets. Suppose that $a \in U$, $a_i \in U_i$ for each i. Then
      (i) If $a \in R(U)$ satisfies $\rho_{UU_i}(a) = 0$ for all i, then $a = 0$.
      (ii) If $\rho_{U_i(U_i \cap U_j)}(a_i) = \rho_{U_j(U_i \cap U_j)}(a_j)$ in $R(U_i \cap U_j)$ for all i, j, then there exists an element $r \in R(U)$ satisfying $\rho_{UU_i}(r) = a_i$ for all i.

   This implies that X is a *sheaf of rings.* We call (X, R) an *affine scheme.*
   [If $x = \mathfrak{p}$, then the local ring $R_{\mathfrak{p}}$ is the *stalk* of X at $\mathfrak{p}$. It is the set of *germs at* $\mathfrak{p}$ on X, i.e., equivalence classes of elements $\overline{r}$ where r is defined in some $R(U)$, U a basic open set, and if r' is defined in $R(U')$, then $r \sim r'$ if there exists a basic open neighborhood V of $\mathfrak{p}$ in $U \cap U'$ such that $\rho_{UV}(r) = \rho_{U'V}(r')$ in $R(V)$.]

93. Integral Extensions of Commutative Rings

As is true in this whole chapter, R is a commutative ring.

In this section, we look further at the notion of integral extensions. We use the material studied in Sections 79 and 80.

Notation 93.1. We begin with some new notation. If A is a commutative R-algebra via the ring homomorphism $\varphi_A : R \to A$, we shall write ra for $\varphi(r)a$, $r \in R$, $a \in A$. If $a_1, \dots, a_n$. are elements of A, then we shall write $R[a_1, \dots, a_n]$ for $\varphi(R)[a_1, \dots, a_n]$ (even if φ_R is not a monomorphism). If $S \subset R$ is a multiplicative set, we shall write $S^{-1}A$ for $\big(\varphi(S)\big)^{-1}A$. So if $r \in R$, $s \in S$, and $a \in A$, we write $\frac{r}{s}a$ for $\frac{\varphi_A(r)}{\varphi_A(s)}a$. If $\mathfrak{p}$ is a prime ideal in R, we write $A_\mathfrak{p}$ for $S^{-1}A$ where $S = R \setminus \mathfrak{p}$.

Definition 93.2. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings, we shall view A as an R-algebra via φ unless otherwise stated. We say

(1) φ is of *finite type* if A is a finitely generated R-algebra.
(2) φ is *finite* if A is a finitely generated R-module (via φ).
[One also calls a finitely generated R-module a *finite R-module*.]
(3) φ is *integral* if $A/\varphi(R)$ is an integral extension.

We leave the various remarks about ring homomorphisms of commutative rings and localization as useful exercises for the reader.

Remarks 93.3. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings.

1. φ is a finite map if and only if it is integral and finite type. (Cf. Exercise 79.15(4).)
2. If φ is an epimorphism, then φ is finite, hence integral.
3. If φ is integral and $\mathfrak{A}$ an ideal in A, then the induced map $\widetilde{\varphi} : R/\varphi^{-1}(\mathfrak{A}) \to A/\mathfrak{A}$ is integral and injective.
4. The composition of ring hommormorphisms of finite type (respectively, finite, integral) are of finite type (respectively, finite, integral).
5. Suppose that φ is the inclusion and $\mathfrak{A} < R$ an ideal. As before, we write A/R. It is possible that $\mathfrak{A}A = A$ and $\mathfrak{A} < (\mathfrak{A}A) \cap R$. For example, (2) is a prime ideal in $\mathbb{Z}$ but $(2)\mathbb{Q} = \mathbb{Q}$. We shall see that this does not happen if A/R is integral.
6. Suppose that φ is the inclusion with R a domain and $K = A$ a field containing R. As in §79, we shall let R_K denote the integral closure of R in K. In this chapter, if R is integrally closed in its quotient field F, i.e., $R = R_F$, we shall use geometric language and say that R is a *normal domain* instead of saying R is integrally closed.

Remarks 93.4. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings and S a multiplicative subset of R.

1. If φ is a monomorphism, so is $\varphi_{S^{-1}R} : S^{-1}R \to S^{-1}A$. In particular, this applies if φ is the inclusion map. We then view $\varphi_{S^{-1}R}$ as the inclusion. If, in addition, $A = K$ is a field $qf(R)$ and $\mathfrak{p}$ a prime ideal in R, then $(R_\mathfrak{p})_K = (R_K)_\mathfrak{p}$.
2. If φ is is of finite type (respectively, finite, integral), then so is $\varphi_{S^{-1}R}$ in which case the integral closure of $S^{-1}R$ in $S^{-1}A$ is $S^{-1}C$ where C is the integral closure of R in A.

In the more elementary part of this tome, we proved the Hilbert Nullstellsatz. We wish to give deeper insight into why this theorem (rather than the proof we gave) is true.

One of the keys to the proof, essentially an equivalent formulation, was Zariski's Lemma 41.10 whose proof we shall derive again later. To do so we must look carefully at integral extensions. We begin with a summary and the methodology that we shall use.

Summary 93.5. Let $i : A \subset B$ be the inclusion of two commutative rings and $S \subset A$ a multiplicative set.

1. $S^{-1}i : S^{-1}A \to S^{-1}B$ is a ring monomorphism, that we view as an inclusion.
2. Let $S^{-1}\varphi_A : A \to S^{-1}A$ by $a \to \frac{a}{1}$, be the canonical homomorphism. Then ${}^a\varphi_A : \operatorname{Spec} S^{-1}A \to \{\mathfrak{p} \in \operatorname{Spec}(A) \mid \mathfrak{p} \cap S = \emptyset\}$ is a homeomorphism. In particular, if $\mathfrak{p}$ is a prime ideal in A such that $\mathfrak{p} \cap S = \emptyset$, then $\mathfrak{p} = (S^{-1}\mathfrak{p}) \cap A := {}^a\varphi_A(S^{-1}\mathfrak{p})$.
3. If $\mathfrak{p}$ is a prime ideal in A, then $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p})$ is a local ring.
4. A major methodology for trying to prove commutative algebra results is by:
   (a) **Localize**: If $\mathfrak{p}$ is a prime ideal, then $A_\mathfrak{p} \subset B_\mathfrak{p}$ with $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p})$ a local ring. Then try to prove the results here and pullback.
   (b) **Quotient**: If $\mathfrak{P}$ is a prime ideal in B, let $\mathfrak{p} = \mathfrak{P} \cap A$ a prime ideal in A with the induced map $A/\mathfrak{p} \to B/\mathfrak{P}$ a homomorphism of domains. Then try to prove the results here and pullback.

Lemma 93.6. *Let A be a commutative R-algebra and u a unit in A. Then u^{-1} is integral over R if and only if u^{-1} lies in $R[u]$.*

PROOF. The element u^{-1} is integral over R if and only if there exists an equation

$$u^{-n} + r_1u^{-n+1} + \cdots + r_n = 0 \text{ in } A \text{ for some } r_1, \ldots, r_n \in R$$

if and only if

$$u^{-1}(r_1 + r_2u + \cdots + r_nu^{n-1}) = -1 \text{ in } A \text{ for some } r_1, \ldots, r_n \in R$$

if and only if u^{-1} lies in $R[u]$. □

Proposition 93.7. *Let $A \subset B$ be domains with B/A integral. Then A is a field if and only if B is a field.*

PROOF. ($\Leftarrow$): Let $x \in A$ be nonzero. Then x^{-1} lies in the field B and is integral over A if and only if $x^{-1} \in A[x] = A$ by the lemma.
($\Rightarrow$): Let $y \in B$ be nonzero. Then there exists $a_1, \ldots, a_n$ in A such that $y^n + a_1y^{n-1} + \cdots + a_n = 0$ in B for some n, as B/A is integral. Since B is a domain, we may assume that a_n is nonzero. Therefore, a_n^{-1} lies in $A \subset qf(B)$, the quotient field of B, as A is a field. Consequently, $y^{-1} = -a_n^{-1}(y^{n-1} + \cdots + a_{n-1})$ in $qf(B)$, hence lies in $A[y] \subset B$. □

Corollary 93.8. *Let $\varphi : A \to B$ be an integral ring homomorphism, $\mathfrak{P}$ a prime ideal in B. Then $\mathfrak{P}$ is a maximal ideal in B if and only if ${}^a\varphi(\mathfrak{P})$ is a maximal ideal in A. In particular, the restriction of ${}^a\varphi$ to $\operatorname{Max}(B)$ gives the map*

$${}^a\varphi|_{\operatorname{Max}(B)} : \operatorname{Max}(B) \to \operatorname{Max}(A).$$

PROOF. let $\mathfrak{P}$ be a prime ideal in B. Then φ induces a monomorphism $\overline{\varphi} : A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$ which is integral by Remark 93.3(3). □

Definition 93.9. Let R be a commutative ring. The *Jacobson radical* of R is defined to be $\operatorname{rad}(R) := \bigcap_{\operatorname{Max}(R)} \mathfrak{m}$.

One of the most useful results is the following lemma, apparently first proved independently by Azumaya and Krull:

Lemma 93.10. (Nakayama's Lemma) *Let $\mathfrak{A}$ be an ideal in* $\operatorname{rad}(R)$ *and M a finitely generated R-module. If $M = \mathfrak{A}M$, then $M = 0$.*

PROOF. Suppose that $M \neq 0$ and $M = \sum_{i=1}^{n} Rm_i$ with n minimal. Then there exist $a_1, \dots, a_n$ in $\mathfrak{A}$ satisfying $m_1 = \sum_{i=1}^{n} a_i m_i$, as $M = \mathfrak{A}M$. Thus $(1 - a_1)m_1 = \sum_{i=2}^{n} a_i m_i$. Since each a_i lies in $\operatorname{rad}(R) \subset \mathfrak{m}$ for every maximal ideal $\mathfrak{m}$ in R, we have $1 - a_1 \notin \mathfrak{m}$ for every maximal ideal $\mathfrak{m}$, hence is a unit in R. It follows that m_1 lies in $\sum_{i=2}^{n} Rm_i$, contradicting the minimality of n. $\square$

The following consequence is also called Nakayama's Lemma.

Corollary 93.11. *Let $\mathfrak{A} \subset \operatorname{rad}(R)$ be an ideal in R, M a finitely generated R-module, and $N \subset M$ a submodule. If $M = N + \mathfrak{A}M$, then $M = N$.*

PROOF. We have M/N is a finitely generated R-module satisfying $M/N = \mathfrak{A}(M/N)$, so $M/N = 0$ by Nakayama's Lemma. Hence $M = N$. $\square$

We leave the proof of the next useful corollary as an exercise.

Corollary 93.12. *Let $(R, \mathfrak{m})$ be a local ring, M a finitely generated R-module, $^{-} : M \to M/\mathfrak{m}M$, the canonical epimorphism,*

$$\mathcal{S} = \{x_1, \dots, x_n\}, \quad \textit{and } \overline{\mathcal{S}} = \{\overline{x_1}, \dots, \overline{x_n}\}.$$

Then

1. *$\mathcal{S}$ generates M if and only if $\overline{\mathcal{S}}$ spans the $R/\mathfrak{m}$-vector space $\overline{M} = M/\mathfrak{m}M$.*
2. *$\mathcal{S}$ is a minimal generating set (obvious definition) for M if and only if $\overline{\mathcal{S}}$ is an $R/\mathfrak{m}$-basis for $\overline{M}$ (and $\dim_{R/\mathfrak{m}} \overline{M} = n$).*

Remark 93.13. We shall see later that if $(R, \mathfrak{m})$ is a Noetherian local ring that

$$\dim R = \operatorname{ht} \mathfrak{m} \leq \dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2 < \infty.$$

If we have equality in the above, then $(R, \mathfrak{m})$ is called a *regular local ring.* This notion of regular is the algebraic replacement for the geometric concept of non-singularity at a point.

A basic theorem in the study of integral extensions, apparently originally due to Krull, is the next result that we now prove. The proof is an excellent illustration of the methods mentioned above.

Theorem 93.14. (Cohen-Seidenberg Theorem)
Let B/A be integral. Then:

1. (Incomparability) *If $\mathfrak{P}_1 \subset \mathfrak{P}_2$ are prime ideals in B, and satisfy $\mathfrak{P}_1 \cap A = \mathfrak{P}_2 \cap A$, then $\mathfrak{P}_1 = \mathfrak{P}_2$.*
2. (Lying Over) *${}^a i : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is surjective, where $i : A \to B$ is the inclusion map, i.e., (in the notation of Section 82 that we shall continue to use) if $\mathfrak{p}$ is a prime ideal in A, there exists a prime ideal $\mathfrak{P}$ in B lying over $\mathfrak{p}$.*

(3) (Going Up). *If $\mathfrak{p}_1 \subset \mathfrak{p}_2$ are prime ideals in A and $\mathfrak{P}_1$ is a prime ideal in B lying over $\mathfrak{p}_1$, then there exists a prime ideal in $V(\mathfrak{P}_1)$ lying over $\mathfrak{p}_2$.*

$$\text{(Picture)} \qquad \begin{array}{ccc} \mathfrak{P}_1 & \subset & \boxed{} \\ | & & | \\ \mathfrak{p}_1 & \subset & \mathfrak{p}_2. \end{array}$$

(4) $\dim A = \dim B$.

Proof. (1): Let $\mathfrak{p} = \mathfrak{P}_1 \cap A = \mathfrak{P}_2 \cap A$ in $\operatorname{Spec}(A)$ and $S = A \setminus \mathfrak{p}$. Then $S \cap \mathfrak{P}_i = \emptyset$ for $i = 1, 2$, so $S^{-1}B/S^{-1}A$ is integral, $(S^{-1}A, S^{-1}\mathfrak{p})$ is a local ring with

$$S^{-1}\mathfrak{P}_i \cap S^{-1}A = S^{-1}(\mathfrak{P}_i \cap A) = S^{-1}\mathfrak{p}.$$

Therefore, $S^{-1}\mathfrak{P}_i \in \operatorname{Spec}(S^{-1}B)$ lies over $S^{-1}\mathfrak{p}$, for $i = 1, 2$. By Corollary 93.8, $S^{-1}\mathfrak{P}_i$ is a maximal ideal in $S^{-1}B$ for $i = 1, 2$. It follows that $\mathfrak{P}_1 \subset \mathfrak{P}_2$ implies $S^{-1}\mathfrak{P}_1 = S^{-1}\mathfrak{P}_2$. Consequently, $\mathfrak{P}_1 = \mathfrak{P}_2$ as $\mathfrak{P}_i \cap S = \emptyset$ for $i = 1, 2$.

(2): Let $S = A \setminus \mathfrak{p}$, then $S^{-1}B/S^{-1}A$ is integral. By the diagram

$$\begin{array}{ccc} B & \xrightarrow{\text{loc}} & S^{-1}B \\ | & & | \\ A & \xrightarrow[\text{loc}]{} & S^{-1}A \end{array} \qquad \text{commutes} \quad \text{and} \qquad \begin{array}{c} S^{-1}\mathfrak{P} \\ | \\ S^{-1}\mathfrak{p} \end{array}$$

if $S^{-1}\mathfrak{P}$ lies over $S^{-1}\mathfrak{p}$, we see that it suffices to replace A by $S^{-1}A$ and assume that $(A, \mathfrak{p})$ is a local ring with B/A integral. In particular, it suffices to show that there exists a prime ideal $\mathfrak{P}$ in B lying over $\mathfrak{p}$. Suppose that $\mathfrak{p}B < B$. Then there exists an ideal $\mathfrak{m} \in V(\mathfrak{p}B) \cap \operatorname{Max}(B)$. Since B/A is integral, $\mathfrak{m} \cap A \in \operatorname{Max}(A) = \{\mathfrak{p}\}$, i.e., $\mathfrak{m} \cap A = \mathfrak{p}$ and we would be done. So we may assume that $\mathfrak{p}B = B$. We then have an equation

$$1 = \sum_{i=1}^{m} p_i b_i \text{ for some } p_i \in \mathfrak{p},\ b_i \in B.$$

Set $B' = A[b_1, \dots, b_m]$ a finitely generated commutative A-algebra. As B/A is integral, B' is a finitely generated A-module satisfying $\mathfrak{p}B' = B'$. It follows by Nakayama's Lemma, that $B' = 0$, which is impossible. Hence $\mathfrak{p}B < B$ and (2) follows.

(3): Let $\mathfrak{P}_1 \in \operatorname{Spec}(B)$ lie over $\mathfrak{p}_1 \in \operatorname{Spec}(A)$. Since the inclusion $A \subset B$ induces an integral monomorphism $A/\mathfrak{p}_1 \to B/\mathfrak{P}_1$, by Lying Over and the Correspondence Principle, there exists a prime ideal $\mathfrak{P}_2$ in $V(\mathfrak{P}_1)$ with $\mathfrak{P}_2/\mathfrak{P}_1$ lying over $\mathfrak{p}_2/\mathfrak{p}_1$ in $\operatorname{Spec}(A/\mathfrak{p}_1)$. Therefore, $\mathfrak{P}_2$ lies over $\mathfrak{p}_1$ by the Correspondence Principle and (3) follows.

(2): Let

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n$$

be a chain of prime ideals in B. By Incomparability,

$$\mathfrak{P}_1 \cap A < \cdots < \mathfrak{P}_n \cap A$$

is a chain of prime ideals in A, so $\dim B \leq \dim A$. If

$$\mathfrak{p}_1 < \cdots < \mathfrak{p}_n$$

is a chain of prime ideals in A, then by Lying Over and Going Up, we can construct a chain of prime ideals

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n$$

in B, so $\dim A \leq \dim B$. Therefore, $\dim A = \dim B$. □

In the theorem, we needed the condition that $A \subset B$, i.e., that we have a monomorphism $A \to B$. Indeed a counterexample is provided by the canonical epimorphism $^{-} : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, $n > 1$ if not.

Corollary 93.15. *Let B/A be an integral extension with B a domain and $\mathfrak{B} < B$ a nonzero ideal. Then $\mathfrak{B} \cap A < A$ is a nonzero ideal.*

Proof. Let $\mathfrak{P} \in V(\mathfrak{B})$. Certainly, $\mathfrak{P} \cap A < A$ as $1 \notin \mathfrak{P}$. Since A and B are domains, the zero ideal in B lies over the zero ideal in A. By Incomparability, $\mathfrak{P}$ does not lie over (0). Suppose that $\mathfrak{B} \cap A = 0$. Then the induced mononomorphism $A \hookrightarrow B/\mathfrak{B}$ is integral. Therefore, by Lying Over and the Correspondence Principle, there exists a prime ideal $\mathfrak{P} \in V(\mathfrak{B})$ satisfying $\mathfrak{P}/\mathfrak{B}$ lies over (0) in $\operatorname{Spec}(A)$. As $\mathfrak{B} \cap A = 0$, this means that $\mathfrak{P}$ lies over (0), a contradiction. □

We need to generalize Proposition 83.1, whose former proof can be adapted to this generalization (essentially by localization). Moreover, we shall give a different proof that uses one of the most useful results, called the Prime Avoidance Lemma that we previously left as an exercise, but now prove.

Lemma 93.16. (Prime Avoidance Lemma) *Let $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ be ideals in R, at least $n-2$ of which are prime. Let $S \subset R$ be a subrng (it does not have to have a 1) contained in $\mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$. Then there exists a j such that $S \subset \mathfrak{A}_j$. In particular, if $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ are prime ideals in R and $\mathfrak{B}$ is an ideal properly contained in S satisfying $S \setminus \mathfrak{B} \subset \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, then S lies in one of the $\mathfrak{p}_i$'s.*

Proof. The case $n = 1$ is trivial, so assume that $n > 1$ and assume that the result is false. By induction, $S \not\subset \mathfrak{A}_1 \cup \cdots \widehat{\mathfrak{A}_i} \cup \cdots \cup \mathfrak{A}_n$ for each i where $\widehat{\ }$ means omit. So for each $i = 1, \dots, n$, there exists an element $x_i \in S \setminus \mathfrak{A}_1 \cup \cdots \widehat{\mathfrak{A}_i} \cup \cdots \cup \mathfrak{A}_n$. Therefore, we have $x_i \in \mathfrak{A}_i$ for every $i = 1, \dots, n$ and $x_i \notin \mathfrak{A}_j$ for all $j \neq i$. If $n > 2$, we may assume that $\mathfrak{A}_1$ is a prime ideal. Let $y = x_1 + x_2 \cdots x_n \in S$. If $n = 2$, then $y = x_1 + x_2$ does not lie in $\mathfrak{A}_1 + \mathfrak{A}_2$, a contradiction. So we may assume that $n > 2$. As $\mathfrak{A}_1$ is a prime ideal and $x_i \notin \mathfrak{A}_1$ for all $i > 1$, we have $x_2 \cdots x_n \notin \mathfrak{A}_1$. Consequently, $y \notin \mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$, a contradiction. □

Our generalization of Proposition 83.1 needs us to replace the condition of a finite Galois extension of fields by an arbitrary normal extension of fields. This requires that we extend the definition of the norm of a finite separable field extension. Let L/F be a finite normal extension of fields, and K the separable closure of F in K. We call $[L:F]_i := [L:K]$ the *purely inseparable degree* of L/F and $[L:F]_s := [K:F]$ the *separable degree* of F in L. So $[L:F] = [L:F]_i[L:F]_s$. Define the norm $\mathrm{N}_{L/F} : L \to F$ by $\mathrm{N}_{L/F}(x) = \left(\prod_{G(L/F)} \sigma(x)\right)^{[L:F]_i}$, i.e., $\mathrm{N}_{L/K}(x) = \mathrm{N}_{K/F}(x^{[L:F]_i})$. Then the norm still satisfies all the expected properties.

Theorem 93.17. *Let A be a normal domain (i.e., an integrally closed domain) with quotient field F and L/F a normal extension of fields. Let $\mathfrak{P}_1$ and $\mathfrak{P}_2$ be prime ideals in A_L lying over $\mathfrak{p}$ in $\operatorname{Spec}(A)$, i.e., $\mathfrak{P}_1 \cap A = \mathfrak{p} = \mathfrak{P}_2 \cap A$. Then there exists an element σ in $G(L/F)$, the Galois group of L/F, satisfying $\mathfrak{P}_1 = \sigma(\mathfrak{P}_2)$, i.e., if $i : A \to A_L$ is the inclusion map, then $G(L/F)$ acts transitively on the fibers of ${}^a i : \operatorname{Spec}(A_L) \to \operatorname{Spec}(A)$.*

Proof. If $b \in A_L$ and $\sigma \in G(L/F)$, then $\sigma(b) \in L$ is integral over $\sigma(A) = A$, i.e., $\sigma(b) \in \big(\sigma(A)\big)_L = A_L$, so $\sigma(A_L) = A_L$. As σ^{-1} lies in $G(L/F)$, the map $\sigma : A_L \to A_L$ is an A-algebra isomomorphism. If $\mathfrak{P}$ is a prime ideal in A_L and $\sigma \in G(L/F)$, then $\sigma(\mathfrak{P}) \in \operatorname{Spec}(A_L)$ and $\mathfrak{P} \cap A = \sigma(\mathfrak{P} \cap A) = \sigma(\mathfrak{P}) \cap A$. So every prime ideal in the orbit $G(L/F)\mathfrak{P} := \{\sigma(\mathfrak{P}) \mid \sigma \in G(L/F)\}$ lies in the fiber $({}^a i)^{-1}(\mathfrak{p})$ of $\mathfrak{p}$.

Case 1. L/F is finite.

Suppose $\mathfrak{P}_2$ lies over $\mathfrak{P}_1 \cap A$. To show that $\mathfrak{P}_2 \in G(L/F)\mathfrak{P}_1$. Suppose not. By Incomparability, $\mathfrak{P}_2 \not\subset \sigma(\mathfrak{P}_1)$ for any $\sigma \in G(L/F)$. By the Prime Avoidance Lemma, as $G(L/F)\mathfrak{P}_1$ is finite, $\mathfrak{P}_2 \not\subset \bigcup_{G(L/F)} \sigma(\mathfrak{P}_1)$. Let a be an element in $\mathfrak{P}_2 \setminus \bigcup_{G(L/F)} \sigma(\mathfrak{P}_1)$. Then $\sigma(a) \notin \mathfrak{P}_1$ for any $\sigma \in G(L/F)$. As A is a normal domain, we have

$$\mathrm{N}_{L/F}(a) = \Big(\prod_{G(L/F)} \sigma(a)\Big)^{[L:F]_i} \text{ lies in } A_L \cap F = A_F = A.$$

Therefore,

$$\mathrm{N}_{L/F}(a) \in \mathfrak{P}_2 \cap A = \mathfrak{p} = \mathfrak{P}_1 \cap A \subset \mathfrak{P}_1.$$

As $\mathfrak{P}_1$ is a prime ideal and $G(L/F)$ is a finite group, there exists an σ in $G(L/F)$ satisfying $\sigma(a) \in \mathfrak{P}_1$, a contradiction.

Case 2. L/F is possibly infinite.

This is left an exercise (use Zorn's Lemma, since L is a union of finite normal extensions). □

Corollary 93.18. *Let A be a normal domain with quotient field F and K/F a finite field extension. Then ${}^a i : \operatorname{Spec}(A_K) \to \operatorname{Spec}(A)$ has finite fibers, where i is the inclusion map of A in A_K, i.e., $({}^a i)^{-1}(\mathfrak{p})$ is a finite set for all prime ideals $\mathfrak{p}$ in A.*

A ring homomorphism whose associated map has finite fibers is called *quasi-finite*.

Proof. Let L/F be a normal closure of K/F. Then L/F is a finite normal extension, hence $G(L/F)$ is finite. Let $\mathfrak{p}$ be a prime ideal in A. By Lying Over, there exists a prime ideal $\mathfrak{P}$ in A_L lying over $\mathfrak{p}$. We have $({}^a i_L)^{-1}(\mathfrak{p}) = \{\sigma(\mathfrak{P}) \mid \sigma \in G(L/F)\}$ where $i_L : A \to A_L$ is the inclusion map. By Lying Over, if $\mathfrak{Q}$ is a prime ideal in A_K, then there exists a prime ideal $\mathfrak{P}'$ in A_L lying over $\mathfrak{Q}$ as A_L/A_K is integral. Hence

$$|({}^a i)^{-1}(\mathfrak{p})| \le |({}^a i_L)^{-1}(\mathfrak{p})| \le |G(L/F)| < \infty$$

where $i : A \to A_K$ is the inclusion. □

Definition 93.19. A commutative ring R with finitely many maximal ideals is called *semi-local ring*.

Corollary 93.20. *Let A be a semi-local normal domain with quotient field F and K/F a finite field extension. Then A_K is semi-local.*

PROOF. As A_K/A is integral, ${}^a i|_{\mathrm{Max}(A)_K} : \mathrm{Max}(A_K) \to \mathrm{Max}(A)$, where i is the inclusion of A in A_K. As ${}^a i$ has finite fibers, the result follows. $\square$

Proposition 93.21. *Let A be a normal domain with quotient field F and K/F a finite field extension. Suppose that $\mathfrak{P}$, a prime ideal in A_K, lies over the prime ideal $\mathfrak{p}$ in A. Let $\overline{F}$ denote the quotient field of $A/\mathfrak{p}$. Then for all $\alpha \in A_K/\mathfrak{P}$, we have $[\overline{F}(\alpha) : \overline{F}] \leq [K : F]$.*

PROOF. Let $^- : A[t] \to (A/\mathfrak{p})[t] \subset \overline{F}[t]$ be the canonical epimorphism and $x \in A_K$ an element satisfying $\alpha = \underline{x + \mathfrak{P}}$ in $A_K/\mathfrak{P}$. Then the minimal polynomial $m_F(x)$ of x in $A_F[t] = A[t]$ is monic and $\overline{m_F(x)}|_{t=\alpha} = 0$, so

$$[\overline{F}(\alpha) : \overline{F}] \leq \deg_{\overline{F}} m_{\overline{F}}(\alpha) \leq \deg_F m_F(x) \leq [K : F]. \qquad \square$$

Corollary 93.22. *Let A be a normal domain with quotient field F and K/F a finite field extension. Suppose that there exists an element x in A_K satisfying $A_K = A[x]$. Let $\mathfrak{P}$ be a prime ideal in A_K lying over the prime ideal $\mathfrak{p}$ in A. Set $\overline{K} = qf(A_K/\mathfrak{P})$ and $\overline{F} = qf(A/\mathfrak{p})$. Then $[\overline{K} : \overline{F}] \leq [K : F]$.*

PROOF. $\overline{K} = qf(A/\mathfrak{p})[x + \mathfrak{P}]$. $\square$

The Cohen-Seidenberg Theorem showed that if B/A is integral, then $\dim A = \dim B$. We are also interested in the heights of primes, as we wish to determine the codimension of irreducible subvarieties. This is more delicate and needs stronger hypotheses. We can now determine one such result.

Theorem 93.23. (Going Down Theorem) *Let B/A be integral with A a normal domain and B a domain. Let $\mathfrak{p}_1 \subset \mathfrak{p}_2$ be prime ideals in A and $\mathfrak{P}_2$ a prime ideal in B lying over $\mathfrak{p}_2$. Then there exists a prime ideal $\mathfrak{P}_1$ lying over $\mathfrak{p}_1$ satisfying $\mathfrak{P}_1 \subset \mathfrak{P}_2$. In particular, if $\mathfrak{P}$ is a prime ideal in B, then $\mathrm{ht}\,\mathfrak{P} = \mathrm{ht}(\mathfrak{P} \cap A)$.*

$$\text{(Picture)} \qquad \begin{matrix} \boxed{} & \subset & \mathfrak{P}_2 \\ | & & | \\ \mathfrak{p}_1 & \subset & \mathfrak{p}_2 \end{matrix}$$

PROOF. Let F be the quotient field of A and K the quotient field of B. Since B/A is integral, K/F is algebraic. Let L/K be a normal closure of K/F. As A_L/A is integral and $B \subset A_L$, we have A_L/B is also integral. Let $\mathfrak{Q}_1 \in \mathrm{Spec}(A_L)$ lie over $\mathfrak{p}_1$. By Going Up, there exists a prime ideal $\mathfrak{Q}_2 \in V(\mathfrak{Q}_1)$ lying over $\mathfrak{p}_2$. By Lying Over there exists a prime ideal $\mathfrak{Q}_2'$ in A_L lying over $\mathfrak{P}_2$ hence over $\mathfrak{p}_2$. Therefore, there exists an element $\sigma \in G(L/F)$ satisfying $\sigma(\mathfrak{Q}_2) = \mathfrak{Q}_2'$. We also have $\sigma(\mathfrak{Q}_1) \cap A = \mathfrak{p}_1 = \mathfrak{Q}_1 \cap A$. It follows that $\mathfrak{P}_1 = \sigma(\mathfrak{Q}_1) \cap B$ works for the first statement.
Let

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n = \mathfrak{P}$$

be a chain of prime ideals in B and $\mathfrak{p} = \mathfrak{P} \cap A$. By Incomparability, we have

$$\mathfrak{P}_1 \cap A < \cdots < \mathfrak{P}_n \cap A = \mathfrak{P} \cap A = \mathfrak{p}$$

a chain of prime ideals in A. Therefore, $\mathrm{ht}(\mathfrak{P} \cap A) \geq \mathrm{ht}\,\mathfrak{P}$. If

$$\mathfrak{p}_1 < \cdots < \mathfrak{p}_n = \mathfrak{p} = \mathfrak{P} \cap A$$

is a chain of prime ideals in A, then, by Going Down, there exists a chain of prime ideals

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n = \mathfrak{P}$$

in B, so $\operatorname{ht}\mathfrak{P} \geq \operatorname{ht}(\mathfrak{P} \cap A)$, and the result follows. $\square$

Corollary 93.24. *Let B/A be integral with A a normal domain and B a domain. Let $\mathfrak{p}$ be a prime ideal in A and $\mathfrak{P}$ a prime ideal in B lying over $\mathfrak{p}$. Then $\operatorname{ht}\mathfrak{p} = 1$ if and only if $\operatorname{ht}\mathfrak{P} = 1$,*

Remark 93.25. The Going Down Theorem 93.23 can be strengthened. In the notation of Theorem 93.23, we can weaken the hypothesis on B to the condition that B be A-torsion-free. Indeed, let

$$S = (A \setminus \{0\})(B \setminus \mathfrak{P}_2) \subset B.$$

Then S is a multiplicative set excluding (0), since A-torsion-free. Hence there exists a prime ideal $\mathfrak{P}$ in B excluding S. It follows that $\mathfrak{P}$ satisfies $\mathfrak{P} \subset \mathfrak{P}_2$ and $\mathfrak{P} \cap A = (0)$. As B/A induces a ring monomorphism $A/(\mathfrak{P} \cap A) \to B/\mathfrak{P}$, which we view as an inclusion, we are in the situation of the Going Down Theorem. In particular, there exists a prime $\mathfrak{P}_1$ containing $\mathfrak{P}$ such that $\mathfrak{P}_1/\mathfrak{P}$ lies over $\mathfrak{p}_1$ and $\mathfrak{P}_1/\mathfrak{P} \subset \mathfrak{P}_2/\mathfrak{P}$. Therefore, $\mathfrak{P}_1$ works.

Exercises 93.26.

1. Prove the assertions in Remarks 93.3.
2. Prove the assertions in Remarks 93.4.
3. Let $\varphi : A \to B$ be an integral homomorphism. Show that the associated map ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a closed map.
4. Prove Corollary 93.12.
5. Let $(R, \mathfrak{m})$ be a local ring. Show that any finitely generated R-projective module (cf. Exercise 39.12(126.1) and exercises that follow) is a free R-module.
6. Suppose that $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ are prime ideals in R and $\mathfrak{P} = \cap_i \mathfrak{p}_i$. Show if $\mathfrak{P}$ is a prime ideal, then there exists an i with $\mathfrak{P} = \mathfrak{p}_i$.
7. Prove Case 2 of Theorem 93.17.
8. (Cayley-Hamilton Theorem) Let R be a commutative ring, $\mathfrak{A} < R$ an ideal, and M an R-module that can be generated by n elements. If $\varphi \in \operatorname{End}_R(M)$ satisfies, $\varphi(M) \subset \mathfrak{A}M$, then there exists a monic polynomial $f = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $R[t]$ with coefficients $a_0, \dots, a_{n-1} \in \mathfrak{A}$, such that the R-endomorphism on M, $f(\varphi) = \varphi^n + a_{n-1}\varphi^{n-1} + \cdots + a_0 1_M = 0$.
   [The characterization of integral elements (as well as Nakayama's Lemma 93.10) easy follow from this.]
9. Let $\mathfrak{A}$ be an ideal in $R[t]$ and $^- : R[t] \to R[t]/\mathfrak{A}$, the canonical epimorphism. Set $x = \bar{t}$ and $A = F[t]/\mathfrak{A}$. Prove the following:
   (i) As an R-module A is generated by at most n elements if and only if there exists a monic polynomial $f \in \mathfrak{A}$ of degree at most n. If this is the case, then $A = \sum_{i=0}^{n-1} Rx_i$, for some $x_0, \dots, x_{n-1}$. In particular, A is a finitely generated R-module if and only if $\mathfrak{A}$ contains a monic polynomial.

(*ii*) A is a free R-module if and only if $\mathfrak{A} = (f)$ for some monic polynomial $f \in R[t]$. If this is the case, say $\operatorname{rank} A = n$. Then there exists an $x \in A$ such that $\mathcal{B} = \{1, x, \dots, x^{n-1}\}$ is a basis for A.

10. Under the conditions of Remark 93.25, show if $\mathfrak{p}$ is a prime ideal in A and $i : A \to B$ is the inclusion map, then
$$({}^a i)^{-1}(\mathfrak{p}) = \{\mathfrak{P} \in \operatorname{Spec}(B) \mid \mathfrak{P} \in V(\mathfrak{p}B) \text{ is minimal}\}.$$
11. Let R be a subring of a commutative ring A. Show that the Going Down Theorem holds for A/R if ${}^a i : \operatorname{Spec}(A) \to \operatorname{Spec}(R)$ is an open map. [The converse of this is true if R is a Noetherian ring.]

94. Primary Decomposition

As throughout this chapter, R will denote a commutative ring.

In this section, we investigate the generalization of unique factorization of ideals that Dedekind domains satisfy called the Lasker-Noether Theorem. As unique factorization of ideals into a product of prime ideals characterize Dedekind domains, the appropriate generalization is, as one would expect, weaker. Multiplication of ideals is replaced by intersections, powers of primes by primary ideals, and not all primary ideals are unique. However, this generalization will apply to any Noetherian ring, not just a Noetherian domain.

We shall need the following lemma whose proof we leave as an exercise:

Recall that the radical of an ideal $\mathfrak{A}$ in R is defined to be
$$\sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ for some } n \in \mathbb{Z}^+\}.$$

Lemma 94.1. *Let $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ be ideals in R. Then*
$$\sqrt{\bigcap_{i=1}^{n} \mathfrak{A}_i} = \bigcap_{i=1}^{n} \sqrt{\mathfrak{A}_i} = \sqrt{\mathfrak{A}_1 \cdots \mathfrak{A}_n}.$$

Recall that an ideal $\mathfrak{Q} < R$ is is called a *primary ideal* if it satisfies one of the following equivalent conditions:

(*i*) If xy lies in $\mathfrak{Q}$ then either x lies in $\mathfrak{Q}$ or there exists a positive integer n with $y^n \in \mathfrak{Q}$.

(*ii*) If xy lies in $\mathfrak{Q}$ then either x lies in $\mathfrak{Q}$ or $y \in \sqrt{\mathfrak{Q}}$.

(*iii*) The set of zero divisors $\operatorname{zd}(R/\mathfrak{A})$ of $R/\mathfrak{A}$ lies in the nilradical $\operatorname{nil}(R/\mathfrak{Q})$ of $R/\mathfrak{A}$.

Lemma 94.2. *Let $\mathfrak{Q}$ be a primary ideal in R. Then $\sqrt{\mathfrak{Q}}$ is a prime ideal in R and is the smallest prime ideal in R containing $\mathfrak{Q}$.*

PROOF. Suppose that xy lies in $\mathfrak{Q}$. Then there exists a positive integer n such that $x^n y^n$ lies in $\mathfrak{Q}$. Hence either $x^n \in \mathfrak{Q}$ or $y^{nm} \in \mathfrak{Q}$ i.e., either $x \in \sqrt{\mathfrak{Q}}$ or $y \in \sqrt{\mathfrak{Q}}$. It follows that $\sqrt{\mathfrak{Q}}$ is a prime ideal in R. As $\mathfrak{Q} \subset \sqrt{\mathfrak{Q}}$ and if $\mathfrak{Q}$ lies in a prime ideal $\mathfrak{p}$ so does $\sqrt{\mathfrak{Q}}$, the second statement also follows. $\square$

Definition 94.3. If $\mathfrak{Q}$ be a primary ideal in R and $\mathfrak{p}$ is the prime ideal $\sqrt{\mathfrak{Q}}$, we say that $\mathfrak{Q}$ is a $\mathfrak{p}$-*primary ideal.*

A useful example is given by the following result:

Lemma 94.4. *Let $\mathfrak{A} < R$ be an ideal satisfying $\sqrt{\mathfrak{A}}$ is a maximal ideal in R. Then $\mathfrak{A}$ is a $\sqrt{\mathfrak{A}}$-primary ideal. In particular, if $\mathfrak{m}$ is a maximal ideal in R, then $\mathfrak{m}^n$ is an $\mathfrak{m}$-primary ideal for each positive integer n.*

Proof. Let $\mathfrak{m} = \sqrt{\mathfrak{A}}$. Then $\mathfrak{m} = \bigcap_{V(\mathfrak{A})} \mathfrak{p}$, hence $\mathfrak{m} \subset \mathfrak{P}$ for every prime ideal in $V(\mathfrak{A})$. Since $\mathfrak{m}$ is maximal, $V(\mathfrak{A}) = \{\mathfrak{m}\}$. It follows that $R/\mathfrak{A}$ is a primary ring, i.e., $|\operatorname{Spec}(R/\mathfrak{A})| = 1$. In particular, we have $\operatorname{zd}(R/\mathfrak{A}) = \operatorname{nil}(R/\mathfrak{A})$, so $\mathfrak{A}$ is $\mathfrak{m}$-primary. $\square$

Examples 94.5. 1. Each prime ideal in R is a primary ideal in R.

2. The ideal (n) in $\mathbb{Z}$ is primary if and only if $n = 0$ or n is a power of a prime.
3. Let F be a field and x, y, z indeterminants over F. Then $(xy - z^2)$ is a prime ideal in the the UFD $F[x, y, z]$. Set $R = F[x, y, z]/(xy - z^2)$ and $^- : F[x, y, z] \to R = F[\overline{x}, \overline{y}, \overline{z}]$ the canonical epimorphism. Then $\mathfrak{p} = (\overline{z})$ is a prime ideal in the domain R. We have
$$\overline{xy} = \overline{z}^2 \in \mathfrak{p}^2 := \mathfrak{p}\mathfrak{p}, \text{ but } \overline{x}^2 \notin \mathfrak{p}^2 \text{ and } \overline{y} \notin \sqrt{\mathfrak{p}^2} = \mathfrak{p}.$$
Therefore, $\mathfrak{p}^2$ is not a primary ideal in R. So, in general, a power of a prime ideal may not be a primary ideal.
4. Let F be a field, x, y indeterminants over F and $R = F[x, y]$. Then $\mathfrak{Q} = (x, y^2)$ is an (x, y)-primary ideal in R as $R/\mathfrak{Q} = F[y]/(y^2)$ satisfies $\operatorname{zd}(R/\mathfrak{Q}) \subset \operatorname{nil}(R/\mathfrak{Q})$. Let $\mathfrak{p} = (x, y)$, then $(x^2, xy, y^2) = \mathfrak{p}^2 < \mathfrak{Q} \subset \mathfrak{p}$. It follows that $\mathfrak{Q} \neq \mathfrak{p}^n$ for any integer n, i.e., a primary ideal may not be a power of a prime ideal.
5. Suppose that $\mathfrak{Q}_i$, $i = 1, \dots, r$, are all $\mathfrak{p}$-primary ideals. Then so is $\mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_r$: We know that $\sqrt{\bigcap \mathfrak{Q}_i} = \bigcap \sqrt{\mathfrak{Q}_i} = \mathfrak{p}$. If $xy \in \bigcap \mathfrak{Q}_i$ with $y \notin \bigcap \mathfrak{Q}_i$, then there exists a j with $y \notin \mathfrak{Q}_j$. As $xy \in \mathfrak{Q}_j$, we have x^n lies in $\mathfrak{Q}_j$ for some n, hence x lies in $\sqrt{\mathfrak{Q}_j} = \mathfrak{p} = \sqrt{\bigcap \mathfrak{Q}_i}$, so $\bigcap \mathfrak{Q}_i$ is $\mathfrak{p}$-primary.

Definition 94.6. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two ideals in R. Set
$$(\mathfrak{A} : \mathfrak{B}) := \{y \in R \mid y\mathfrak{B} \subset \mathfrak{A}\} \subset R,$$
an ideal in R called the *colon ideal*.

If $\mathfrak{B} = (x)$, we write $(\mathfrak{A} : x)$ for $\big(\mathfrak{A} : (x)\big)$. If $\mathfrak{A} = 0$, then $(0 : x) = \{y \in R \mid yx = 0\}$ is the annihilator $\operatorname{ann}_R(x)$ of x in R.

Note the following:

Remarks 94.7. We have:

1. $\operatorname{zd}(R) = \bigcup_{x \neq 0} \sqrt{\operatorname{ann}_R(x)} = \bigcup_{x \neq 0} \sqrt{(0 : x)}$.
2. If $\mathfrak{A}_i \subset R$ are ideals for $i = 1, \dots, n$, then
$$(\bigcap_{i=1}^n \mathfrak{A}_i : x) = \bigcap_{i=1}^n (\mathfrak{A}_i : x).$$

Computation 94.8. Let $\mathfrak{Q}$ be a $\mathfrak{p}$-primary ideal in R and $x \in R$. Then we have:

(i) If $x \in \mathfrak{Q}$, then $(\mathfrak{Q} : x) = R$.

(ii) If $x \notin \mathfrak{Q}$, then $(\mathfrak{Q} : x)$ is $\mathfrak{p}$-primary. In particular, $\sqrt{(\mathfrak{Q} : x)} = \mathfrak{p}$.
(iii) If $x \notin \mathfrak{p}$, then $(\mathfrak{Q} : x) = \mathfrak{Q}$.

PROOF. (i) is immediate.
(iii): If $x \notin \mathfrak{p}$ and $y \in (\mathfrak{Q} : x)$, then $xy \in \mathfrak{Q}$, hence $y \in \mathfrak{Q}$ by definition.
(ii): Let $y \in (\mathfrak{Q} : x)$. As $x \notin \mathfrak{p}$ and $xy \in \mathfrak{Q}$, we have $y \in \sqrt{\mathfrak{Q}} = \mathfrak{p}$. Hence $\mathfrak{Q} \subset (\mathfrak{Q} : x) \subset \mathfrak{p}$, so

$$\mathfrak{p} = \sqrt{\mathfrak{Q}} \subset \sqrt{(\mathfrak{Q} : x)} \subset \sqrt{\mathfrak{p}} = \mathfrak{p}.$$

Next let $y, z \in R$ with $yz \in (\mathfrak{Q} : x)$ and $y \notin \sqrt{\mathfrak{Q}} = \mathfrak{p}$. Then $yzx \in \mathfrak{Q}$ implies that $xz \in \mathfrak{Q}$, hence $z \in (\mathfrak{Q} : x)$ showing that $(\mathfrak{Q} : x)$ is $\mathfrak{p}$-primary. $\square$

With the above definitions, we can now define the type of decomposition of ideals in which we shall be interested.

Definition 94.9. Let $\mathfrak{A} < R$ be an ideal. A *primary decomposition* of $\mathfrak{A}$ is an equation

$$(*) \qquad \mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \text{ with } \mathfrak{Q}_i \text{ primary for } i = 1, \ldots n.$$

We say that (*) is *irredundant* if, in addition,
(i) $\sqrt{\mathfrak{Q}_1}, \ldots, \sqrt{\mathfrak{Q}_n}$ are all distinct.
(ii) $\bigcap_{j \neq i} \mathfrak{Q}_j \not\subset \mathfrak{Q}_i$ for each $i = 1, \ldots, n$.

Remark 94.10. Suppose that an ideal $\mathfrak{A}$ has a primary decomposition (*). By Example 94.5(5), if $\mathfrak{Q}_{i_1}, \ldots, \mathfrak{Q}_{i_s}$ have the same radical, then $\mathfrak{Q}_{i_1} \cap \cdots \cap \mathfrak{Q}_{i_s}$ is $\sqrt{\mathfrak{Q}_{i_1}}$-primary, so any primary decomposition (*) can be made to satisfy (i). Throwing out those $\mathfrak{Q}_i$'s not satisfying (ii), then shows that any $\mathfrak{A}$ having a primary decomposition has an irredundant primary decomposition.

We show our first uniqueness statement for irredundant primary decompositions.

Theorem 94.11. *Suppose that $\mathfrak{A} < R$ is an ideal that has an irredundant primary decomposition*

$$\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \textit{ with } \mathfrak{p}_i = \sqrt{\mathfrak{Q}_i} \textit{ for } i = 1, \ldots, n.$$

Then

$$\{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\} = \{\sqrt{(\mathfrak{A} : x)} \mid x \in R\} \cap \operatorname{Spec}(R)$$

and is independent of the irredundant primary decomposition, i.e., the radicals of the primary ideals giving any irredundant primary decomposition of $\mathfrak{A}$ are unique.

PROOF. Let $x \in R$. We have $(\mathfrak{A} : x) = (\bigcap \mathfrak{Q}_i : x) = \bigcap (\mathfrak{Q}_i : x)$, so by the Computation 94.8,

$$\sqrt{(\mathfrak{A} : x)} = \bigcap \sqrt{(\mathfrak{Q}_i : x)} = \bigcap_{x \notin \mathfrak{Q}_i} \mathfrak{p}_i.$$

If $\sqrt{(\mathfrak{A} : x)}$ is a prime ideal, it follows that there exists a $\mathfrak{Q}_i$ with $x \notin \mathfrak{Q}_i$ such that $\sqrt{(\mathfrak{A} : x)} = \mathfrak{p}_i$ by Exercise 93.26(6).

Conversely, since our primary decomposition is irredundant, for all i and all x_i satisfying $x_i \in \bigcap_{j \neq i} \mathfrak{Q}_j \setminus \mathfrak{Q}_i$, we have $\sqrt{(\mathfrak{A} : x_i)} = \mathfrak{p}_i$ by Computation 94.8. $\square$

The proof of the above theorem and Computation 94.8 show

Corollary 94.12. *If $\mathfrak{A} < R$ is an ideal having an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$, then there exists an element x_i in $\mathfrak{Q}_i$ such that $(\mathfrak{A} : x_i)$ is $\mathfrak{p}_i$-primary.*

Definition 94.13. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition of $\mathfrak{A}$ in R with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. We set

$$\operatorname{Ass}_R V(\mathfrak{A}) := \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\},$$

the set of *associated prime ideals of* $\mathfrak{A}$ [actually of $R/\mathfrak{A}$ in modern language]. We partially order $\operatorname{Ass}_R V(\mathfrak{A})$ by set inclusion $\subset$. Elements that are minimal in $\operatorname{Ass}_R V(\mathfrak{A})$ under this partial order are called *isolated prime ideals* of $\mathfrak{A}$. The others (if any) are called *embedded prime ideals* of $\mathfrak{A}$.

Isolated prime ideals in $\operatorname{Ass}_R V(\mathfrak{A})$ are, in fact, precisely the prime ideals minimally containing $\mathfrak{A}$ in R, which we shall show next. In the sequel we shall write $\mathfrak{p} \in V(\mathfrak{A})$ is minimal for such a prime ideal.

Proposition 94.14. *Suppose that the ideal $\mathfrak{A}$ has an irredundant primary decomposition. Then a prime ideal $\mathfrak{p}$ in R is an isolated prime of $\mathfrak{A}$ if and only if $\mathfrak{p} \in V(\mathfrak{A})$ is minimal if and only if $\mathfrak{p}/\mathfrak{A} \in \operatorname{Min}(R/\mathfrak{A})$.*

Proof. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition of $\mathfrak{A}$ in R with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. Suppose that $\mathfrak{p}$ lies in $V(\mathfrak{A})$. Then

$$\mathfrak{p} = \sqrt{\mathfrak{p}} \supset \sqrt{\mathfrak{A}} = \bigcap \sqrt{\mathfrak{Q}_i} = \bigcap \mathfrak{p}_i.$$

Hence there exists an i such that $\mathfrak{p}_i \subset \mathfrak{p}$, i.e., $\mathfrak{p}$ contains an isolated prime of $\mathfrak{A}$. □

Because of the proposition, isolated primes are also called *minimal primes* of $\mathfrak{A}$.

Proposition 94.15. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. Then*

$$\bigcup \mathfrak{p}_i = \{x \in R \mid (\mathfrak{A} : x) > \mathfrak{A}\}.$$

In particular, if $\mathfrak{A} = (0)$, then

$$\operatorname{zd}(R) = \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}.$$

Proof. Let $^{-} : R \to R/\mathfrak{A}$ be the canonical epimorphism.

Check. $(\overline{0}) = \overline{\mathfrak{Q}_1} \cap \cdots \cap \overline{\mathfrak{Q}_n}$ is an irredundant primary decomposition in $\overline{R}$. So it suffices to prove $\operatorname{zd}(R) = \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}$.

By the proof of Theorem 94.11, each $\mathfrak{p}_j = \sqrt{(0 : x)}$ for some $x \in R$. The result follows. □

So we now know if (0) has an irredundant primary decomposition, then we have

$$\begin{aligned}\operatorname{zd}(R) &= \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}\\ \operatorname{nil}(R) &= \bigcap_{\operatorname{Spec}(R)} \mathfrak{p} = \bigcap_{\operatorname{Min}(R)} \mathfrak{p}\\ \operatorname{Min}(R) &\subset \operatorname{Ass}_R V(0).\end{aligned}$$

Remark 94.16. If they exist, irredundant primary decompositions are not necessarily unique. An example is given as follows: Let $R = F[x, y]$ with F a field and x, y indeterminants. Then

$$(x) \cap (x, y)^2 = (x^2, xy) = (x) \cap (x^2, y)$$

are two irredundant primary decompositions for (x^2, xy).

We shall obtain a 'weaker' uniqueness statement. We leave the proof of the following as an exercise.

Lemma 94.17. *Primary ideals satisfy:*

(1) *Let S be a multiplicative set in R and $\varphi_R : R \to S^{-1}R$ the canonical ring homomorphism given by $r \mapsto r/1$. Then φ_R induces a bijection:*
$$\{\mathfrak{Q} \mid \mathfrak{Q} < R \text{ primary with } \mathfrak{Q} \cap S = \emptyset\} \longrightarrow \{\mathfrak{Q} \mid \mathfrak{Q} < S^{-1}R \text{ primary}\}$$
via $\mathfrak{Q} \mapsto S^{-1}\mathfrak{Q}$.

(2) *If $\psi : A \to B$ is a ring homomorphism of commutative rings, $\mathfrak{Q} < B$ a primary ideal, then $\psi^{-1}(\mathfrak{Q}) \subset A$ is a primary ideal.*

Proposition 94.18. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \dots, n$ and S is a multiplicative set in R satisfying $S \cap \mathfrak{Q}_i = \emptyset$ for $1 \le i \le m$, and $S \cap \mathfrak{Q}_i \ne \emptyset$ for $i > m$. Then*

$$S^{-1}\mathfrak{A} = S^{-1}\mathfrak{Q}_1 \cap \cdots \cap S^{-1}\mathfrak{Q}_m$$

is an irredundant primary decomposition for $S^{-1}\mathfrak{A}$.

Proof. We have

$$S^{-1}\mathfrak{A} = S^{-1}(\bigcap_{i=1}^{n} \mathfrak{Q}_i) = \bigcap_{i=1}^{n} S^{-1}\mathfrak{Q}_i = \bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i$$

with $S^{-1}\mathfrak{Q}_i$ being $S^{-1}\mathfrak{p}_i$-primary, $i = 1, \dots, m$ and the $S^{-1}\mathfrak{p}_1, \dots, S^{-1}\mathfrak{p}_m$ distinct. It follows that $S^{-1}\mathfrak{A} = \bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i$ is an irredundant primary decomposition. $\square$

Remark 94.19. In the above, if $\varphi_R : R \to S^{-1}R$ is the canonical ring homomorphism $r \mapsto r/1$, then

$$(\varphi_R)^{-1}(S^{-1}\mathfrak{A}) = (\varphi_R)^{-1}(\bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i) = \bigcap_{i=1}^{m}(\varphi_R)^{-1}(S^{-1}\mathfrak{Q}_i) = \bigcap_{i=1}^{m} \mathfrak{Q}_i.$$

Definition 94.20. Let $\mathfrak{A} < R$ be an ideal with an irredundant primary decomposition. A subset $I \subset \operatorname{Ass}_R V(\mathfrak{A})$ is called *isolated* if whenever $\mathfrak{p}, \mathfrak{p}'$ lie in $\operatorname{Ass}_R V(\mathfrak{A})$ and satisfy $\mathfrak{p} \subset \mathfrak{p}'$, then $\mathfrak{p}' \in I$ implies $\mathfrak{p} \in I$, i.e., all descending chains of prime ideals lying in $\operatorname{Ass}_R V(\mathfrak{A})$ beginning with an element of I have all elements in the chain lying in I.

Remark 94.21. Let $I \subset \operatorname{Ass}_R V(\mathfrak{A})$ be an isolated set where $\mathfrak{A}$ has an irredundant primary decomposition. Let $S = R \setminus \bigcup_I \mathfrak{p}$. Then S is a saturated multiplicative set. Moreover, if $\mathfrak{p}' \in \operatorname{Ass}_R V(\mathfrak{A})$, then we have:

(i) If $\mathfrak{p}' \in I$, then $\mathfrak{p}' \cap S = \emptyset$.

(ii) If $\mathfrak{p}' \notin I$, then $\mathfrak{p}' \not\subset \bigcup_I \mathfrak{p}$ (by the Prime Avoidance Lemma 93.16), hence $\mathfrak{p}' \cap S \neq \emptyset$.

In particular, we obtain our second uniqueness result:

Theorem 94.22. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$ and $I = \{\mathfrak{p}_{i_1}, \ldots, \mathfrak{p}_{i_m}\} \subset \operatorname{Ass}_R V(\mathfrak{A})$ an isolated subset. Then $\bigcap_{j=1}^m \mathfrak{Q}_{i_j}$ is independent of an irredundant primary decomposition for $\mathfrak{A}$. In particular, the $\mathfrak{Q}_i$ with $\mathfrak{p}_i \in V(\mathfrak{A})$ minimal are uniquely determined by $\mathfrak{A}$, i.e., the isolated primes determine unique primary ideals.*

Proof. Let $S = R \setminus \bigcup_{j=1}^m \mathfrak{p}_{i_j}$. Then by the above remark,

$$\mathfrak{Q}_{i_1} \cap \cdots \cap \mathfrak{Q}_{i_m} = \varphi_R^{-1}(S^{-1}\mathfrak{A})$$

where $\varphi_R : R \to S^{-1}R$ is the canonical ring homomorphism $r \mapsto r/1$. □

Example 94.23. Let $\mathfrak{A} < R$ be a radical ideal, i.e., $\mathfrak{A} = \sqrt{\mathfrak{A}}$. Suppose that $|\operatorname{Min}(R/\mathfrak{A})|$ is finite. Then $\mathfrak{A}$ has an irredundant primary decomposition and $\operatorname{Ass}_R V(\mathfrak{A}) = \{\mathfrak{p} \in V(\mathfrak{A}) \mid \mathfrak{p} \text{ minimal}\}$, i.e., there are no embedded primes:
We have

$$\mathfrak{A} = \sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{\mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}} \mathfrak{p}$$

with the right hand side an irredundant primary decomposition of $\mathfrak{A}$.
If R is Noetherian, we have established in Corollary 92.27 and will establish again below that $|\operatorname{Min}(R/\mathfrak{A})|$ is always finite.

We wish to show that every ideal in a Noetherian ring has an irredundant primary decomposition. To do this we shall use Exercise 30.22(21) which asked to show

Lemma 94.24. *An ideal $\mathfrak{C}$ in a commutative ring R is called irreducible if whenever $\mathfrak{C} = \mathfrak{A} \cap \mathfrak{B}$ for some ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R, then either $\mathfrak{C} = \mathfrak{A}$ or $\mathfrak{C} = \mathfrak{B}$. If R is Noetherian, then every ideal $\mathfrak{A} < R$ is a finite intersection of irreducible ideals of R, i.e., $\mathfrak{A} = \mathfrak{C}_1 \cap \cdots \cap \mathfrak{C}_n$, for some irreducible ideals $\mathfrak{C}_i$ in R. We call such a decomposition an irreducible decomposition of $\mathfrak{A}$.*

This is what is needed as:

Lemma 94.25. *Let R be Noetherian ring and $\mathfrak{A} < R$ an irreducible ideal. Then $\mathfrak{A}$ is a primary ideal.*

PROOF. Since $R/\mathfrak{A}$ is also Noetherian, by the Correspondence Theorem, it suffices to show that if (0) is irreducible, then it is primary. So suppose that (0) is irreducible and $xy = 0, \ \ x, y \in R$, with $y \neq 0$. By the ascending chain condition,

$$(0 : x) \subset (0 : x^2) \subset \cdots \subset (0 : x^n) \subset \cdots$$

stabilizes, say $(0 : x^N) = (0 : x^{N+i})$ for all $i \geq 1$.

Claim. $(0) = (y) \cap (x^N)$.

If we prove the claim, the result will follow as (0) being irreducible implies $x^N = 0$, since $(y) \neq 0$. To show the claim, let $z \in (y) \cap (x^N)$. As $z \in (y)$, we have $xz = 0$ and as $z \in (x^N)$, we have $z = ax^N$ for some $a \in R$. Consequently, $0 = xz = ax^{N+1}$, so $a \in (0 : x^{N+1}) = (0 : x^N)$. Therefore, $z = ax^N = 0$ as needed. □

Putting all of what we have done together establishes:

Theorem 94.26. (Lasker-Noether Theorem) *Let R be a commutative Noetherian ring and $\mathfrak{A} < R$ an ideal. Then $\mathfrak{A}$ has an irredundant primary decomposition, say*

$$\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \ \textit{ with } \ \mathfrak{p}_i = \sqrt{\mathfrak{Q}_i} \ \textit{ for } \ i = 1, \ldots, n.$$

Moreover,

(1) *$\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are unique.*
(2) *$\mathfrak{Q}_i$ is unique if $\mathfrak{p}_i$ is isolated.*

In particular, $|\operatorname{Min}(R/\mathfrak{A})| \leq |\operatorname{Ass}_R \mathfrak{A}|$ is finite.

We derive some consequences.

Corollary 94.27. *Let R be a Noetherian ring. Then $\operatorname{Min}(R)$ is finite.*

Lemma 94.28. *Let R be a Noetherian ring and $\mathfrak{A} < R$ an ideal. Then there exists a positive integer n satisfying $(\sqrt{\mathfrak{A}})^n \subset \mathfrak{A}$.*

PROOF. Since ideals in a Noetherian ring are finitely generated, $\mathfrak{A} = (a_1, \ldots, a_n)$ for some $a_i \in \sqrt{\mathfrak{A}}$. It follows that there exists a positive integer N such that $a_i^N \in \mathfrak{A}$ for all i, so $(\sqrt{\mathfrak{A}})^N \subset \mathfrak{A}$. □

Corollary 94.29. *Let R be a Noetherian ring. Then $\operatorname{nil}(R)$ is nilpotent, i.e., there exists a positive integer N such that $\big(\operatorname{nil}(R)\big)^N = 0$.*

Corollary 94.30. *Let R be a Noetherian ring, $\mathfrak{m} \in \operatorname{Max}(R)$, and $\mathfrak{Q} < R$ an ideal. Then the following are equivalent:*

(1) *$\mathfrak{Q}$ is $\mathfrak{m}$-primary.*
(2) *$\sqrt{\mathfrak{Q}} = \mathfrak{m}$.*
(3) *There exists a positive integer n such that $\mathfrak{m}^n \subset \mathfrak{Q} \subset \mathfrak{m}$.*

PROOF. We have shown all but the implication (3) $\Rightarrow$ (2) which follows from $\mathfrak{m} = \sqrt{\mathfrak{m}^n} = \sqrt{\mathfrak{Q}} = \sqrt{\mathfrak{m}} = \mathfrak{m}$. □

Proposition 94.31. *Let R be a Noetherian ring and $\mathfrak{A} < R$ an ideal. Then*

$$\operatorname{Ass}_R V(\mathfrak{A}) = \{(\mathfrak{A} : x) \mid x \in R\} \cap \operatorname{Spec}(R)$$

Proof. Replacing R with $R/\mathfrak{A}$, we may assume that $\mathfrak{A} = 0$. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \dots, n$. As the decomposition is irredundant, $\mathfrak{A}_i = \bigcap_{j \neq i} \mathfrak{Q}_j > 0$. By the proof of Theorem 94.11, $\mathfrak{p}_i = \sqrt{(0 : x)}$ for all nonzero x in $\mathfrak{A}_i$, since $\mathfrak{A}_i \cap \mathfrak{Q}_i = 0$. Hence

$$(0 : x) = \operatorname{ann}_R x \subset \mathfrak{p}_i \text{ for all nonzero } x \in \mathfrak{A}_i.$$

As $\mathfrak{Q}_i$ is $\mathfrak{p}_i$-primary, there exists a positive integer m satisfying $\mathfrak{p}_i^m \subset \mathfrak{Q}_i$ by Lemma 94.28. Consequently, $\mathfrak{A}_i\mathfrak{p}_i^m \subset \mathfrak{A}_i \cap \mathfrak{p}_i^m \subset \mathfrak{A}_i \cap \mathfrak{Q}_i = 0$.

Choose $m > 0$ minimal with $\mathfrak{A}_i \cap \mathfrak{p}_i^m = 0$. If $x \in \mathfrak{A}_i\mathfrak{p}_i^{m-1} \subset \mathfrak{A}_i$, then $x\mathfrak{p}_i = 0$, so we conclude that $\mathfrak{p}_i \subset \operatorname{ann}_R x = (0 : x)$ if x is nonzero.

Conversely, if $\mathfrak{p} = (0 : x)$ is a prime ideal, then $\sqrt{(0 : x)} = \sqrt{\operatorname{ann}_R x} = \mathfrak{p}$, and $\mathfrak{p} \in \operatorname{Ass}_R V(0)$. □

Exercises 94.32.

1. Prove Lemma 94.1
2. Prove Lemma 94.4
3. Show that the colon ideal $(\mathfrak{A} : \mathfrak{B})$ of the ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R is an ideal in R and the largest ideal $\mathfrak{C}$ in R satisfying $\mathfrak{B}\mathfrak{C} \subset \mathfrak{A}$.
4. Prove Remark 94.7
5. Prove Lemma 94.17

95. Addendum: Associated Primes of Modules

In the Section §94, we gave the best generalization of the Fundamental Theorem of Algebra, thereby concluding this part of our investigations into analogues of unique factorization domains. Some of the constructions and ideas in that section have become more important than primary decomposition itself. Many are easier to develop than in that section, so we generalize the approach and get alternative proofs of some of the results in Section 94. In particular, if $\mathfrak{A}$ is an ideal in a commutative ring R, then the set of associated primes of $\mathfrak{A}$ were denoted $\operatorname{Ass}_R V(\mathfrak{A})$. In this section, this will become $\operatorname{Ass}_R(R/\mathfrak{A})$ as we shall define the associated primes of an R-module. Throughout this section R is a commutative ring.

Definition 95.1. Let R be a commutative ring and M an R-module. A prime ideal $\mathfrak{p}$ in R is called an *associated prime of M if there exists a nonzero element m in M such that* $\mathfrak{p} = \operatorname{ann}_R m$. *Let*

$$\operatorname{Ass}_R(M) := \{\mathfrak{p} \in \operatorname{Spec} R \mid \mathfrak{p} \textit{ is an associated prime of } M\}.$$

Remarks 95.2. Let M be an R-module and $\mathfrak{p}$ a prime ideal in R.

1. The prime ideal $\mathfrak{p}$ lies in $\operatorname{Ass}_R(M)$ if and only if there exists an R-monomorphism $R/\mathfrak{p} \hookrightarrow M$, since $R/\operatorname{ann}_R m \cong Rm$ for all m in M.
2. $R/\mathfrak{p}$ is a domain, so $\mathfrak{p} = \operatorname{ann}_R(r + \mathfrak{p})$ for all $r \in R \setminus \mathfrak{p}$, hence

$$\operatorname{Ass}_R(R/\mathfrak{p}) = \{\mathfrak{p}\}.$$

Lemma 95.3. *Let R be a commutative ring and*

$$0 \to M' \to M \to M'' \to 0$$

be an exact sequence of R-modules. Then

$$\mathrm{Ass}_R(M') \subset \mathrm{Ass}_R(M) \subset \mathrm{Ass}_R(M') \cup \mathrm{Ass}_R(M'').$$

PROOF. In the exact sequence, we may assume that $M' \subset M$, $M'' = M/M'$, and $^- : M \to M/M'$ is the canonical R-homomorphism.

If $\mathfrak{p} \in \mathrm{Ass}_R(M')$, there exists an R-monomorphism $R/\mathfrak{p} \hookrightarrow M' \subset M$, so $\mathfrak{p} \in \mathrm{Ass}_R(M)$.

Suppose that $\mathfrak{p} \in \mathrm{Ass}_R(M)$. Let N be the image of the R-monomorphism $R/\mathfrak{p} \hookrightarrow M$. If $N \cap M' = 0$, then the restriction of $^- : M \to M/M'$ to N gives an R-monomorphism $N \hookrightarrow M/M'$, so $\mathfrak{p} \in \mathrm{Ass}_R(M/M')$. If $0 \neq N \cap M' \subset N$, since $\mathfrak{p} = \mathrm{ann}_R\, n$ for all nonzero n in N, we have $\mathfrak{p} \in \mathrm{Ass}_R(N \cap M') \subset \mathrm{Ass}_R(M')$. □

We shall see that the second set inclusion in the lemma can be proper. [The first set inclusion can easily seen to be proper in general.]

The generalization of Proposition 94.31 is easy to prove.

Proposition 95.4. *Let R be a commutative ring and M be a nonzero R-module. Suppose that*

$$\mathfrak{A} \in \{\mathrm{ann}_R\, m \mid 0 \neq m \in M\} \quad \textit{is maximal relative to } \subset.$$

Then $\mathfrak{A} \in \mathrm{Ass}_R(M)$. In particular if R is Noetherian and N is an R-module, then

$$\mathrm{Ass}_R(N) = \emptyset \quad \textit{if and only if} \quad N = 0.j$$

PROOF. Suppose that there exist a and b in R with $b \notin \mathfrak{A}$ but $ab \in \mathfrak{A}$. By definition $\mathfrak{A} = \mathrm{ann}_R\, m$ for some nonzero element m in M. As $bm \neq 0$, we have $\mathrm{ann}_R\, m \subset \mathrm{ann}_R\, bm < R$. By maximality, $\mathrm{ann}_R\, m = \mathrm{ann}_R\, bm$, so $abm = 0$ implies that $a \in \mathrm{ann}_R\, bm \subset \mathrm{ann}_R\, m = \mathfrak{A}$. Therefore, $\mathfrak{A} \in \mathrm{Spec}(R)$. Furthermore, if R is Noetherian and N is nonzero, then $\{\mathrm{ann}_R\, n \mid 0 \neq n \in N\}$ has a maximal element. □

Definition 95.5. Let R be a ring and M a nonzero R-module. We call $r \in R$ a *zero divisor* on M if there exists a nonzero element $m \in M$ such that $rm = 0$. We let $\mathrm{zd}(M)$ denote the set of zero divisors on M.

Using the proposition, we get a simple proof to the generalization of the conclusion of Proposition 94.15.

Corollary 95.6. *Let R be a Noetherian ring and M a nonzero R-module. Then $\mathrm{zd}(M) = \bigcup_{\mathrm{Ass}_R(M)} \mathfrak{p}$.*

Proposition 95.7. *Let R be a Noetherian ring and M a nonzero finitely generated R-module. Then there exists a chain of R-modules*

$$0 = M_0 < M_1 < \cdots < M_n = M$$

for some n satisfying

$$M_i/M_{i-1} \cong R/\mathfrak{p}_i \quad \textit{for some prime ideals } \mathfrak{p}_i,\ i = 1, \ldots, n.$$

PROOF. Let

$$\mathcal{S} = \{N \mid 0 < N \subset M \ \ R\text{-modules with } N \text{ satisfying the Proposition}\}.$$

Since $M \neq 0$, there exists a prime ideal $\mathfrak{p} \in \operatorname{Ass}_R(M)$ and an R-monomorphism $\varphi : R/\mathfrak{p} \hookrightarrow M$. Therefore, $\operatorname{im}\varphi$ lies in $\mathcal{S}$, i.e., $\mathcal{S}$ is nonempty. As R is Noetherian and M is finitely generated, M is a a Noetherian R-module. Hence there there exists a maximal element $N \in \mathcal{S}$ (relative to $\subset$). Suppose that $N < M$. Then there exists a prime ideal in $\operatorname{Ass}_R(M/N)$. Consequently, there exists an R-module N' satisfying $N < N' \subset M$ and an R-isomorphism $R/\mathfrak{p} \to N'/N$. As $N \in \mathcal{S}$ and $N'/N \cong R/\mathfrak{p}$, we have $N' \in \mathcal{S}$ by Lemma 95.3, contradicting the maximality of N. Therefore, $N = M$. □

Corollary 95.8. *Let R be a Noetherian ring and M a nonzero finitely generated R-module. Then $\operatorname{Ass}_R(M)$ is a finite set.*

PROOF. By Proposition 95.7, there exists a chain of R-modules $0 = M_0 < M_1 < \cdots < M_n = M$ satisfying $M_i/M_{i-1} \cong R/\mathfrak{p}_i$ with $\mathfrak{p}_i$ a prime ideal and $\operatorname{Ass}_R(M_i/M_{i-1}) = \{\mathfrak{p}_i\}$ for $i = 1, \ldots, n$. By Lemma 95.3, we have $\operatorname{Ass}_R(M_i) \subset \operatorname{Ass}_R(M_{i-1}) \cup \operatorname{Ass}_R(M_i/M_{i-1})$, it follows that

$$(*) \qquad \begin{aligned} \operatorname{Ass}_R(M) &\subset \operatorname{Ass}_R(M_1) \cup \operatorname{Ass}_R(M_2/M_1) \cup \cdots \cup \operatorname{Ass}_R(M_n/M_{n-1}) \\ &= \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\}. \end{aligned}$$

□

Remark 95.9. In (*) in the above proof, it is possible that

$$\operatorname{Ass}_R M < \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\}.$$

For example, let $M = R$ be a Noetherian domain. Then $\operatorname{Ass}_R(R) = \{0\}$. Let $\mathfrak{p}$ be a nonzero prime ideal in R. Then $0 < \mathfrak{p} < R$ is a sequence of R-modules. If $\mathfrak{p}$ is a principal prime ideal in R, then $\mathfrak{p}$ is R-free. In particular, $\mathfrak{p}/(0) = \mathfrak{p} \cong R$ is R-free, giving such an example, e.g., if R is a UFD (not a field).

Kaplansky believed the following corollary is one of the most useful facts.

Corollary 95.10. *Let R be a Noetherian ring, M a nonzero finitely generated R- module, and S a subrng of R in $\operatorname{zd}(M)$. Then there exists an associated prime $\mathfrak{p}$ of M satisfying $S \subset \mathfrak{p}$. In particular, there exists a nonzero element m in M such that $Sm = 0$.*

PROOF. We have $S \subset \operatorname{zd}(M) = \bigcup_{\operatorname{Ass}_R(M)} \mathfrak{p}$ and $|\operatorname{Ass}_R(M)| < \infty$, so there exists a prime ideal $\mathfrak{p} \in \operatorname{Ass}_R(M)$ with $S \subset \mathfrak{p}$ by the Prime Avoidance Lemma 93.16. If $\mathfrak{p} = \operatorname{ann}_R m$, then $Sm = 0$. □

Definition 95.11. Let M be a nonzero R-module and $\mathfrak{p}$ an associated prime of M. We say that $\mathfrak{p}$ is a *minimal prime* or *isolated prime* if $\mathfrak{p}$ is minimal in $\operatorname{Ass}_R(M)$ and an *embedded prime* otherwise.

We shall characterize minimal associated primes of a finitely generated module over a Noetherian ring below. Embedded associated primes are more mysterious. We identify one type of embedded associated primes (when they exist) in the following:

Corollary 95.12. *Let R be a Noetherian ring and M a nonzero finitely generated R-module. Then the following sets are identical:*

(1) $\{\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Ass}_R(M)$ *is maximal*$\}$.
(2) $\{\mathfrak{p} \in \operatorname{Spec}(R) \mid \mathfrak{p} \subset \operatorname{zd}(M)$ *is maximal*$\}$.
(3) $\{\mathfrak{A} < R$ *an ideal* $\mid \mathfrak{A} \subset \operatorname{zd}(M)$ *is maximal*$\}$.

Proof. If $\mathfrak{A} \subset \operatorname{zd}(M)$ is a ideal, then there exists a prime ideal $\mathfrak{p} \in \operatorname{Ass}_R(M)$ satisfying $\mathfrak{A} \subset \mathfrak{p}$ by Lemma 95.10. The result now follows easily. □

We need three lemmas to characterize minimal elements in $\operatorname{Ass}_R(M)$ when M is a finitely generated module over a Noetherian ring. We leave a proof of the first lemma as an exercise.

Lemma 95.13. *Let M be a finitely generated R module and S a multiplicative set in R. Then* $\operatorname{ann}_{S^{-1}R}(S^{-1}M) = S^{-1}\big(\operatorname{ann}_R(M)\big)$.

Lemma 95.14. *Let M be a finitely generated R module over a Noetherian ring R and S a multiplicative set in R. Then*

$$\operatorname{Ass}_{S^{-1}R}(S^{-1}M) = \{S^{-1}\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Ass}_R M \ \textit{ with } \ \mathfrak{p} \cap S = \emptyset\}.$$

Proof. ($\supset$): We know that

$$\operatorname{Spec}\big(S^{-1}(R)\big) = \{S^{-1}\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Spec}_R(M),\, \mathfrak{p} \cap S = \emptyset\}$$

and that $\operatorname{ann}_{S^{-1}R}\big(\frac{m}{1}\big) = S^{-1}(\operatorname{ann}_R m)$ for all m in M by Lemma 95.13, so $\supset$ follows.

($\subset$): Let $\mathfrak{P} \in \operatorname{Ass}_{S^{-1}R}(S^{-1}M)$, say $\mathfrak{P} = \operatorname{ann}_{S^{-1}R}\big(\frac{m}{s}\big)$ for some $m \in M$ and $s \in S$. Let $\mathfrak{p} \in \operatorname{Spec}(R)$ satisfy $S^{-1}\mathfrak{p} = \mathfrak{P}$ with $\mathfrak{p} \cap S = \emptyset$.

Claim. $\mathfrak{p} \in \operatorname{Ass}_R(M)$ (hence we are done).

Since R is Noetherian, $\mathfrak{p}$ is finitely generated, say $\mathfrak{p} = \big(a_1, \dots, a_n\big)$. By assumption,

$$\frac{a_i}{1}\frac{m}{s} = 0 \ \text{ in } \ S^{-1}M, \ \text{ so there exist } \ s_i \in S \ \text{ with } \ s_i a_i m_i = 0$$

for $i = 1, \dots, n$. Let $\widetilde{s} = s_1 \cdots s_n$, so $\widetilde{s} a_i m = 0 = a_i \widetilde{s} m$ for $i = 1, \dots, n$. Therefore, $\mathfrak{p} = (a_1, \dots, a_n) \subset \operatorname{ann}_R(\widetilde{s} m)$. Let $\varphi : R \to S^{-1}R$ be the natural ring homomorphism given by $r \mapsto \frac{r}{1}$. Since $\frac{s}{1}$ and $\frac{\widetilde{s}}{1}$ lie in $(S^{-1}R)^\times$, we have $\operatorname{ann}_{S^{-1}R}\big(\frac{\widetilde{s}\, m}{1}\big) = \operatorname{ann}_{S^{-1}R}\big(\frac{m}{s}\big)$. Using Lemma 95.13, we see that

$$\begin{aligned}\varphi^{-1}\big(S^{-1}(\operatorname{ann}_R \widetilde{s}m)\big) &= \varphi^{-1}\big(\operatorname{ann}_{S^{-1}R} \frac{\widetilde{s}m}{1}\big) \\ &= \varphi^{-1}\big(\operatorname{ann}_{S^{-1}R} \frac{m}{s}\big) = \varphi^{-1}(\mathfrak{P}) = \mathfrak{p}.\end{aligned}$$

Therefore, $\operatorname{ann}_R \widetilde{s}\, m \subset \varphi^{-1}\big(S^{-1}(\operatorname{ann}_R \widetilde{s}\, m)\big) \subset \mathfrak{p}$. Consequently, $\mathfrak{p} = \operatorname{ann}_R \widetilde{s}\, m$. □

Definition 95.15. Let M be an R-module. Then the *support of* M is defined to be the set $\operatorname{Supp}(M) := \{\mathfrak{p} \in \operatorname{Spec} R \mid M_\mathfrak{p} \neq 0\}$.

Lemma 95.16. *Let R be a commutative ring and M and R-module. Then*

(1) $M = 0$ *if and only if* $\operatorname{Supp}(M) = \emptyset$.

(2) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-modules, then
$$\mathrm{Supp}_R(M) = \mathrm{Supp}_R(M') \cup \mathrm{Supp}_R(M'').$$
(3) If M is finitely generated, then $\mathrm{Supp}(M) = V\big(\mathrm{ann}_R(M)\big)$.

Proof. (1) and (2) follow by Exercise 92.31(7).
(3): Let $\mathfrak{p} \in V\big(\mathrm{ann}_R(M)\big)$ and $S = R \setminus \mathfrak{p}$. Suppose that $\mathfrak{p} \notin \mathrm{Supp}(M)$, then $M_{\mathfrak{p}} = 0$; so for all $m \in M$ there exists an element $s_m \in S$ satisfying $s_m m = 0$. Since M is finitely generated, there exists an $s \in S$ satisfying $s\,M = 0$, hence $s \in \mathrm{ann}_R(M) \subset \mathfrak{p}$, a contradiction. Therefore, $V\big(\mathrm{ann}_R(M)\big) \subset \mathrm{Supp}(M)$

Suppose that $\mathfrak{p} \in \mathrm{Supp}(M)$ and $S = R \setminus \mathfrak{p}$. Then there exists an $m \in M$ satisfying $\dfrac{m}{s}$ is nonzero in $M_{\mathfrak{p}}$ for some $s \in S$. It follows that $\widetilde{s}\,m \neq 0$ in M for all $\widetilde{s} \in S$. Therefore, $S \cap \mathrm{ann}_R(m) = \emptyset$, i.e., $\mathrm{ann}_R(m) \subset \mathfrak{p}$. □

Theorem 95.17. *Let R be a commutative ring and M a nonzero R-module. Then $\mathrm{Ass}_R(M) \subset \mathrm{Supp}(M)$. If, in addition, R is Noetherian and M is a finitely generated R-module, then the following sets are identical:*
(1) $\{\mathfrak{p} \mid \mathfrak{p} \in \mathrm{Ass}_R(M)$ is minimal$\}$.
(2) $\{\mathfrak{p} \mid \mathfrak{p} \in \mathrm{Supp}_R(M)$ is minimal$\}$.
(3) $\{\mathfrak{p} \mid \mathfrak{p} \in V\big(\mathrm{ann}_R(M)\big)$ is minimal$\}$.

Proof. Let $\mathfrak{p} \in \mathrm{Ass}_R(M)$, say $\mathfrak{p} = \mathrm{ann}_R\, m$ for some nonzero m in M. Then $Rm \subset M$ and, as Rm is finitely generated, by Lemma 95.16,
$$\mathfrak{p} \in V\big(\mathrm{ann}_R(Rm)\big) = \mathrm{Supp}(Rm) \subset \mathrm{Supp}(M).$$
So $\mathrm{Ass}_R(M) \subset \mathrm{Supp}(M)$ as needed.
Now assume that R is Noetherian and M is finitely generated. Then by Lemma 95.16, the set of elements in (2) and (3) are equal. We must show that the set of elements in (1) and (2) are equal. Since $\mathrm{Ass}_R(M) \subset \mathrm{Supp}(M)$, it suffices to show:
Claim. If $\mathfrak{p} \in \mathrm{Supp}(M)$ is minimal, then $\mathfrak{p} \in \mathrm{Ass}_R(M)$.
By Lemma 95.14, it suffices to show that $S^{-1}\mathfrak{p} \in \mathrm{Ass}_{S^{-1}R}\, M_{\mathfrak{p}}$. As $\mathfrak{p} \in \mathrm{Supp}(M)$ is minimal, clearly $S^{-1}\mathfrak{p} \in \mathrm{Supp}_{S^{-1}R}(M_{\mathfrak{p}})$ is also minimal. Therefore, we may assume that R is a local ring with maximal ideal of $\mathfrak{p}$. Then $M = M_{\mathfrak{p}}$ is nonzero and $\mathfrak{p}$ in $\mathrm{Supp}(M)$ is minimal, we must have $\mathrm{Supp}(M) = \{\mathfrak{p}\}$. Since R is Noetherian, we have $\emptyset \neq \mathrm{Ass}_R(M) \subset \mathrm{Supp}(M) = \{\mathfrak{p}\}$ by Proposition 95.4, so $\mathrm{Ass}_R((M) = \{\mathfrak{p}\}$ as needed. □

Corollary 95.18. *Let R be a Noetherian ring, $\mathfrak{A} < R$ and ideal. Then $\mathfrak{p} \in \mathrm{Ass}_R(R/\mathfrak{A})$ is minimal if and only if $\mathfrak{p} \in V(\mathfrak{A})$ is minimal. In particular, $\mathrm{Min}(R) \subset \mathrm{Ass}_R(R)$ and are the minimal elements of $\mathrm{Ass}_R(R)$.*

Proof. This is immediate as $\mathrm{ann}_R(R/\mathfrak{A}) = \mathfrak{A}$ □

Note that this gives another proof that $\mathrm{Min}(R)$ is finite if R is Noetherian. (Cf. Corollary 92.27.)

Corollary 95.19. *Let R be a Noetherian ring and M a finitely generated R-module. Then $\mathrm{Supp}(M) = \bigcup_{\mathrm{Ass}_R(M)} V(\mathfrak{p})$.*

PROOF. Minimal elements in $\mathrm{Supp}(M)$ and $\mathrm{Ass}_R(M)$ coincide. □

Remark 95.20. The Primary Decomposition Theorem when $M = R$ is a Noetherian ring and $N = \mathfrak{A} < R$ an ideal can now be generalized as follows — we leave the verification of the details to the reader: Let R be a commutative ring, M a finitely generated R-module, and $N < M$ a submodule. If $\mathfrak{p} \in \mathrm{Spec}(R)$, we say N is $\mathfrak{p}$-primary if $\mathrm{Ass}_R(M/N) = \{\mathfrak{p}\}$. If $\mathfrak{p} \in \mathrm{Spec}(R)$ and M is an R-module that is the intersection of $\mathfrak{p}$-primary submodules of M, then M is also $\mathfrak{p}$-primary. A decomposition of N as an intersection

$$(*) \qquad N = Q_1 \cap \cdots \cap Q_n$$

with each $Q_i \subset M$ a $\mathfrak{p}_i$-primary submodule, $i = 1, \ldots, n$, is called a *primary decomposition* of N in M. Such a primary decomposition is called *irredundant* if no Q_i can be omitted and $\mathfrak{p}_i \neq \mathfrak{p}_j$ for $i, j = 1, \ldots, n$.

The Primary Decompositon Theorem generalize to: If R is a Noetherian ring, M a finitely generated R-module and $N < M$ a submodule, then N has an irredundant primary decomposition in M. Moreover, $\mathrm{Ass}_R(M/N) = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\}$, with the $\mathfrak{p}_i$-primary submodule Q_i unique if $\mathfrak{p}_i \in \mathrm{Ass}_R(M/N)$ is minimal, i.e., $\mathfrak{p}_i \in \big(V(\mathrm{ann}_r(M/N)\big)$ is minimal.

Exercises 95.21.

1. Show if $\mathfrak{A}$ is an ideal in R, then $\mathrm{Ass}_R\big(V(\mathfrak{A})\big) = \mathrm{Ass}_R(R/\mathfrak{A})$.
2. Prove Lemma 95.13.
3. Prove the assertions in Remark 95.20.
4. Let R be a Noetherian domain and M be a finitely generated free R-module. Show that $\mathrm{Ass}_R(M) = \{0\}$ but if M is R-torsion-free but not R-free, then the set of primes ideals appearing in the filtration of M Proposition 95.9 for M is $\mathrm{Ass}_R(M)$.

96. Akizuki and Krull-Akizuki Theorems

As is true in this chapter, all rings are commutative, although straight-forward modifications of the beginning of this chapter can be shown to be true if R is arbitrary.

Recall that a commutative ring is called *Artinian ring* if it satisfies the *descending chain condition*. Similarly, a left R-module is called *Artinian* if submodules of it satisfies the descending chain condition. We wish first with to prove a theorem of Akizuki that characterizes Artinian commutative rings as those that are Noetherian of dimension zero. We have previously noted that the analogue of the Jordan-Hölder Theorem for vector spaces holds and used it to show that dimension was a well defined invariant for finite dimensional vector spaces. We expand on these observations.

Definition 96.1. Let R be a commutative ring. An R-module M is said to have a *composition series* if the exists a finite filtration of submodules of M

$$(*) \qquad M = M_0 > M_1 > \cdots > M_n = 0$$

with each quotient M_i/M_{i+1} *irreducible* (or *simple*), i.e., has no proper submodules. We say that a module having a composition series has *finite length* . The composition series in (*) is called a series of *length* n, i.e., the number of *links* in (*).

Just as in the group theoretical case, we have:

Theorem 96.2. (Jordan-Hölder Theorem) *Suppose that an R-module M has a composition of length n. Then every composition series of M is of length n. Moreover, if M has a composition series, any proper chain (i.e, all links proper) of M can be refined to a composition series with unique quotients up to order an isomorphism.*

We omit the proof which is similar, but easier than the group theoretic case.

If M is an R-module having a composition series, the common length of a composition series for M is called the *length* of R and denoted by $l(M)$. We set the length of the zero module as zero. If a module M has no composition series, for convenience, we let $l(M) = \infty$ (with $\infty + n = \infty$ for any integer n). An immediate consequence is:

Corollary 96.3. *Let M be an R-module having a composition series and N a submodule of M, Then $l(N) \leq l(M)$ with equality if and only if $N = M$.*

More generally,

Corollary 96.4. *Let*

$$0 \to M' \to M \to M'' \to 0$$

be an exact sequence of R-modules. Then l is additive, i.e.,

$$l(M) = l(M') + l(M''),$$

whose proof we leave as an exercise. For a module to have finite length is a very strong condition. Indeed:

Lemma 96.5. *Let M be an R-module. Then M has finite length if and only if it is both Noetherian and Artinian.*

Proof. If M has finite length, all proper chains of submodules of M have length bounded by the length of M by the Jordan-Hölder Theorem. Conversely, suppose that M is both Noetherian and Artinian. Since it is Noetherian, there exists a maximal proper submodule $M_1 < M$, i.e., M/M_1 is irreducible by the Correspondence Principle. Since M is Noetherian so is M_1. Continuing gives a descending chain of modules $M = M_0 > M_1 > M_2 > \cdots$ with M_i/M_{i+1} irreducible. This must stop, since M is also Artinian. □

An easy consequence of this (that is left as an exercise) is:

Proposition 96.6. *Let V be a vector space over a field K. Then the following are equivalent:*

(1) V is finite dimensional.
(2) V has finite length.
(3) V is Noetherian.
(4) V is Artinian.

Moreover, the dimension of V is just its length.

Corollary 96.7. *Suppose $\mathfrak{m}_1, ..., \mathfrak{m}_n$ are maximal ideals in R (not necessarily distinct) and further that $\mathfrak{m}_1 \cdots \mathfrak{m}_n = 0$. Then R is Noetherian if and only if R is Artinian.*

Proof. Each $\mathfrak{m}_1 \cdots \mathfrak{m}_i/\mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ is an $(R/\mathfrak{m}_{i+1})$-vector space so is Artinian if and only if it is Noetherian. We know that $\mathfrak{m}_1 \cdots \mathfrak{m}_i$ is Noetherian (resp., Artinian) if and only if both of the ideals $\mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ and $\mathfrak{m}_1 \cdots \mathfrak{m}_i/\mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ are. [Cf. Proposition 40.5 and Exercise 40.12(4).] The result now follows, since

$$R \supset \mathfrak{m}_1 \supset \mathfrak{m}_1\mathfrak{m}_2 \supset \cdots \supset \mathfrak{m}_1 \cdots \mathfrak{m}_n = 0. \qquad \square$$

Proposition 96.8. *Let R be a nonzero commutative Artinian ring. Then R is semi-local of dimension zero. In particular,* $\mathrm{rad}(R) = \mathrm{nil}(R)$.

Proof. Let $\mathfrak{p}$ be a prime ideal of R. Then $R/\mathfrak{p}$ is an Artinian domain. Thus to show that $\dim R$ is zero, i.e., $\mathrm{Spec}(R) = \mathrm{Max}(R)$, we need only show that any Artinian domain is a field. But if R is an Artinian domain and x nonzero in R, then the descending chain

$$Rx \supset Rx^2 \supset Rx^3 \supset \cdots$$

must stabilize. Hence $Rx^n = Rx^{n+1}$ for some n. In particular, $yx^{n+1} = x^n$ for some y in the domain R. Thus $yx = 1$ and R is a field.

Now suppose that R is an arbitrary Artinian ring. By the Minimal Principle, there exists a minimal element in the set of finite intersections of maximal ideals,

$$\{\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \mid \mathfrak{m}_1, \ldots, \mathfrak{m}_m \in \mathrm{Max}(R),\ \text{some } m\}.$$

Let $\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n$ be such a minimal element. If $\mathfrak{m}$ is a maximal ideal, then by minimality

$$\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n = \mathfrak{m} \cap \mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \subset \mathfrak{m}.$$

It follows that $\mathfrak{m} = \mathfrak{m}_i$, for some i, hence $\mathrm{Max}(R) = \{\mathfrak{m}_1, \ldots, \mathfrak{m}_n\}$. $\square$

Theorem 96.9. (Akizuki) *Let R be a nonzero commutative ring. Then the following are equivalent.*

(1) R is Artinian.
(2) R is Noetherian of dimension zero.
(3) Every finitely generated R-module has finite length.

Proof. Every module over an Artinian (resp., Noetherian) ring R is Artinian (resp., Noetherian), since it is a quotient of R^n for some n. Thus by the above, we know that R satisfies both (1) and (2) if and only if R satisfies (3). So we need only show that R satisfies (1) if and only if R satisfies (2).

Suppose that (1) holds, i.e., that R is Artinian. Then R is semi-local of dimension zero. Let $\mathrm{Max}(R) = \{\mathfrak{m}_1, ..., \mathfrak{m}_n\}$. By the corollary above, it suffices to show that $0 = \prod_{1=1}^{n} \mathfrak{m}_i^k$ for some k. But $\prod_{i=1}^{n} \mathfrak{m}_i \subset \bigcap_{1=1}^{n} \mathfrak{m}_i = \mathrm{rad}(R) = \mathrm{nil}(R)$ by the proposition. So it suffices to prove the following:

Claim 96.10. *If R is Artinian, then* $\mathrm{nil}(R)$ *is nilpotent:*

By the descending chain condition, $\big(\mathrm{nil}(R)\big)^k = \big(\mathrm{nil}(R)\big)^{k+i}$ for some k and all i. Let $\mathfrak{A} := \big(\mathrm{nil}(R)\big)^k$. We must show that $\mathfrak{A} = 0$. Suppose not. Let

$$\mathcal{S} = \{\mathfrak{B} < R \,|\, \mathfrak{B} \text{ an ideal of } R \text{ with } \mathfrak{A}\mathfrak{B} \neq 0\}.$$

Since $\mathfrak{A} \in \mathcal{S}$ and R is Artinian, there exists an ideal $\mathfrak{B}$ that is minimal in $\mathcal{S}$. In particular, there exists an x in $\mathfrak{B}$ such that $x\mathfrak{A} \neq 0$. By minimality, $\mathfrak{B} = Rx$. Since $(x\mathfrak{A})\mathfrak{A} = x\mathfrak{A}^2 = x\mathfrak{A} \neq 0$, we also have $x\mathfrak{A} = \mathfrak{B}$ by minimality. Choose $y \in \mathfrak{A}$ such that $x = xy$. Then

$x = xy^N$ for all positive integers N. But $y \in \mathfrak{A} \subset \mathrm{nil}(R)$, so $y^N = 0$ for some N and hence $x = 0$ also. This is a contradiction. Thus $\mathfrak{A} = 0$ and the Claim is established.

Now suppose that R is Noetherian of dimension zero. By the corollary above, it suffices to show that there exist maximal ideals $\mathfrak{m}_1, ..., \mathfrak{m}_n$ in R, not necessarily distinct, so that $0 = \mathfrak{m}_1 \cdots \mathfrak{m}_n$. Since R is Noetherian the zero ideal contains a finite product of prime ideals. (Cf. Exercise 30.22(20).) Since $\dim R = 0$, these prime ideals are maximal. The result follows. □

Corollary 96.11. *Let R be a domain. Then the following are equivalent:*

(1) R is Noetherian of dimension at most one.
(2) If $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ has finite length.
(3) If $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ is Artinian.

Proof. If (1) holds and $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ is Noetherian of dimension zero so (2) holds. Clearly, (2) implies (3), so we need only show that (3) implies (1).

If (3) holds then $R/\mathfrak{A}$ is a Noetherian ring of dimension zero for any ideal $0 < \mathfrak{A} < R$. In particular, if $x \in \mathfrak{A}$ is nonzero, then $\mathfrak{A}/Rx$ is a finitely generated ideal in R/Rx. It follows that $\mathfrak{A}$ is finitely generated as an ideal in R, i.e., R is Noetherian. If $0 < \mathfrak{p}_1 \subset \mathfrak{p}_2$ is a chain of primes in R then $\mathfrak{p}_2/\mathfrak{p}_1 = 0$ in the Artinian domain, hence field, $R/\mathfrak{p}_1$. Thus R has dimension at most one. □

Lemma 96.12. *Let M be a nontrivial R-module and $\mathfrak{p}$ a prime ideal in R containing $\mathrm{ann}_R(M)$ and a minimal such prime ideal, i.e., no prime ideal containing $\mathrm{ann}_R(M)$ properly lies in $\mathfrak{p}$. Then $\mathfrak{p}$ consists of zero divisors of M. In particular,*

$$\bigcup_{\mathrm{Min}(R)} \mathfrak{p} \subset \mathrm{zd}(R).$$

[There is no Noetherian condition or finite generation condition.]

Proof. Let S be the multiplicative set in R defined by

$$S := \{ab \,|\, a \in R \setminus \mathfrak{p} \text{ and } b \in R \setminus \mathrm{zd}(M)\}.$$

Claim 96.13. $S \cap \mathrm{ann}_R(M) = \emptyset$.

Suppose not. Then there exists an a in $R \setminus \mathfrak{p}$ and b in $R \setminus \mathrm{zd}(M)$ satisfying $abM = 0$. Since b is not a zero divisor on M, we have $aM = 0$, and hence $a \in \mathrm{ann}_R(M) \subset \mathfrak{p}$, a contradiction. This establishes the Claim.

Thus there exists a prime $\mathfrak{P}$ containing $\mathrm{ann}_R(M)$ such that $\mathfrak{P}$ excludes S and is maximal with respect to this property. Since 1 lies in R but not in $\mathfrak{p}$ or $\mathrm{zd}(M)$ and $S = \big(R \setminus \mathrm{zd}(M)\big) \cdot (R \setminus \mathfrak{p})$, we have

$$S \supset \big(R \setminus \mathrm{zd}(M)\big) \text{ and } S \supset (R \setminus \mathfrak{p}).$$

Therefore,

$$\mathfrak{P} \subset \mathrm{zd}(M) \cap \mathfrak{p} \subset \mathfrak{p}.$$

The minimality condition on $\mathfrak{p}$ implies that $\mathfrak{p} = \mathfrak{P}$, so $\mathfrak{p} \subset \mathrm{zd}(M)$ as desired. For the last statement, let $M = R$. Then every prime contains $0 = \mathrm{ann}_R(R)$. It follows from the first part that if $\mathfrak{p}$ is a minimal prime ideal in R, then $\mathfrak{p}$ consists of zero divisors of R as $\operatorname{ann}_R(R) = 0$. □

We need the lemma in the following special case.

Corollary 96.14. *If* $\dim R = 0$ *then* $R^\times = R \setminus \operatorname{zd}(R)$.

PROOF. We have

$$(*) \qquad \operatorname{zd}(R) \supset \bigcup_{\operatorname{Min}(R)} \mathfrak{p} = \bigcup_{\operatorname{Spec}(R)} \mathfrak{p} = \bigcup_{\operatorname{Max}(R)} \mathfrak{m}.$$

Since $R^\times = R \setminus \bigcup_{\operatorname{Max}(R)} \mathfrak{m}$, in the above (*) is an equality. □

Lemma 96.15. *Let R be a Noetherian domain of dimension one. Let a and c be non-zero elements of R. Let*

$$\mathfrak{A} = \bigcup_{n=0}^{\infty} (Rc : Ra^n) := \{x \in R \mid xa^n \in Rc \textit{ for some integer } n\}.$$

Then

$$\mathfrak{A} + Ra = R.$$

PROOF. Let $\mathfrak{A}_k = (Rc : Ra^k) := \{x \in R \,|\, xa^k \in Rc\}$. Since $\mathfrak{A}_k \subset \mathfrak{A}_{k+1}$, for all k, we know that $\mathfrak{A}$ is an ideal. Since R is Noetherian, there exists an integer n such that $\mathfrak{A}_n = \mathfrak{A}_{n+i} = \mathfrak{A}$ for all positive integers i. As c lies in $\mathfrak{A}_k$ for all k, we have $c \in \mathfrak{A}$. In particular, $\mathfrak{A}$ is not trivial.

Let $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism. Since $\mathfrak{A} > 0$ and R is a domain, it is clear that $\dim R/\mathfrak{A} = 0$. [We do not need Akizuki's Theorem.] By the corollary above, it suffices to establish the following:

Claim 96.16. *$\overline{a}$ is not a zero divisor in $R/\mathfrak{A}$:*

If this were false, then there would exist a y in $R \setminus \mathfrak{A}$ satisfying $\overline{ay} = 0$, i.e., $ay \in \mathfrak{A} = \mathfrak{A}_n$. We would then have $(ay)a^n \in Rc$ and that would imply that $y \in \mathfrak{A}_{n+1} = \mathfrak{A}_n = \mathfrak{A}$, a contradiction. This establishes the Claim. □

Theorem 96.17. (Krull-Akizuki Theorem) *Let A be a Noetherian domain of dimension one. Let F be the quotient field of A and let K/F be a finite field extension. Let B be a ring such that $A \subset B \subset K$. Then B is a Noetherian domain of dimension at most one and if $\mathfrak{B}$ is a nontrivial ideal of B, then $B/\mathfrak{B}$ is a finitely generated $\big(A/(\mathfrak{B} \cap A)\big)$-module of finite length.*

PROOF. We first make two reductions.

Reduction 1. We may assume that $F = K$:

Certainly, we may assume that K is the quotient field of B and, in fact, $K = F(x_1, \ldots, x_n)$ for some $x_i \in B$. Choose $0 \neq c \in A$ such that each cx_i is integral over A. Let $C = A[cx_1, ..., cx_n]$. Then C is integral over A and a finitely generated A-module. Thus C is a Noetherian domain of dimension one with quotient field K and contained in B. If $\mathfrak{C}$ is an ideal in C then $C/\mathfrak{C}$ is integral over $A/(\mathfrak{C}\cap A)$ and a finitely generated $\big(A/(\mathfrak{C}\cap A)\big)$-module. So $C/\mathfrak{C}$ and $A/(\mathfrak{C}\cap A)$ are Noetherian rings of the same dimension. In particular, one is Artinian if and only if the other is. Consequently, if $\mathfrak{B}$ is a nontrivial ideal in B then by clearing denominators and multiplying by an appropriate power of c, we see that $B\cap C$ is a nontrivial ideal in C. In particular, $B/\mathfrak{B}$ is a finitely generated $\big(A/(\mathfrak{B}\cap A)\big)$-module if it is a finitely generated $\big(C/(\mathfrak{B}\cap C)\big)$-module and has finite length as an $\big(A/(\mathfrak{B}\cap A)\big)$-module if it has finite length as a $\big(C/(\mathfrak{B}\cap C)\big)$-module. This completes the reduction.

Reduction 2. It suffices to show that the (A/Aa)-module B/Ba is finitely generated for any $0 \neq a \in A$:

To show that B is Noetherian of dimension at most one, it suffices, by Corollary 96.11 to Akizuki's Theorem, to show that $B/\mathfrak{B}$ is Artinian for any nonzero ideal $\mathfrak{B}$ of B. Let $0 < \mathfrak{B}$ be an ideal of B. We must also show that $B/\mathfrak{B}$ has finite length over $A/(\mathfrak{B}\cap A)$.

Claim. $0 < \mathfrak{B}\cap A$:

Let $\mathfrak{B}'$ be any nontrivial finitely generated A-submodule of $\mathfrak{B}$. Then there exists $0 \neq c \in A$ such that $c\mathfrak{B}'$ lies in the domain A by the first reduction. This establishes the Claim.

Let $0 \neq a$ lie in $\mathfrak{B}\cap A$. Then A/Aa is Artinian by Corollary 96.11 to Akizuki's Theorem. By assumption, B/Ba is a finitely generated (A/Aa)-module. Since $B/\mathfrak{B}$ is a cyclic (B/Ba)-module, it is also finitely generated as an (A/Aa)-module, hence has finite length over the Artinian ring A/Aa so also over the Artinian ring $A/(\mathfrak{B}\cap A)$. This also implies that $B/\mathfrak{B}$ is Artinian.

So to finish we are in the following situation. We have $0 \neq a \in A$ is fixed, and we must show that B/Ba is finitely generated as an (A/Aa)-module. We do this in a number of steps. Note, as before, that A/Aa is an Artinian ring.

Step 1. Let $x \in B\,(\subset F)$. Then there exists a positive integer n such that $x \in Aa^{-n}+Ba$:

Write $x = \dfrac{b}{c}$ with $b, c \in A$ and $c \neq 0$. Set

$$\mathfrak{B} = \bigcup_{n=0}^{\infty}(Ac : Aa^n) := \{y \in A \mid ya^n \in Ac,\ \text{ some } n\}.$$

By Lemma 96.15, we have $\mathfrak{B} + Aa = A$ so $1 = y + za$, some $y \in \mathfrak{B}$ and $z \in A$. Consequently, $x = yx + zax$. Since $y \in \mathfrak{B}$, by definition, there exists an integer n such that $ya^n \in Ac$, so

$$x = yx + zax = \frac{ya^n}{a^n}\frac{b}{c} + zax \ \text{ lies in } \ Aa^{-n} + Ba$$

as needed.

Step 2. Let $\mathfrak{A}_n := (Ba^n \cap A) + Aa$, an ideal of A. Then there exists a positive integer m such that $\mathfrak{A}_m = \mathfrak{A}_{m+i}$ for all positive integers i:

Each $\mathfrak{A}_n$ contains a so is nontrivial. Moreover, it is clear that the $\mathfrak{A}_n$ form a descending chain of ideals. Since A/Aa is Artinian, the descending chain of ideals

$$\cdots \supset \mathfrak{A}_n/Aa \supset \mathfrak{A}_{n+1}/Aa \supset \cdots$$

stabilizes, hence so does the corresponding chain of $\mathfrak{A}_n$'s.

Step 3. Let m be the integer in Step 2. Then $B \subset Aa^{-m} + Ba$:

Let $x \in B$ be fixed. Then by Step 1, there exists a minimal positive integer n so that $x \in Aa^{-n} + Ba$. If we show that $m \geq n$ then $a^m \in Aa^n$ and $Aa^{-n} \subset Aa^{-m}$ as needed. So we may assume that $n > m$. Write $x = ra^{-n} + ba$, with $r \in A$ and $b \in B$. Then

$$r = a^n(x - ba) \in Ba^n \cap A \subset (Ba^n \cap A) + Aa = \mathfrak{A}_n$$

and $\mathfrak{A}_n = \mathfrak{A}_{n+1} = \mathfrak{A}_m$. Hence $r = b_1 a^{n+1} + r_1 a$, for some $r_1 \in A$ and $b_1 \in B$ so $x = ra^{-n} + ba = (b_1 a^{n+1} + r_1 a)a^{-n} + ba$ lies in $Aa^{-n+1} + Ba$. This contradicts the minimality of n, so completes the step.

Step 4. B/Ba is a finitely generated (A/Aa)-module (and hence we are done):

By Step 3, we know that $B/Ba \subset (Aa^{-m} + Ba)/Ba$. Moreover, we know that the A-module $(Aa^{-m} + Ba)/Ba \cong Aa^{-m}/(Aa^{-m} \cap Ba)$ is cyclic, hence Noetherian. Thus B/Ba is a finitely generated A-module as needed. □

Corollary 96.18. *Let A be a Dedekind domain with quotient field F. If K/F is a finite field extension then A_K is also a Dedekind domain.*

Remarks 96.19. 1. If $K = F$ in the theorem then clearly F is the only field between A and F, i.e., the only such B with $\dim B = 0$ is F.

2. If $\mathfrak{B}$ is zero and $B = F$ then F is not a finitely generated A-module when $A < F$.

Exercises 96.20.

1. Prove Corollary 96.4
2. Prove Proposition 96.6
3. Let R be a commutative Artinian ring and M a finitely generated free R-module. Let N be a submodule of M that is also R-free. Show that $\operatorname{rank} N \leq \operatorname{rank} M$.
4. Let R be a commutative ring and M a finitely generated free R-module. Let N be a submodule of M that is also R-free. Show that $\operatorname{rank} N \leq \operatorname{rank} M$.
5. Why does it suffice to prove Claim 96.16.
6. Let A be a domain with quotient field F. Suppose that every ring B with $A \subset B \subset F$ is Noetherian. Show that $\dim A = 1$.

97. Affine Algebras

Throughout this section F will denote a field and R a commutative ring.

Recall a finitely generated commutative F-algebra, with F a field, is called an *affine F-algebra* . By the Hilbert Basis Theorem, such an algebra is a Noetherian ring. We wish to investigate the dimension theory attached to such algebras. We know that an integral extension of rings preserves dimension. As an affine F algebra A is isomorphic to a quotient of a polynomial ring $F[t_1, \dots, t_n]$ for some n, it seems reasonable to assume that if A is also a domain, then the transcendence degree of $qf(A)/F$ should play a crucial role in such an investigation. A fundamental theorem (whose proof we give is due to Nagata) about affine F-algebras bringing together these two observations is the following:

Theorem 97.1. (Noether Normalization Theorem) *Let F be a field and $A = F[x_1, \ldots, x_n]$ an affine F-algebra that is also a domain. Let $r = \operatorname{tr\,deg}_F qfA$. Then there exist $y_1, \ldots, y_r$ in A algebraically independent over F satisfying $A/F[y_1, \ldots, y_r]$ is integral. In particular, A is a finitely generated $F[y_1, \ldots, y_r]$-module. Let $\varphi : F[t_1, \ldots, t_r] \to A$ be the evaluation map given by $t_i \mapsto y_i$. Then*

(1) φ is an F-algebra monomorphism.
(2) ${}^a\varphi : \operatorname{Spec}(A) \to \operatorname{Spec}(F[t_1, \ldots, t_r])$ is surjective and has finite fibers.
(3) $\dim A = \dim F[y_1, \ldots, y_r] = \dim F[t_1, \ldots, t_r]$.

Proof. Let $A = F[t_1, \ldots, t_n]$. If $qf(A)/F$ is a finite field extension, then it is integral. Consequently, we may assume that $qf(A)/F$ is not algebraic. Relabeling, we may also assume that x_1 is transcendental over F. In addition, we may assume that $x_1, \ldots, x_n$ are algebraically dependent over F.

Let $(\mathbf{j}) := (j_1, \ldots, j_n)$ with j_i non-negative integers for $i = 1, \ldots, n$. As the x_i are algebraically dependent, we have an equation

$$(*) \qquad \sum_{(\mathbf{j})} a_{(\mathbf{j})} x_1^{j_1} \cdots x_n^{j_n} = 0$$

for some $a_{(\mathbf{j})} = a_{j_1, \ldots, j_n} \in F$ not all zero. Let $m_2, \ldots, m_n$ be positive integers (to be chosen) and set

$$y_i = x_i - x_1^{m_i} \ \text{ for } \ i = 2, \ldots n$$

and $m_1 = 1$. Define

$$(\mathbf{j}) \cdot (\mathbf{m}) := j_1 + j_2 m_2 + \cdots + j_n m_n.$$

Plugging $x_i = y_i + x_1^{m_i}$, $i = 2, \ldots, n$, into (*) yields an equation

$$(**) \qquad \sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})} + f(x_1, y_2, \ldots, y_n) = 0$$

with $f \in F[x_1, y_2, \ldots y_n]$ and having no monomial solely in x_1. In addition, if there is no cancellation in the term $\sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})}$, then then degree in x_1 of the equation in (**) satisfies.

$$(\dagger) \qquad \deg_{x_1} f < \deg_{x_1} \big(\sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})}\big).$$

Choose $d > \max\{j_i \mid a_{(\mathbf{j})} \neq 0\}$. and set

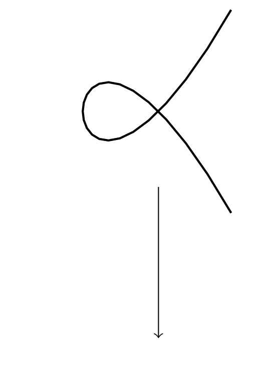
$$(\mathbf{m}) = (1, d, \ldots, d^{n-1}), \ \textit{i.e.}, \ m_i = d^{i-1} \ \text{ for } \ i = 0, \ldots, n-1.$$

As the $(\mathbf{j})$'s are distinct n-tuples, the $(\mathbf{j}) \cdot (\mathbf{m})$'s are all distinct, since they are the d-adic expansions of the $(\mathbf{j}) \cdot (\mathbf{m})$ and x_1 is transcendental over F. Thus ($\dagger$) holds, so (**) produces an integral equation for x_1 over $F[y_2, \ldots, y_r]$, as the leading coefficient of (**) in x_1 say $a_{(\mathbf{j})}$ is a unit in F. Since $x_i - (x_1^{m_i} + y_i) = 0$ for $i = 2, \ldots, n$ by definition, x_i is integral over $F[x_1, y_2, \ldots, y_n]$ for $i > 1$, hence over $F[y_2, \ldots, y_n]$. Consequently, $F[x_1, x_2, \ldots, x_n]/F[y_2, \ldots, y_n]$ is integral and $F[x_1, x_2, \ldots, x_n]$ is a finitely generated $F[y_2, \ldots, y_n]$-module.

If $y_2, \ldots, y_n$ are algebraically independent over F, we are done. If not we can repeat the process to finish this part of the theorem (as being integral and being a finitely generated module persists in towers).

The other statements in the theorem follow from the above and our previous work, e.g., using Lying Over, as $qf\big(A/F(y_1,\ldots,y_r)\big)$ is a finite algebraic extension and $F[y_1,\ldots,y_r] \cong F[t_1,\ldots,t_r]$ is a normal domain. $\square$

Remark 97.2. Using the Hilbert Nullstellensatz (cf. the discussion following Lemma 41.14 or another proof of it in 97.10 below), a geometric picture for the Noether Normalization Theorem for a curve when $F = \mathbb{C}$ is given by the following picture, where $f(x,y) = 0$ is the curve in $\mathbb{C}^2$ given by $f = y^2 - x^2(x+1)$ in $\mathbb{C}[x,y]$ with x, y variables the *nodal cubic* and A is the affine $\mathbb{C}$-algebra $\mathbb{C}[x,y]/(f)$:



$$\begin{array}{ccc} \mathrm{Max}(A) & & A = \mathbb{C}[x,y]/(f) \\ & & \uparrow \\ \mathrm{Spec}(\mathbb{C}[t]) \setminus \{(0)\} & & \mathbb{C}[t] \end{array}$$

Corollary 97.3. *Let F be a field. Then*

$$\dim F[t_1,\ldots,t_n] = n = \text{tr deg}_F F(t_1,\ldots,t_n).$$

PROOF. Since $(0) < (t_1) < \cdots < (t_1,\ldots,t_n)$ is a chain of prime ideals in $F[t_1,\ldots,t_n]$, we have $\dim F[t_1,\ldots,t_n] \geq n$. Suppose that $(0) < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m$ is a chain of prime ideals in $F[t_1,\ldots,t_n]$ and $A = F[t_1,\ldots,t_n]/\mathfrak{p}_1$, an affine F-domain, with $^- : F[t_1,\ldots,t_n] \to A$ the canonical epimorphism, so $A = F[\overline{t_1},\ldots,\overline{t_n}]$. By Noether Normalization, there exist $y_1,\ldots,y_r$ in A, algebraically independent over F satisfying $A/F[y_1,\ldots,y_r]$ is integral. Since $0 < \mathfrak{p}_1$, we must have $\overline{t_1},\ldots,\overline{t_n}$ are algebraically dependent over F, i.e., $r < n$. By induction, we have

$$\begin{aligned}\dim A &= \dim F[y_1,\ldots,y_r] = \dim F[t_1,\ldots,t_r] \\ &= r = \dim F[t_1,\ldots,t_n]/\mathfrak{p}_1 \geq m-1,\end{aligned}$$

since $\overline{\mathfrak{p}_1} < \cdots < \overline{\mathfrak{p}_m}$ is a chain of prime ideals in A. Consequently, $n \geq m$, and the result follows. $\square$

Notation 97.4. Let A be an affine F-algebra that is also a domain and $K = qf(A)$. Then there exists a transcendence basis of K over F in A by clearing denominators (or using Noether Normalization). We shall set

$$\text{tr deg}_F A := \text{tr deg}_F K.$$

Corollary 97.5. *Let A be an affine F-algebra that is also a domain. Then* $\dim A =$ tr $\deg_F A$.

Proof. Let $r' = \text{tr deg}_F A$. By Noether Normalization there exist $y_1, \dots, y_r$ in A algebraically independent over F with $A/F[y_1, \dots, y_r]$ integral. Therefore, $qf(A)/F(y_1, \dots, y_r)$ is finite algebraic, so $r = r'$, and we have

$$\dim A = \dim F[y_1, \dots, y_r] = \text{tr deg}_F F[y_1, \dots y_r] = r = \text{tr deg}_F A.$$

□

Corollary 97.6. *Let A be an affine F-algebra and $\mathfrak{p}$ a prime ideal in A. Then $\dim V(\mathfrak{p}) = \text{tr deg}_F A/\mathfrak{p}$.*

Proof. $\dim V(\mathfrak{p}) = \dim A/\mathfrak{p}$. □

As a consequence of Noether Normalization, we also obtain another proof of Zariski's Lemma.

Corollary 97.7. (Zariski's Lemma) *Let A be an affine F-algebra. If A is a field, then A/F is a finite field extension.*

Proof. We have $0 = \dim A = \text{tr deg}_F A$, so the finitely generated field extension A/F is algebraic, hence finite. □

Corollary 97.8. *Let $\varphi : A \to B$ be an F-algebra homomorphism of affine F-algebras. Then the restriction of ${}^a\varphi$ to $\text{Max}(B)$ satisfies*

$${}^a\varphi|_{\text{Max}(B)} : \text{Max}(B) \to \text{Max}(A).$$

Proof. Let $\mathfrak{m}$ be a maximal ideal in B. Then $\mathfrak{p} = {}^a\varphi(\mathfrak{m})$ is a prime ideal in A and φ induces a ring monomorphism $A/\mathfrak{p} \to B/\mathfrak{m}$. As the quotient of a affine F-algebra is an affine F-algebra, $B/\mathfrak{m}$ is a finite field extension of F by by Zariski's Lemma. Thus $F \subset A/\mathfrak{p} \subset B/\mathfrak{m}$ which implies that $A/\mathfrak{p}$ is a field. Therefore, $\mathfrak{p}$ a maximal ideal in A. □

Recall from §36, if $A = F[t_1, \dots, t_n]$ and $\mathfrak{A} \subset A$ is an ideal, the *affine variety* of $\mathfrak{A}$ in F^n is defined by

$$Z_F(\mathfrak{A}) = \{\underline{a} = (a_1, \dots, a_n) \in F^n \mid f(\underline{a}) = 0 \text{ for all } f \in \mathfrak{A}\},$$

the collection of all such forms a system of closed sets for the geometric Zariski topology of F^n. We have seen in Section 41, how Zariski's Lemma implies the (Weak) Hilbert Nullstellensatz:

Theorem 97.9. (Hilbert Nullstellensatz) (Weak Form) *Suppose that F is an algebraically closed field, $A = F[t_1, \dots, t_n]$, and $\mathfrak{A} = (f_1, \dots, f_r)$ is an ideal in A. Then $Z_F(\mathfrak{A})$ is the empty set if and only if $\mathfrak{A}$ is the unit ideal. In particular, if $\mathfrak{A} < R$, then there exists a point $\underline{a} \in F^n$ satisfying $f_1(\underline{a}) = 0, \dots, f_r(\underline{a}) = 0$.*

and using the Rabinowitch Trick, how this implies the (Strong) Hilbert Nullstellensatz):

Theorem 97.10. (Hilbert Nullstellensatz) (Strong Form) *Suppose that F is an algebraically closed field and $A = F[t_1, \dots, t_n]$. Let $f, f_1, ..., f_r$ be elements in A and $\mathfrak{A} = (f_1, ..., f_r) \subset R$. Suppose that $f(\underline{a}) = 0$ for all $\underline{a} \in Z_F(\mathfrak{A})$. Then there exists an integer m such that $f^m \in \mathfrak{A}$, i.e., $f \in \sqrt{\mathfrak{A}}$. In particular, if $\mathfrak{A}$ is a prime ideal, then $f \in \mathfrak{A}$.*

We give another proof of Strong Form of the Hilbert Nullstellensatz. We begin this proof by reducing it to another form of the theorem.

Proof. Let $A = F[t_1, \ldots, , t_n]$ and $\operatorname{Max}(A) \subset \operatorname{Spec}(A)$ have the induced topology. The Weak Form of the Hilbert Nullstellensatz implies that the map

$$F^n \to \operatorname{Max}(A) \text{ given by } \underline{a} = (a_1, \ldots, a_n) \mapsto \mathfrak{m}_{\underline{a}} := (t_1 - a_1, \ldots, t_n - a_n)$$

is a homeomorphism. Let $\mathfrak{A} < A$ be an ideal. Recall that $\mathcal{I}(V) := \bigcap_V \mathfrak{p}$ if $V \subset \operatorname{Spec}(A)$, so $\sqrt{\mathfrak{A}} = \mathcal{I}\big(V(\mathfrak{A})\big) = \bigcap_{V(\mathfrak{A})} \mathfrak{p}$. Therefore, it suffices to prove that

$$\sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{V(\mathfrak{A}) \cap \operatorname{Max}(A)} \mathfrak{m}$$

in the above.

In fact, this is independent of the fact that F be algebraically closed and is an algebraic version of the Strong Form of the Hilbert Nullstellensatz that we shall now state and prove.

Theorem 97.11. (Hilbert Nullstellensatz) (Algebraic Form) *Let A be an affine F-algebra and $\mathfrak{A} < A$ an ideal. Then*

$$\sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A}) \cap \operatorname{Max}(A)} \mathfrak{m},$$

i.e., the closed points in $\operatorname{Spec}(A)$ *determines the variety* $V(\mathfrak{A})$.

Proof. Let $\mathfrak{B} = \bigcap_{V(\mathfrak{A}) \cap \operatorname{Max}(A)} \mathfrak{m}$. Therefore, $\sqrt{\mathfrak{A}} \subset \mathfrak{B}$. Suppose that $\sqrt{\mathfrak{A}} < \mathfrak{B}$. Let $b \in \mathfrak{B} \setminus \sqrt{\mathfrak{A}}$ and $S = \{b^n \mid n \geq 0\}$. Then $S \bigcap \sqrt{\mathfrak{A}} = \emptyset$. Let $\varphi : A \to S^{-1}A$ be the canonical homomorphism given by $a \mapsto \frac{a}{1}$. Since $S^{-1}A = A[b^{-1}]$, it is also an affine F-algebra, hence φ is an F-algebra homomorphism of affine F-algebras, so ${}^a\varphi$ takes maximal ideals to maximal ideals. By choice, $S^{-1}\mathfrak{A} < S^{-1}A$, so there exists a maximal ideal $\mathfrak{n}$ in $V(S^{-1}\mathfrak{A}) \cap \operatorname{Max}(S^{-1}A)$. Since $\varphi(b)$ is a unit in $S^{-1}A$, we know that $\varphi(b) \notin \mathfrak{n}$. We also have $\mathfrak{A} = \varphi^{-1}(S^{-1}\mathfrak{A}) \subset \varphi^{-1}(\mathfrak{n}) = {}^a\varphi(\mathfrak{n})$, so ${}^a\varphi(\mathfrak{n})$ lies in $V(\mathfrak{A}) \cap \operatorname{Max}(A)$. Therefore, we have

$$b \in \mathfrak{B} = \bigcap_{V(\mathfrak{A}) \cap \operatorname{Max}(A)} \mathfrak{m} \subset {}^a\varphi(\mathfrak{n}),$$

hence $\varphi(b) \in \mathfrak{n}$, a contradiction. $\square$

A ring satisfying the conclusion of this form of the Nullstellensatz is called a *Jacobson* (or *Hilbert*) ring.

Corollary 97.12. *Let A be an affine F-algebra. Then* $\operatorname{Max}(A)$ *is a dense subset of* $\operatorname{Spec}(A)$.

Proof. Let $\mathfrak{A} < A$ be an ideal and $U = \operatorname{Spec}(A) \setminus V(\mathfrak{A})$. If $\operatorname{Max}(A) \subset V(\mathfrak{A})$, then we have

$$\operatorname{nil}(A) = \bigcap_{\operatorname{Spec}(A)} \mathfrak{p} = \bigcap_{\operatorname{Max}(A)} \mathfrak{m} = \bigcap_{\operatorname{Max}(A) \cap V(\mathfrak{A})} \mathfrak{m} = \sqrt{\mathfrak{A}}$$

by the Algebraic Form of the Hilbert Nullstellensatz, so $\operatorname{Spec}(A) = V(\sqrt{\mathfrak{A}}) = V(\mathfrak{A})$ and U is empty. The result follows. $\square$

Let A be an affine F-algebra. Then $A \cong F[t_1, \ldots, t_N]/\mathfrak{A}$ for some N and ideal $\mathfrak{A} < F[t_1, \ldots, t_N]$. As $V(\mathfrak{A}) \cong \operatorname{Spec}(F[t_1, \ldots, t_N]/\mathfrak{A}) \cong \operatorname{Spec}(A)$ and A is a a Noetherian ring by the Hilbert Basis Theorem, $\operatorname{Spec}(A)$ is a Noetherian space. Therefore, it decomposes into finitely many indecomposable components, i.e., the irreducible varieties $V(\mathfrak{p})$ with $\mathfrak{p}$ in the finite set $\operatorname{Min}(A)$. Of course, each prime ideal (respectively, minimal prime ideal, maximal prime ideal) $\mathfrak{p}$ in $\operatorname{Spec}(A)$ is isomorphic to $\mathfrak{P}/\mathfrak{A}$ for some prime ideal $\mathfrak{P}$ (and respectively, minimal, maximal) in $V(\mathfrak{A})$. If $\mathfrak{p} \in \operatorname{Min}(A)$, then $\dim V(\mathfrak{p}) = \operatorname{tr\,deg}_F A/\mathfrak{p}$. Different minimal primes can result in different dimensions. Of course, one is interested in determining when all irreducible components have the same dimension, a subject that we shall not pursue. Suppose that $\mathfrak{P}$ is a prime ideal in A. We define $\operatorname{codim}_{\operatorname{Spec}(A)} V(\mathfrak{P}) := \dim A - \dim V(\mathfrak{P})$, the *codimension* of $V(\mathfrak{P})$ in $\operatorname{Spec}(A)$ and if $\mathfrak{p} \subset \mathfrak{P}$ is another prime ideal, hence $V(\mathfrak{P}) \subset V(\mathfrak{p})$, we define $\operatorname{codim}_{V(\mathfrak{p})} V(\mathfrak{P}) := \dim V(\mathfrak{p}) - \dim V(\mathfrak{P})$, the *codimension* of $V(\mathfrak{P})$ in $V(\mathfrak{p})$. We wish to show in all cases that

$$\operatorname{ht}\mathfrak{p} = \operatorname{codim}_{V(\mathfrak{p})} V(\mathfrak{P}) = \operatorname{tr\,deg}_F A/\mathfrak{p} - \operatorname{tr\,deg}_F A/\mathfrak{P}. \tag{97.13}$$

In particular,

$$\dim A/\mathfrak{P} + \operatorname{ht}\mathfrak{P} = \dim A = \dim V(\mathfrak{P}) + \operatorname{ht}\mathfrak{P}.$$

Of course, if height is to be codimension, we would also want if $\mathfrak{p} \subset \mathfrak{P}$ are prime ideals in A that

$$\operatorname{ht}\mathfrak{P} = \operatorname{ht}\mathfrak{p} + \operatorname{ht}\mathfrak{P}/\mathfrak{p}.$$

Intuitively, still more should be expected. If $\mathfrak{B} = (f_1, \ldots, f_n)$ is an ideal in A, geometrically, we would want to view the f_i as hypersurfaces in $V(\mathfrak{B})$, e.g., if $A = F[t_1, \ldots, t_N]$, then the f_i are polynomials in $F[t_1, \ldots, t_N]$, and we can view this as the hypersurface $f_i = 0$ in F^N. Taking our cue from linear algebra, we would then expect if $V(\mathfrak{p})$ is an irreducible component of $V(\mathfrak{B})$ that

$$\dim V(\mathfrak{p}) \geq \dim A - n, \tag{97.14}$$

i.e., $\operatorname{codim}_{\operatorname{Spec}(A)} V(\mathfrak{p}) \leq n$ or if height is the same as codimension that

$$\operatorname{ht}\mathfrak{p} \leq n.$$

We shall, in fact, show all of this to be true for an affine F-algebra. Note that if F is algebraically closed, the above still will hold with $Z(\mathfrak{A})$ replacing $V(\mathfrak{A})$, etc. It will therefore follow that the algebraic notions of Krull dimension and height translate into the correct geometric notions.

Most of the above will follow by our next theorem. We first make the following definition:

Definition 97.15. A chain of prime ideals

$$\mathfrak{p}_0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$

in a commutative ring is called *saturated* if no further new primes can be added to this chain.

We establish equation (97.13).

Theorem 97.16. *Let F be a field and A be an affine F-algebra that is a domain, $\mathfrak{p}$ a prime ideal in A, and*
$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$
a saturated chain of prime ideals in A. Then

(1) $\operatorname{ht}\mathfrak{p} = m$.

(2) $\dim A = \dim A/\mathfrak{p} + \operatorname{ht}\mathfrak{p}$.

Moreover, $\operatorname{ht}\mathfrak{p} = \operatorname{codim}_{\operatorname{Spec}(A)} V(\mathfrak{p})$ *and all maximal saturated chains of prime ideals in* $\mathfrak{p}$ *(hence starting from* 0 *and ending at* $\mathfrak{p}$*) have the same number of links. In particular,* $\dim A = \operatorname{ht}\mathfrak{m}$ *for all maximal ideals in A.*

Proof. Clearly, we need only prove (1) and (2). Let $n = \dim A$, finite by Noether Normalization, and $A = F[x_1, \dots, x_r]$.

Case 1. If $m = 1$, then (1) and (2) hold.

As $0 < \mathfrak{p}$ is a saturate chain of prime ideals, $\operatorname{ht}\mathfrak{p} = 1$, so (1) holds.

Subcase 1. $n = r$.

Since $\dim A = \operatorname{tr}\deg_F A$, the elements $x_1, \dots, x_n$ are algebraically independent over F. Therefore, $A \cong F[t_1, \dots, t_n]$ is a UFD. In particular, $\mathfrak{p}$ must contain a prime element f, so we must have $\operatorname{ht}\mathfrak{p} = 1$ and $\mathfrak{p} = (f)$. Since f is nonzero, there exists an i such that x_i occurs nontrivially in f, say $i = n$. Let $^- : A \to A/\mathfrak{p}$ be the canonical epimorphism, so $\overline{f(x_1, \dots, x_n)} = f(\overline{x_1}, \dots, \overline{x_n}) = 0$ in the domain $\overline{A} = F[\overline{x_1}, \dots, \overline{x_n}]$, i.e., $\overline{x_1}, \dots, \overline{x_n}$ are algebraically dependent over F. Therefore, $\operatorname{tr}\deg_F \overline{A} = \operatorname{tr}\deg F[\overline{x_1}, \dots, \overline{x_n}] < n$. If $g \in A$ satisfies $\overline{g} = g(\overline{x_1}, \dots, \overline{x_n}) = 0$, then $g \in \ker{}^- = (f)$, so $f \mid g$ in A. In particular, $g = 0$ or x_n occurs nontrivially in g. In particular, if $h \in F[x_1, \dots, x_{n-1}] \cong F[t_1, \dots, t_{n-1}]$ satisfies $h(\overline{x_1}, \dots, \overline{x_{n-1}}) = 0$, then $h = 0$. It follows that $\overline{x_1}, \dots, \overline{x_{n-1}}$ are algebraically independent over F, so $\dim \overline{A} = \operatorname{tr}\deg_F \overline{A} = n - 1$ and $\dim A = \dim \overline{A} + 1 = \dim \overline{A} + \operatorname{ht}\mathfrak{p}$.

Subcase 2. $n \neq r$.

By Noether Normalization there exist $y_1, \dots, y_n$ in A algebraically independent over F satisfying $A/F[y_1, \dots, y_n]$ is integral. Let $R = F[y_1, \dots, y_n]$ and $\mathfrak{p}' = \mathfrak{p} \cap R$, a prime ideal in R. We have
$$\dim A = \dim R \text{ and } \operatorname{ht}\mathfrak{p} = \operatorname{ht}\mathfrak{p}',$$
the second equality by Going Down as R is a normal domain. We also have the monomorphism $R/\mathfrak{p}' \to A/\mathfrak{p}$ is integral, hence $\dim R/\mathfrak{p}' = \dim A/\mathfrak{p}$. By Subcase 1 applied to R, we have
$$\dim A = \dim R = \dim R/\mathfrak{p}' + \operatorname{ht}\mathfrak{p}' = \dim A/\mathfrak{p} + \operatorname{ht}\mathfrak{p} = \dim(A/\mathfrak{p}) + 1,$$
which is (2).

Case 2. Suppose that $\mathfrak{p}$ is a maximal ideal in A. Then $\operatorname{ht}\mathfrak{p} = m = \dim A$.

We have a saturated chain
$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}.$$
We show that $m = \operatorname{ht}\mathfrak{p} = \dim A$ by induction on $\dim A$. If $\dim A = 1$, the result is immediate, so we may assume that $\dim A > 1$. By Noether Normalization, there exist $y_1, \dots, y_n$ in A algebraically independent over F satisfying $A/F[y_1, \dots, y_n]$ is integral. Let $R = F[y_1, \dots, y_n]$, a normal domain, so by Going Down, we have $\operatorname{ht}\mathfrak{p} = \operatorname{ht}(\mathfrak{p} \cap R)$.

Moreover, the inclusion $R \hookrightarrow A$ is an integral map, so $\mathfrak{p} \cap R$ is a maximal ideal in R. Since $\dim A = \dim R$ and

$$0 < \mathfrak{p}_1 \cap R < \cdots < \mathfrak{p}_m \cap R = \mathfrak{p} \cap R$$

by Comparability, it suffices to show that $m = n$ in the case that $A = R$. So assume that $A = R$. We know, by the definition of $\dim A$, that

$$m \leq \dim A = \operatorname{tr\,deg}_F A = n.$$

By the argument in Case 1, Subcase 1, we have

$$n - 1 = \dim A/\mathfrak{p}_1 = \operatorname{tr\,deg}_F A - 1.$$

Let $^- : A \to A/\mathfrak{p}_1$, the canonical epimorphism. Then we have

$$0 = \overline{\mathfrak{p}_1} < \cdots < \overline{\mathfrak{p}_m} = \overline{\mathfrak{p}}$$

with $\overline{\mathfrak{p}}$ a maximal ideal in $\overline{A}$ by the Correspondence Principle. By induction on n, we have $m - 1 = n - 1$, so $m = n$. This establishes Case 2.

Case 3. For any prime ideal $\mathfrak{p}$ in A, (1) and (2) hold.

Let $\mathfrak{m} \in V(\mathfrak{p}) \cap \operatorname{Max}(A)$. By Case 2, every saturated chain of prime ideals in $\mathfrak{m}$ has $n = \operatorname{ht} \mathfrak{m} = \dim A$ links. Extend the saturated chain

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$

in A to a saturated chain of prime ideals

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m < \cdots < \mathfrak{p}_n = \mathfrak{m}$$

in A using the fact that A is Noetherian. Since $\mathfrak{m}/\mathfrak{p}_m$ is a maximal ideal in $A/\mathfrak{p}_m$, by Case 2, we have

$$\operatorname{ht} \mathfrak{m}/\mathfrak{p}_m = \dim A/\mathfrak{p}_m = n - m.$$

Clearly,

$$\dim A \geq \dim A/\mathfrak{p}_m + \operatorname{ht} \mathfrak{p}_m \text{ and } \operatorname{ht} \mathfrak{p}_m \geq m,$$

so

$$\begin{aligned}\dim A &\geq \dim A/\mathfrak{p}_m + \operatorname{ht} \mathfrak{p}_m \\ &= (n - m) + \operatorname{ht} \mathfrak{p}_m \geq (n - m) + m = n = \dim A,\end{aligned}$$

hence

$$\dim A = \dim A/\mathfrak{p} + \operatorname{ht} \mathfrak{p} \text{ and } \operatorname{ht} \mathfrak{p} = m$$

as needed. □

A Noetherian ring R is called *catenary* if all prime ideals $\mathfrak{p} < \mathfrak{P}$ in R and saturated chains of prime ideals beginning at $\mathfrak{p}$ and ending at $\mathfrak{P}$ have the same finite number of links and is called *universally catenary* if every finitely generated commutative R-algebra is catenary. It is easy to show the following:

Lemma 97.17. *If R is a catenary ring and $\mathfrak{A} < R$ an ideal, then $R/\mathfrak{A}$ is a catenary ring.*

As every affine F-algebra that is also a domain, e.g., $F[t_1, \ldots, t_n]$, is catenary by the theorem, by the lemma, we conclude that:

Corollary 97.18. *Every field is universally catenary.*

We turn to the last fact that we wish to prove, viz., if $\mathfrak{A} = (f_1, \dots, f_n)$ is an ideal in an affine F-algebra, then $\operatorname{ht}\mathfrak{p} \leq n$. We prove this in greater generality. This result, due to Krull, is one of the foundational theorems in commutative algebra.
Recall if $\mathfrak{A} < R$ is an ideal, the *height* of $\mathfrak{A}$ is

$$\operatorname{ht}\mathfrak{A} := \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A})\} = \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}\}.$$

Theorem 97.19. (Principal Ideal Theorem) *Let R be a commutative Noetherian ring, $\mathfrak{A} < R$ an ideal and $\mathfrak{p} \in V(\mathfrak{A})$ minimal. If $\mathfrak{A} = (x_1, \dots, x_n)$, then $\operatorname{ht}\mathfrak{p} \leq n$. In particular, $\operatorname{ht}\mathfrak{A} \leq n$.*

Proof. The crucial result is the case that $n = 1$, hence the theorem's name.

Case 1: $\mathfrak{A} = (x)$ is principal.

If $x = 0$ or even nilpotent, then $\mathfrak{p} \in \operatorname{Min}(R)$ and $\operatorname{ht}\mathfrak{p} = 0$. So we may assume that x is nonzero. Suppose that the result is false, i.e., $\operatorname{ht}\mathfrak{p} > 1$. Then we have a chain of prime ideals in R,

$$\mathfrak{p}_0 < \mathfrak{p}_1 < \mathfrak{p} \text{ with } \mathfrak{p}_0, \mathfrak{p}_1 \notin V(\mathfrak{A}),$$

since $\mathfrak{p} \in V(\mathfrak{A})$ is minimal. We make some reductions. By the Correspondence Principle, we have $\mathfrak{p}/\mathfrak{p}_0 \in V\Big(\big((x) + \mathfrak{p}_0\big)/\mathfrak{p}_0\Big)$ is minimal and $\operatorname{ht}\mathfrak{p}/\mathfrak{p}_0 \geq 2$. Replacing R by $R/\mathfrak{p}_0$, we may assume that R is a domain and $\mathfrak{p}_0 = 0$. Next let $S = R \setminus \mathfrak{p}$. Then $R_\mathfrak{p} = S^{-1}R$ satisfies $S^{-1}\mathfrak{p} \in V(R_\mathfrak{p}x)$ is minimal and $0 < S^{-1}\mathfrak{p}_1 < S^{-1}\mathfrak{p}$. So replacing R by $R_\mathfrak{p}$ and changing notation (considerably), we are reduced to R satisfying the following conditions:

(1) R is a Noetherian domain
(2) $(R, \mathfrak{m})$ is a local ring.
(3) $\operatorname{ht}\mathfrak{m} \geq 2$, so there exists a chain of prime ideals $0 < \mathfrak{P} < \mathfrak{m}$ in R.
(4) The element $x \in \mathfrak{m}$ satisfies $V(x) = \{\mathfrak{m}\}$. In particular, $x \notin \mathfrak{P}$.

Let $\mathfrak{A}_i = (\mathfrak{P}R_\mathfrak{P})^i \cap R < R$ a finitely generated ideal in R for every non-negative integer i. Note that $(R_\mathfrak{P}, \mathfrak{P}R_\mathfrak{P})$ is a Noetherian local domain with $R \subset R_\mathfrak{P} \subset qf(R)$.

We have a descending chain of ideals in R:

$$\mathfrak{P} = \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_i \supset \cdots.$$

As $\mathfrak{m} \in V(x)$ is minimal, $\operatorname{Spec}\big(R/(x)\big) = \{\mathfrak{m}/(x)\}$, hence $R/(x)$ is a Noetherian ring of dimension zero. By Akizuki's Theorem 96.9, $R/(x)$ is an Artinian ring. Therefore, the descending chain of ideals

$$\big(\mathfrak{A}_1 + (x)\big)/(x) \supset \cdots \supset \big(\mathfrak{A}_i + (x)\big)/(x) \supset \cdots$$

in $R/(x)$ stabilizes, say $\big(\mathfrak{A}_j + (x)\big)/(x) = \big(\mathfrak{A}_{j+n} + (x)\big)/(x)$ for all $n \geq 0$. By the Correspondence Theorem, $\big(\mathfrak{A}_j + (x)\big) = \big(\mathfrak{A}_{j+n} + (x)\big)$. Let $a \in \mathfrak{A}_j$. Then

$$a = b + rx, \text{ for some } b \in \mathfrak{A}_{j+1}, \text{ and } r \in R.$$

In particular, $rx = a - b \subset \mathfrak{A}_j \subset (\mathfrak{P}R_\mathfrak{P})^j \subset R_\mathfrak{P}$. As $x \notin \mathfrak{P}$, it must be a unit in $R_\mathfrak{P}$, consequently, $r \in (\mathfrak{P}R_\mathfrak{P})^j$. It follows that $r \in (\mathfrak{P}R_\mathfrak{P})^j \cap R = \mathfrak{A}_j$, i.e.,

$$\mathfrak{A}_j = \mathfrak{A}_{j+1} + \mathfrak{A}_j x \subset \mathfrak{A}_{j+1} + \mathfrak{A}_j\mathfrak{m} \subset \mathfrak{A}_j.$$

Therefore, $\mathfrak{A}_j = \mathfrak{A}_{j+1} + \mathfrak{A}_j\mathfrak{m}$ in the local ring $(R, \mathfrak{m})$ with $\mathfrak{A}_j$ finitely generated. By alternate form of Nakayama's Lemma, Corollary 93.11, this implies that we have $\mathfrak{A}_j = \mathfrak{A}_{j+1}$. Thus,

$$(\mathfrak{P}R_{\mathfrak{P}})^j = \mathfrak{A}_j R_{\mathfrak{P}} = \mathfrak{A}_{j+1}R_{\mathfrak{P}} = (\mathfrak{P}R_{\mathfrak{P}})^{j+1}$$

in the local Noetherian ring $(R_{\mathfrak{P}}, \mathfrak{P}R_{\mathfrak{P}})$. Hence $(\mathfrak{P}R_{\mathfrak{P}})^j = 0$ in the domain $R_{\mathfrak{P}}$, so $\mathfrak{P}R_{\mathfrak{P}} = 0$ and $\mathfrak{P} = 0$, a contradiction. This proves Case 1.

Case 2. General case (the proof is due to Akizuki).

Let $\mathfrak{A} = (a_1, \dots, a_n) < R$ be an ideal with $\mathfrak{p} \in V(\mathfrak{A})$ minimal. Suppose that $\operatorname{ht}\mathfrak{p} > n$. Choose a prime ideal $\mathfrak{p}_1$ in R with $\mathfrak{p}_1 < \mathfrak{p}$, and $\operatorname{ht}\mathfrak{p}_1 \geq n$. Replacing R by $R_{\mathfrak{p}}$ if necessary, we may assume that $(R, \mathfrak{m})$ is a local ring with $\mathfrak{m} = \mathfrak{p}$, so $V(\mathfrak{A}) = \{\mathfrak{m}\}$. As R is Noetherian, we may assume that $\mathfrak{p}_1 < \mathfrak{m}$ is saturated, i.e., there exists no prime ideal $\mathfrak{q}$ in R with $\mathfrak{p}_1 < \mathfrak{q} < \mathfrak{m}$. (Why?) Since $\mathfrak{p}_1 \notin V(\mathfrak{A})$, there exists an i such that $a_i \notin \mathfrak{p}_1$. We may assume that $a_1 \notin \mathfrak{p}_1$. Therefore, we have

$$\mathfrak{p}_1 < \mathfrak{p}_1 + (a_1) \subset \mathfrak{p} = \mathfrak{m}.$$

Hence $V\big(\mathfrak{p}_1 + (a_1)\big) = \{\mathfrak{m}\}$, as $\mathfrak{p}_1 < \mathfrak{m}$ is saturated. Consequently, we have

$$(*) \qquad \sqrt{\mathfrak{p}_1 + (a_1)} = \bigcap_{V\left(\mathfrak{p}_1 + (a_1)\right)} \mathfrak{P} = \mathfrak{m}.$$

Since $\mathfrak{m}$ is finitely generated, there exists a positive integer k satisfying $\mathfrak{m}^k \subset \mathfrak{p}_1 + (a_1)$ by (*) (and the definition of the radical). As we also have $\mathfrak{A} \subset \mathfrak{m}$, we conclude that $\mathfrak{A}^k \subset \mathfrak{m}^k \subset \mathfrak{p}_1 + (a_1)$. Therefore, we have equations

$$a_i^k = b_i + c_i a_1 \text{ for some } b_i \in \mathfrak{p}_i \text{ and } c_i \in R,\ i = 2, \dots, n.$$

Let $\mathfrak{B} = (b_2, \dots, b_n) \subset \mathfrak{p}_1$, so $\mathfrak{p}_1 \in V(\mathfrak{B})$. As $\operatorname{ht}\mathfrak{p}_1 \geq n$, by induction, we conclude that there exists a prime ideal $\mathfrak{q} \in V(\mathfrak{B})$ minimal with $\mathfrak{B} \subset \mathfrak{q} < \mathfrak{p}_1$. However, $a_1^k, \dots, a_n^k$ all lie in $\mathfrak{B} + (a_1) \subset \mathfrak{q} + (a_1)$, so

$$\{\mathfrak{m}\} = V(\mathfrak{A}) = V(a_1, \dots, a_n) = V(a_1^k, \dots a_n^k) \supset V\big(\mathfrak{q} + (a_1)\big).$$

Therefore, $V\big(\mathfrak{q} + (a_1)\big) = \{\mathfrak{m}\}$ and $\mathfrak{m}/\mathfrak{q} \in V\big((\mathfrak{q} + (a_1))/\mathfrak{q}\big)$ is minimal (working in the ring $R/\mathfrak{q}$). By Case 1, we know that $\operatorname{ht}\mathfrak{m}/\mathfrak{q} \leq 1$, which contradicts the existence of the chain of primes $0 = \mathfrak{q}/\mathfrak{q} < \mathfrak{p}_1/\mathfrak{q} < \mathfrak{m}/\mathfrak{q}$ in $R/\mathfrak{q}$. □

The Principal Ideal Theorem allows up to show equation (97.14) is valid.

Corollary 97.20. *Let F be a field and A be an affine F-algebra that is a domain of dimension n. If $f_1, \dots f_m \in A$ with $m < n$, then* $\dim V(f_1, \dots, f_n) \geq n - m > 0$.

Proof. Let $\mathfrak{p}$ a minimal prime in $(f_1, \dots, f_m)$. Then $\operatorname{ht}\mathfrak{p} \leq m$ by the Principal Ideal Theorem. The result follows by Theorem 97.16. □

Since ideals in a Noetherian ring are finitely generated, we have:

Corollary 97.21. *Let R be a Noetherian ring and $\mathfrak{p}$ a prime ideal in R. Then* $\operatorname{ht}\mathfrak{p}$ *is finite. In particular, the set of prime ideals in R satisfies the descending chain condition.*

Corollary 97.22. *Every local Noetherian ring has finite dimension.*

Corollary 97.23. *Let R be a Noetherian domain. Then R is a UFD if and only if every height one prime in R is principal if and only if the set of height one primes is equal to the set of nonzero principal prime ideals.*

Proof. We know by Kaplansky's Theorem 31.1 that R is a UFD if and only if every nonzero prime ideal in R contains a prime element. As every nonzero prime ideal in R contains a prime ideal of height one by Corollary 97.21, the result follows. $\square$

We shall also need a further generalization of the Principal Ideal Theorem.

Theorem 97.24. *Let R be a Noetherian ring, $\mathfrak{A} < R$ an ideal generated by n elements, and $\mathfrak{p} \in V(\mathfrak{A})$ a prime satisfying $\operatorname{ht}_{R/\mathfrak{A}} \mathfrak{p}/\mathfrak{A} = m$. Then*

$$\operatorname{ht}_R \mathfrak{p} \leq m + n = h_{R/\mathfrak{A}}\, \mathfrak{p}/\mathfrak{A} + n.$$

Proof. We induct on m (and all n simultaneously).
$m = 0$: As $\mathfrak{p} \in V(\mathfrak{A})$ is minimal, this follows from the Principal Ideal Theorem.
$m > 1$: As R is Noetherian, there exist finitely prime ideals minimal over $\mathfrak{A}$, say $\mathfrak{p}_1, , \dots, \mathfrak{p}_r$. As $\mathfrak{p}$ is not minimal, $\mathfrak{p} \not\subset \bigcap_{i=1}^r \mathfrak{p}_i$ by the Prime Avoidance Lemma 93.16. Let $a \in \mathfrak{p} \setminus \bigcap_{i=1}^r \mathfrak{p}_i$ and set $\mathfrak{B} = \mathfrak{A} + Ra$. Then $\mathfrak{p}_i \notin V(\mathfrak{B})$ for $i = 1, \dots, r$, so there exist no chain of prime ideals in $V(\mathfrak{B})$ starting at a $\mathfrak{p}_i$ and ending in $\mathfrak{p}$. It follows that $\operatorname{ht}_{R/\mathfrak{B}} \mathfrak{p} \leq m - 1$, since $V(\mathfrak{B}) \subset V(\mathfrak{A})$. By induction, we have

$$\operatorname{ht} \mathfrak{p} \leq \operatorname{ht}_{R/\mathfrak{B}} \mathfrak{p}/\mathfrak{B} + (n+1) \leq n + m. \qquad \square$$

Corollary 97.25. *Let R be a Noetherian ring and $\mathfrak{p}$ a prime ideal in R of height m. Suppose that x is an element in $\mathfrak{p}$.*

(1) *We have $m - 1 \leq \operatorname{ht}_{R/(x)} \mathfrak{p}/(x) \leq m$.*
(2) *Suppose that x does not lie in any minimal prime of R. Then $\operatorname{ht}_{R/(x)} \mathfrak{p}/(x) = m - 1$.*
(3) *If x is not a zero divisor in R, then $\operatorname{ht}_{R/(x)} \mathfrak{p}/(x) = m - 1$.*

Proof. Let $s = \operatorname{ht}_{R/(x)} \mathfrak{p}/(x)$ with $\mathfrak{p} \in V(x)$. By the Principal Ideal Theorem 97.19, $m = \operatorname{ht} \mathfrak{p} \leq s + 1$, and it is immediate that $s = \operatorname{ht}_{R/(x)} \mathfrak{p}/(x) \leq \operatorname{ht} \mathfrak{p} = m$. Therefore, (1) follows. Suppose that the set $\operatorname{Min}(R) \cap V(x)$ is empty. Then $\operatorname{ht}_{R/(x)} \mathfrak{p}/(x) < \operatorname{ht} \mathfrak{p}$, so (2) follows. Finally, as $\bigcup_{\operatorname{Min}(R)} \mathfrak{p} \subset \operatorname{zd}(R)$ by Proposition 94.15, we have (2) implies (3). $\square$

Remarks 97.26. Let R be a Noetherian ring.

1. Let $\mathfrak{A} = (a_1, \dots, a_n) < R$ be of height n. Then there exist a minimal prime ideal $\mathfrak{p} \in V(\mathfrak{A})$ satisfying $\operatorname{ht} \mathfrak{p} = \operatorname{ht} \mathfrak{A} = n$ and $\mathfrak{A}$ cannot be generated by a fewer number of elements.
2. It is possible for an ideal $\mathfrak{A} < R$ to have $\mathfrak{p}_1, \mathfrak{p}_2 \in V(\mathfrak{A})$ both minimal but of different heights. For example, let $R = F[X, Y]$ with F a field and X, Y indeterminants. Set $\mathfrak{A} = \big(X(X-1), XY\big)$. Then

$$\mathfrak{A} = (X) \cup (X - 1, Y), \ (X), (X - 1, Y) \in V(\mathfrak{A})$$

where (X) and $(X-1, Y)$ are minimal primes of $\mathfrak{A}$ of heights 1 and 2 respectively, and $V(\mathfrak{A}) = V(X) \cup V(X-1, Y)$, i.e., $V(\mathfrak{A})$ is the union of the Y-axis and the point $(1, 0)$ in F^2.

The next result can be viewed as a converse to the Principal Ideal Theorem.

Corollary 97.27. *Let R be a Noetherian ring, $\mathfrak{A} < R$ an ideal, and $\mathfrak{p} < R$ a prime ideal.*

(1) *If $\operatorname{ht}\mathfrak{A} = n \geq 1$, then there exist $a_1, \ldots, a_n \in \mathfrak{A}$ satisfying $\operatorname{ht}(a_1, \ldots, a_i) = i$ for $i = 1, \ldots, n$.*
(2) *If $\operatorname{ht}\mathfrak{p} = n \geq 1$, then there exist $a_1, \ldots, a_n \in \mathfrak{p}$ satisfying $\mathfrak{p} \in V(a_1, \ldots, a_n)$ is minimal and $\operatorname{ht}(a_1, \ldots, a_i) = i$ for $i = 1, \ldots, n$.*

Proof. (1): As $\operatorname{ht}\mathfrak{A} \geq 1$ and $\operatorname{Min}(R)$ is finite, there exists an element $a_1 \in \mathfrak{A} \setminus \bigcup_{\operatorname{Min}(R)} \mathfrak{p}$ by the Prime Avoidance Lemma. Therefore, $\mathfrak{p} \notin V(a_1)$ for all primes $\mathfrak{p}$ in $\operatorname{Min}(R)$, so $\operatorname{ht}(a_1) > 0$. By the Principal Ideal Theorem, $\operatorname{ht}(a_1) \leq 1$, hence $\operatorname{ht}(a_1) = 1$. By induction, there exists $a_1, \ldots, a_i$ in $\mathfrak{A}$ satisfying $\operatorname{ht}(a_1, \ldots, a_i) = i < n$. Let $\mathfrak{p} \in V(a_1, \ldots, a_i)$ be minimal. By the Principal Ideal Theorem, $\operatorname{ht}\mathfrak{p} \leq i$ and hence $\operatorname{ht}\mathfrak{p} = i$. As $\{\mathfrak{P} \mid \mathfrak{P} \in V(a_1, \ldots, a_i) \text{ minimal}\}$ is a finite set, there exists an element

$$a_{i+1} \in \mathfrak{A} \setminus \bigcup_{\substack{\mathfrak{P} \in V(a_1, \ldots a_i) \\ \text{minimal}}} \mathfrak{P}$$

by the Prime Avoidance Lemma. Let $\mathfrak{B} = (a_1, \ldots, a_{i+1})$ and $\mathfrak{Q} \in V(\mathfrak{B}) \subset V(a_1, \ldots, a_i)$. Then there exists a prime ideal $\mathfrak{P}$ in $V(a_1, \ldots, a_i)$ minimal satisfying $\mathfrak{P} \subset \mathfrak{Q}$. Consequently, $\operatorname{ht}\mathfrak{Q} > \operatorname{ht}\mathfrak{P} = i$. By the Principal Ideal Theorem, $\operatorname{ht}\mathfrak{B} \leq i+1$, hence $\operatorname{ht}\mathfrak{B} = i+1$ and (1) follows by induction.

(2): Choose a_i as in (1) applied to $\mathfrak{A} = \mathfrak{p}$. Since $\mathfrak{p} \in V(A_1, \ldots, a_n)$ satisfies $\operatorname{ht}\mathfrak{p} = \operatorname{ht}(a_1, \ldots, a_n)$, we have $\mathfrak{p} \in V(a_1, \ldots, a_n)$ is minimal. □

Corollary 97.28. *Let R be an affine F-domain of dimension n and $\mathfrak{P}$ a prime ideal in R of height $m \geq 1$. Then there exist $a_1, \ldots, a_m$ in R such that $\mathfrak{P}$ is minimal over $(a_1, \ldots, a_m)$. In particular, if $\mathfrak{p} \in V(a_1, \ldots, a_m)$ is minimal, then $\dim R/\mathfrak{p} \geq n - m$ with equality if $\mathfrak{p} = (a_1, \ldots, a_m)$.*

Proof. The result follows first applying Theorem 97.16 then by localizing at $\mathfrak{p}$ and applying Corollary 97.27. □

The last corollary implies if F a field over an algebraically closed field (respectively any field) and Z is a irreducible affine variety (respectively, abstract irreducible variety) of codimension r in F^n, then Z is an irreducible component of an affine variety (respectively abstract variety) defined as the intersection of r hypersurfaces in F^n.

Exercises 97.29.

1. Prove the following generalization of the Normalization Theorem 97.1: Let F be a field and A an affine F-algebra (not necessarily a domain). If

$$\mathfrak{A}_1 < \cdots < \mathfrak{A}_m < A$$

is a chain of ideals in A, then there exists a non-negative integer n and elements $y_1, \ldots, y_n$ in A algebraically independent over F and integers

$$0 \leq h_1 \leq \cdots \leq h_m \leq n$$

satisfying the following:

(a) $F[y_1, \dots, y_n]$ is integral in A.
(b) A is a finitely generated $F[y_1, \dots, y_n]$-module.
(c) $\mathfrak{A}_i \cap F[y_1, \dots, y_n] = (y_1, \dots, y_{h_i})$ for $i = 1, \dots m$.

2. Prove Lemma 97.17.
3. Let R be a catenary ring and S a multiplicative set in R. Then $S^{-1}R$ is catenary. [Note. If A is an affine F-algebra, this says that localizations of it are catenary. Such a localization may not be an affine F-algebra.]
4. Let R be a Noetherian ring of dimension greater than one. Prove that $\operatorname{Spec}(R)$ is infinite.

98. Regular Local Rings

In geometry, the notion of a simple (or non-singular) point of an algebraic variety is a fundamental notion. It depends on the Jacobian matrix at the point being of the maximal rank (cf. Remark 98.31 below). Indeed one of the interesting aspects of algebraic geometry is that all points need not be simple. Unfortunately, the algebraic analogue does not have all the properties that such geometric points have. We shall study the analogue in this section. In geometry, if a point is simple then the ring of germs at the point (i.e., the functions defined in a neighborhood of the point with two such being identified if they agree in some neighborhood of the point) is a UFD. [If all points are simple, this leads to the theory of Weil divisors.] This is also true of the algebraic analogue. Unfortunately, the general proof depends on developing sufficient homological algebra, which we do not do. Instead we shall only prove the algebraic case corresponding to the geometric case.

Recall if R is a local ring with maximal ideal $\mathfrak{m}$, we write this as $(R, \mathfrak{m})$ is a local ring.

Definition 98.1. Let $(R, \mathfrak{m})$ be a Noetherian local ring of dimension n. By Corollary 97.27, there exist elements $x_1, \dots, x_n$ in $\mathfrak{m}$ satisfying $\operatorname{ht}(x_1, \dots, x_n) = \operatorname{ht} \mathfrak{m}$. Hence $\mathfrak{m} \in V(x_1, \dots, x_n)$ is minimal. We say that the elements $x_1, \dots, x_n$ form a *system of parameters* for R. As $V(x_1, \dots, x_n) = \{\mathfrak{m}\}$, the ideal $\mathfrak{m}/(x_1, \dots, x_n)$ in $R/(x_1, \dots, x_n)$ is nilpotent. Therefore, we have $\mathfrak{m}^k \subset (x_1, \dots, x_n) \subset \mathfrak{m}$ for some integer k, since $\mathfrak{m}$ is finitely generated. If there exists a system of parameters that generate $\mathfrak{m}$, such a system of parameters is called a *regular system of parameters* for R and R is called a *regular local ring.*.

Define

$$\operatorname{V-dim} R := \dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2.$$

By Corollary 93.12 of Nakayama's Lemma, a minimal generating set for $\mathfrak{m}$ is precisely one whose images give a basis for the $R/\mathfrak{m}$-vector space $\mathfrak{m}/\mathfrak{m}^2$. Therefore, $\operatorname{V-dim} R$ is the size of a minimal generating set for $\mathfrak{m}$. By the Principal Ideal Theorem 97.19,

$$\dim R = \operatorname{ht} \mathfrak{m} \leq \operatorname{V-dim} R.$$

Therefore, R is a regular local ring if and only if a minimal generating set for $\mathfrak{m}$ consists of $\dim R$ elements if and only if there exists a regular system of parameters for $\mathfrak{m}$. Regular local rings will be the analogue of a simple point (more exactly, the germs of functions at a simple point).

Example 98.2. 1. Let F be a field and $R = F[[t_1, \dots, t_n]]$, the formal power series in $t_1, \dots, t_n$. Then R is a local ring with maximal ideal $\mathfrak{m} = (t_1, \dots, t_n)$ by Exercise 37.11(2) and is Noetherian by Example 30.13(2). Therefore, it is a regular local ring by Exercise 37.11(2). By Theorem 37.9, the regular local ring $R = F[[t_1, \dots, t_n]]$ is a UFD.

We shall use this to prove that a regular local ring arising from a simple point of an affine variety in the geometric case is a UFD. More specifically, we shall show if $(R, \mathfrak{m})$ is a regular local ring of dimension n containing a copy of the field $F = R/\mathfrak{m}$, then R embeds into $F[[t_1, \dots, t_n]]$. Using the fact that $F[[t_1, \dots, t_n]]$ is a UFD, we shall show that R is also a UFD.

2. Let K be field and $R = K[t]/(t^2)$. R is called the *ring of dual numbers over* K. The ring R is the image of the ring epimorphism $\overline{} : K[t] \to R$ by $t^2 \mapsto 0$. The ideal $\mathfrak{m} = R\bar{t}$ is the unique prime ideal in R. Therefore, R is a Noetherian local ring with $\dim R = 0$ and $K = R/\mathfrak{m}$. As $\mathfrak{m}^2 = 0$ and $\mathfrak{m} = R\bar{t}$ is one dimensional as a K-vector space, R is not a regular local ring. Note that R is not a domain.

Our first goal is to show that a regular local ring is a domain. To do so we need another (important) theorem of Krull. [This is another form of the Krull Intersection Theorem, Exercise 40.12(5).] Recall that a nonzero polynomial $f \in R[t_1, \dots, t_n]$ over a commutative ring R is called *homogeneous* if every monomial in f has the same total degree. We say it is homogeneous of *degree* d if every monomial in F is of total degree d, e.g., $t_1^2 t_3^3 t_4 + t_2^3 t_4^3$ is homogeneous of degree 6. [We also will say the zero polynomial is homogeneous.]

Theorem 98.3. (Krull) *Let R be a Noetherian ring and $\mathfrak{A}$ an ideal in R. Suppose that $1 + \mathfrak{A}$ consists of nonzero divisors. Then $\bigcap_{k=0}^{\infty} \mathfrak{A}^k = 0$.*

PROOF. Let $\mathfrak{A} = (a_1, \dots, a_n)$ and x an element of $\bigcap_{k=0}^{\infty} \mathfrak{A}^k$. Let $\underline{a} = (a_1, \dots, a_n) \in R^n$ and $e_{\underline{a}} : R[t_1, \dots, t_n] \to R$ the evaluation at $\underline{a}$. Clearly, for each $k > 0$, there exists a homogeneous element f_k in $R[t_1, \dots, t_n]$ of degree k satisfying $x = f_k(\underline{a})$. Let $\mathfrak{B}$ be the ideal in $R[t_1, \dots, t_n]$ generated by the f_k, $k > 0$. Since $R[t_1, \dots, t_n]$ is a Noetherian ring by the Hilbert Basis Theorem 41.1, it satisfies ACC, so there exist $f_1, \dots, f_m$ and $x = f_i(\underline{a})$ for $i = 1, \dots, m$ generating $\mathfrak{B}$. In particular,

$$f_{m+1} = g_1 f_1 + \cdots + g_m f_m,$$

with $g_i \in R[t_1, \dots, t_n]$ homogeneous of degree $m + 1 - i \geq 1$. It follows that $g_i(\underline{a})$ lies in $\mathfrak{A}^{m+1-i} \subset \mathfrak{A}$ for $i = 1, \dots, m$, and

$$x = f_{m+1}(\underline{a}) = \sum_{i=1}^{m} g_i(\underline{a}) f_i(\underline{a}) = (\sum_{i=1}^{m} g_i(\underline{a}))x.$$

Set $y = \sum_{i=1}^{m} g_i(\underline{a})$. Then $(1 - y)x = 0$, and by hypothesis, $x = 0$. $\square$

Corollary 98.4. *Let R be a Noetherian domain and $\mathfrak{A} < R$ an ideal. Then $\bigcap_{k=0}^{\infty} \mathfrak{A}^k = 0$.*

Corollary 98.5. *Let R be a Noetherian ring and $\mathfrak{A} < R$ an ideal lying in the Jacobson radical* $\operatorname{rad}(R)$ *of R. Then for every ideal $\mathfrak{B} \subset R$, we have $\bigcap_{k=0}^{\infty} (\mathfrak{B} + \mathfrak{A}^k) = \mathfrak{B}$.*

Proof. Let $^{-} : R \to R/\mathfrak{B}$ be the canonical ring epimorphism. Then $\overline{\mathfrak{A}}^k = (\mathfrak{A}^k + \mathfrak{B})/\mathfrak{B}$. For all $a \in \mathfrak{A} \subset \operatorname{rad}(R)$, so $1 + a \in R^\times$ by Exercise 28.19(12). Therefore, $1 + \overline{a}$ in a nonzero divisor in $\overline{R}$ for all $a \in \mathfrak{A}$. Hence

$$\bigcap_{k=0}^{\infty} \overline{\mathfrak{A}}^k = \bigcap_{k=0}^{\infty} \Big(\frac{\mathfrak{B} + \mathfrak{A}^k}{\mathfrak{B}}\Big) = \overline{0}.$$

The result follows. □

Corollary 98.6. *Let R be a local Noetherian ring with maximal ideal $\mathfrak{m}$. Then $\bigcap_{k=0}^{\infty} \mathfrak{m}^k = 0$.*

We need the corollary to prove the following lemma:

Lemma 98.7. *Let R be a local Noetherian ring. Suppose that R is not a domain and $\mathfrak{p} = (p)$ is a prime ideal. Then $\mathfrak{p}$ is a minimal prime of R.*

Proof. By Corollary 98.6, we have $\bigcap_{k=0}^{\infty} \mathfrak{p}^k = 0$. As R is not a domain, it suffices to show if $\mathfrak{q}$ is a prime ideal in R satisfying $\mathfrak{q} < \mathfrak{p}$, then $\mathfrak{q} \subset \bigcap_{k=0}^{\infty} \mathfrak{p}^k$. But if $q \in \mathfrak{q}$, then $q = pq_1$ in $\mathfrak{p}$. As $p \notin \mathfrak{q}$, $q_1 \in \mathfrak{q}$. Iterating this shows that $q \in \bigcap_{k=0}^{\infty} \mathfrak{p}^k$. □

Lemma 98.8. *Let R be a regular local ring of dimension n with maximal ideal $\mathfrak{m}$. Let $x \in \mathfrak{m} \setminus \mathfrak{m}^2$. Then $R/(x)$ is a regular local ring of dimension $n - 1$.*

Proof. Let $^{-} : R \to R/(x)$ be the canonical epimorphism. Let $x_1, \dots, x_m$ lie in $\mathfrak{m}$ be chosen such that $\overline{x_1}, \dots, \overline{x_m}$ is a minimal generating set for $\overline{\mathfrak{m}}$.

We show that $x, x_1, \dots, x_m$ is a minimal generating set for $\mathfrak{m}$. Certainly they generate, so suppose that we have a R-linear combination $ax + r_1x_1 + \cdots + r_mx_m$ that lies in $\mathfrak{m}^2$. We must show each coefficient lies in $\mathfrak{m}$. As $\overline{r_1}\,\overline{x_1} + \cdots + \overline{r_m}\,\overline{x_m}$ lies in $\overline{\mathfrak{m}}^2$, each $\overline{r_i}$ lies in $\overline{\mathfrak{m}}$ by the minimality of $x_1, \dots, x_m$ (cf. Corollary 93.12 of Nakayama's Lemma). It follows that $r_1, \dots, r_m$ all lie in $\mathfrak{m}$, hence $ax \in \mathfrak{m}^2$. Since $x \notin \mathfrak{m}^2$, $a \in \mathfrak{m}$. Thus $x, x_1, \dots, x_m$ is a minimal generating set. This means that $\operatorname{V-dim} R = m + 1$.

So to finish, we must show that $\operatorname{ht} \overline{\mathfrak{m}} = n - 1$. We know that

$$\dim R/(x) \le (\operatorname{V-dim} R) - 1 = (\dim R) - 1 \le \dim R/(x),$$

and the result follows by Corollary 97.25. □

Corollary 98.9. *Let $(R, \mathfrak{m})$ be a Noetherian local ring with $x \in \mathfrak{m} \setminus \mathfrak{m}^2$ and lying in no minimal prime ideal of R. Then R is a regular local ring if and only if $R/(x)$ is a regular local ring.*

Proof. By the Lemma, we need only show if $R/(x)$ is regular, then R is. But this follows from Corollary 97.25. □

We can now attain our goal.

Theorem 98.10. *Let R be a regular local ring. Then R is a domain.*

Proof. Let R be of dimension n. If $n = 0$, then R is a field so a domain, so we may assume that $n > 0$. Let $\mathfrak{m}$ be the maximal ideal of R and nonzero. By Nakayama's Lemma, $\mathfrak{m} \neq \mathfrak{m}^2$. Let $x \in \mathfrak{m} \setminus \mathfrak{m}^2$. Then $R/(x)$ is a regular local ring of dimension $n - 1$

by Lemma 98.8, so a domain by induction. Suppose that R is not a domain. Then by Lemma 98.7, (x) is a minimal prime ideal. Since this is true for all $x \in \mathfrak{m} \setminus \mathfrak{m}^2$, we have $\mathfrak{m} \setminus \mathfrak{m}^2 \subset \bigcup_{\mathrm{Min}(R)} \mathfrak{p}$. Since $\mathrm{Min}(R)$ is a finite set as R is Noetherian, $\mathfrak{m} \subset \mathfrak{p}$, for some $\mathfrak{p} \in \mathrm{Min}(R)$ by the Prime Avoidence Lemma 93.16. But this means R is of dimension 0, a contradiction. $\square$

Corollary 98.11. *Let R be a regular local ring and $x \in \mathfrak{m} \setminus \mathfrak{m}^2$. Then (x) is a prime ideal in R. In particular, if $x_1, \dots, x_n$ is a regular system of parameters for R, then $R/(x_1, \dots, x_i)$ is a regular local ring of dimension $n - i$ and $(x_1, \dots, x_i)$ is a prime ideal of height i for $i = 0, \dots, n$.*

Remark 98.12. It is a fact that if R is a regular local ring and $\mathfrak{p}$ a prime ideal in R that $R_{\mathfrak{p}}$ is a regular local ring. The proof uses homological algebra methods used to prove that R is shown to be a UFD for an arbitrary regular local ring.

Remark 98.13. We have seen that R is a regular local ring if and only if its maximal ideal is generated by a system of parameters. We comment on the generalization of this property. We omit proofs. Let R be a Noetherian ring. In 95.6, we have seen that if M is an R-module, then $\mathrm{zd}(M) = \bigcup_{Ass_R(M)} \mathfrak{p}$ and understand the minimal primes in $\mathrm{Ass}_R(M)$. The embedded primes do have an effect. Suppose that R is also a local ring. We constructed a system of parameters $x_1, \dots, x_n$ for R based on the condition that x_i is not in any minimal prime containing $(x_1, \cdots, x_{i-1})$ for $i = 1, \dots, n$. This does not involve all of the associated primes of $(x_1, \dots, x_n)$, i.e., all the zero divisors of $R/(x_1, \dots, x_n)$. It turns out that all the zero divisors, i.e., all elements of the associated primes of $R/(x_1, \dots, x_n)$ are important.

A more general approach would be the following: Let R be a commutative ring and M an R-module. An *R-sequence on M* is an ordered sequence $x_1, \dots, x_n$ of elements in R satisfying:

1. The ideal $(x_1, \dots, x_n)M < M$.
2. The image of x_i in $\mathrm{zd}\big(M/(x_1, \dots, x_{i-1})M\big)$ under the natural epimorphism is not a zero divisor for $i = 1, \dots, n$.

If $M = R$, we call an R-sequence $x_1, \dots, x_n$ on R just an *R-sequence*. The first condition is included so that all the quotients $M/(x_1, \dots, x_i)M$ are not zero. As an example, if R_0 is a commutative ring, then $t_1, \dots, t_n$ is an R-sequence in $R = R_0[t_1, \dots, t_n]$.

In general, if M is an R-module and $x_1, \dots, x_n$ is an R-sequence on M, then a permutation of this sequence may not be an R-sequence on M. If R is a Noetherian ring and M a nonzero finitely generated R-module, then it can be shown that any permutation of $x_1, \dots, x_n$ is still an R-sequence on M if $x_1, \dots, x_n$ lie in the Jacobson radical of R. [In general, the problem will arise if the image of x_i is a zero divisor in $\mathrm{zd}\big(M/(x_1, \dots, x_{i-1})M\big)$ for some i.] In particular, if R is a local Noetherian ring and M a nonzero finitely generated R-module, then any permutation of an R-sequence on M is an R-sequence on M; and it can also be shown that, in this case, all maximal R-sequences on M have the same length. In particular, in this case the maximal length of an R-sequence M is well-defined and called the *depth* of M. For example, if $M = R$ the *depth* of R, written $\operatorname{depth} R$, exists and is the independent of a maximal R-sequence in R. [In general, if R is any Noetherian ring with $\mathfrak{A}$ an ideal in R and M a finitely generated module with $M \neq \mathfrak{A}M$, then one can show that any two maximal R-sequences on M lying in $\mathfrak{A}$ have the same finite length.]

For example, suppose that R is a regular local ring and $x_1, \dots, x_n$ is a regular system of parameters. Then $R/(x_1, \dots, x_i)$ is a domain for $i = 1, \dots, n$. It follows that $x_1, \dots, x_n$ is an R-sequence. In particular, $\operatorname{depth} R = \dim R$. In general, $\operatorname{depth} R \leq \dim R$. A Noetherian local ring is called a *Cohen-Macaulay ring* if $\operatorname{depth} R = \dim R$. [A Noetherian ring is called a *Cohen-Macaulay ring* if $R_{\mathfrak{m}}$ is Cohen-Macaulay ring for every maximal ideal $\mathfrak{m} \in \operatorname{Spec}(R)$.] Being a Cohen-Macaulay ring is weaker than being a regular local ring. In fact, for a local domain to be a Cohen-Macaulay ring is weaker than being a normal domain. For example, the ring $\mathbb{C}[X,Y]/\big(Y - X^2(X+1)\big)$ giving the nodal curve of Remark 97.2 localized at $(X,Y)/(Y - X^2(X+1)$ (the ring at the origin of the curve in $\mathbb{C}^2$) can be shown to be Cohen-Macaulay but not normal, as the images of X, Y is checked to give a regular sequence. One can also show that a Noetherian local ring is a Cohen-Macaulay if and only if every system of parameters is an R-sequence. The notion of being Cohen-Macaulay is quite useful in algebraic geometry as well as in commutative algebra.

We next turn to another useful characterization of a regular local ring. We shall use *dehomogenization* of a homogeneous polynomial, e.g., substitute $t/t, t_1/t, \dots, t_n/t$ for the variables in such an f thereby obtaining a polynomial in n variables. Note if R is a domain, $f \in R[t, t_1, \dots, t_n]$ a homogeneous of degree d, and $f(x, x_1, \dots, x_n) = 0$ with $x, x_1, \dots x_n \in R$ nonzero if and only if $x^N f(1, x_1/x, \dots x_n/x) = 0$ in R for all $N \geq d$. We use this to prove the following lemma:

Lemma 98.14. *Let $(R, \mathfrak{m})$ be a Noetherian local domain of dimension $n+1$. Suppose that $\mathfrak{m} \in V(x, x_1, \dots, x_n)$ is minimal (e.g., R is a regular local ring) and $f \in R[t, t_1, \dots, t_n]$ is a homogeneous polynomial. If $f(x, x_1, \dots, x_n) = 0$, then all coefficients of f lie in $\mathfrak{m}$.*

Proof. Set $\mathfrak{P} = \mathfrak{m} R[t_1, \dots, t_n]$, the extension of $\mathfrak{m}$ to $R[t_1, \dots, t_n]$. We know that $R[t_1, \dots, t_n]/\mathfrak{P} \cong (R/\mathfrak{m})[t_1, \dots, t_n]$, hence it is a domain and $\mathfrak{P}$ is a prime ideal in $R[t_1, \dots, t_n]$ and minimal in $V_{R[t_1,\dots,t_n]}(x, x_1, \dots, x_n)$. In particular, $\operatorname{ht} \mathfrak{P} \leq n+1$ by the Principal Ideal Theorem 97.19. As any proper chain of primes for $\mathfrak{m}$ gives a proper chain of primes for $\mathfrak{P}$, we have $\operatorname{ht} \mathfrak{P} = \operatorname{ht} \mathfrak{m} = \dim R = n+1$.

let $\varphi : R[t_1, \dots, t_n] \to R[x_1/x, \dots, x_n/x]$ be the R-algebra epimorphism determined by $t_i \mapsto x_i/x$, which lies in the quotient field of R. To prove the result, it suffices to show $\ker \varphi \subset \mathfrak{P}$.

Let $\mathfrak{A} = (xt_1 - x_1, \dots, xt_n - x_n) \subset \mathfrak{P}$. Since $\operatorname{ht} \mathfrak{P} = \operatorname{ht} \mathfrak{m}$, $\mathfrak{P} \in V(\mathfrak{A})$ cannot be minimal. Let $\mathfrak{p} \subset \mathfrak{P}$ with $\mathfrak{p} \in V(\mathfrak{A})$ minimal.

We show that $x \notin \mathfrak{p}$. If $x \in \mathfrak{p}$, then $x, x_1, \dots, x_n \in \mathfrak{p}$, and we have $(x, x_1, \dots, x_n) \subset \mathfrak{p} \cap R$. Thus $\mathfrak{p} \cap R = \mathfrak{m}$ and $\mathfrak{P} \subset \mathfrak{p}$, a contradiction. Therefore, $x \notin \mathfrak{p}$.

Now suppose that $g \in \ker \varphi$. In particular, if $k >> 0$, then $x^k g(x_1, \dots, x_n) = 0$. Let R_x denote the domain obtained by localizing R at $S = \{x^n \mid n \geq 0\}$. Applying the Remainder Theorem (of the Division Algorithm) to the monic polynomials $t_i - \dfrac{x_i}{x}$ to g sequentially, we get polynomials h_i in $R_x[t_i, \dots, t_n]$, for $i = 1, \dots, n$, and equations in

$R_x[t_1, \dots, t_n]$:

$$g = (t_1 - \frac{x_1}{x})h_1 + g(\frac{x_1}{x}, t_2, \dots, t_n)$$
$$\vdots$$
$$g(\frac{x_1}{x}, \dots, \frac{x_{i-1}}{x}, t_i, \dots, t_n) = (t_i - \frac{x_i}{x})h_i + g(\frac{x_1}{x}, \dots, \frac{x_i}{x}, t_{i+1}, \dots, t_n)$$
$$\vdots$$
$$g(\frac{x_1}{x}, \dots, \frac{x_{n-1}}{x}, t_n) = (t_n - \frac{x_n}{x})h_n + g(\frac{x_1}{x}, \dots, \frac{x_n}{x}).$$

Lifting these equations back to $R[t_1, \dots, t_n]$, i.e., clearing all denominators of all these polynomials, we see that there exits $k > 0$ such that $x^k g \in \mathfrak{P}$. Therefore, $\ker \varphi \subset \mathfrak{P}$ as needed. □

Definition 98.15. A commutative ring R will be called a *graded ring* (on index set $\mathbb{N} = \{i \in \mathbb{Z} \mid i \geq 0\}$) if $R = \bigoplus_{i=0}^{\infty} R_i$ as additive groups with multiplication satisfying $R_i R_j \subset R_{i+j}$ for all $i, j \geq 0$. (Note in this case, R_0 is a commutative ring.) An element r in R_d is called a *homogeneous element of degree* d. A *homomorphism of graded rings* is a ring homomorphism $\varphi : \bigoplus_{i\geq 0} R_i \to \bigoplus_{i\geq 0} S_i$ that preserves degree, i.e., $\varphi(R_i) \subset S_i$ for all $i \geq 0$ and, a *graded ring isomorphism* if, in addition, φ is also a ring isomorphism.

Examples 98.16. Let R_0 be a commutative ring.

1. The polynomial ring $R = R_0[t_1, \dots, t_n]$ is a graded ring with R_d the group generated by the monomials of total degree d.
2. $R = R_0[[t_1, \dots, t_n]]$ is a graded ring with R_d the group generated by the monomials of total degree d. Note that the natural map $R_0[t_1, \dots, t_n] \to R_0[[t_1, \dots, t_n]]$ induced by $t_i \mapsto t_i$ is an injective graded ring homomorphism.
3. Let R be a commutative ring and $\mathfrak{A}$ an ideal in R. The graded ring of R determined by $\mathfrak{A}$, called the *associated graded ring relative to* $\mathfrak{A}$, is defined to be
$$G_{\mathfrak{A}}(R) = \bigoplus_{i=0}^{\infty} \mathfrak{A}^i/\mathfrak{A}^{i+1}$$
where $\mathfrak{A}^0 = R$ and the multiplication is induced by
$$(x + \mathfrak{A}^{i+1})(y + \mathfrak{A}^{j+1}) := xy + \mathfrak{A}^{i+j+1} \quad \text{for all} \quad x \in \mathfrak{A}^i,\, y \in \mathfrak{A}^j.$$
The elements in $\mathfrak{A}^d/\mathfrak{A}^{d+1}$ are the homogeneous elements of $G_{\mathfrak{A}}(R)$ of degree d.

Using Lemma 98.14, we have the following characterization of regular local rings.

Theorem 98.17. *Let $(R, \mathfrak{m})$ be a Noetherian local ring of dimension n. Then the following are equivalent.*

(1) The ring $(R, \mathfrak{m})$ is a regular local ring.
(2) The rings $(R/\mathfrak{m})[t_1, \dots, t_n]$ and $G_{\mathfrak{m}}(R)$ are isomorphic graded rings.
(3) The vector space $\mathfrak{m}/\mathfrak{m}^2$ over the field $R/\mathfrak{m}$ is n-dimensional.

PROOF. (1) $\Rightarrow$ (2): As R is a regular local ring, $\mathfrak{m} = (x_1, \dots, x_n)$ for some $x_1, \dots, x_n$. The canonical ring epimorphism $^- : R \to R/\mathfrak{m}$ induces the natural graded ring epimorphism $R[t_1, \dots, t_n] \to (R/\mathfrak{m})[t_1, \dots, t_n]$ and the graded ring epimorphism $\varphi : (R/\mathfrak{m})[t_1, \dots, t_n] \to G_{\mathfrak{m}}(R)$ induced by $t_i \mapsto x_i + \mathfrak{m}^2$. Then φ is injective by Lemma 98.14.

(2) $\Rightarrow$ (3): This is clear.

(3) $\Rightarrow$ (1): By (3), a minimal generating set for $\mathfrak{m}$ consists of $n = \dim R$ elements, so is a regular system of parameters. Therefore, R is regular.

□

We next turn to our main goal, to show that a regular local ring in the geometric case is a UFD. The algebraic condition we need is the following:

Definition 98.18. Let $(R, \mathfrak{m})$ be a Noetherian local ring. We say that R is *equicharacteristic* if R contains an isomorphic copy of the field $R/\mathfrak{m}$. For convenience of notation, we shall assume that $F = R/\mathfrak{m}$ lies in R if this is the case.

We wish to show that under the hypothesis of an equicharacteristic Noetherian local ring, we can assign a power series to each element of R. In general, such an assignment will not be unique. However, for a regular local ring, it shall be as we shall see.

The idea is the following: We have defined a graded ring isomorphism $G_{\mathfrak{m}}(R) \to (R/\mathfrak{m})[t_1, \dots, t_n]$ extending the canonical map $R \to R/\mathfrak{m}$ and sending $x_i + \mathfrak{m}^2 \mapsto t_i$. In particular, this defines an $R/\mathfrak{m}$-isomorphism $\mathfrak{m}^d/\mathfrak{m}^{d+1} \to (R/\mathfrak{m})[t_1, \dots, t_n]_d$. Since $F \subset R$, this map fixes F. We also have the natural graded ring monomorphism $(R/\mathfrak{m})[t_1, \dots, t_n] \to F[[t_1, \dots, t_n]]$. As power series can be truncated to give polynomials of various degrees, we wish to mimic the construction of Taylor series as in calculus. We then will show that the construction is unique when R is a regular local ring. For example, if R is of dimension one and $x \in \mathfrak{m}$ then we would expect that the inverse of the unit $1 - x$ should be assigned the power series $\sum_{i=0}^{\infty} t^i$ corresponding to $(1, 1 + \mathfrak{m}, 1 + \mathfrak{m}^2, \dots)$ with truncations lying in $G_{\mathfrak{m}}(R)$. We first do the construction.

Construction 98.19. Let $(R, \mathfrak{m})$ be an equicharacteristic local Noetherian ring with $F = R/\mathfrak{m}$ and $^- : R \to F$ the canonical epimorphism. We assume that F also lies in R and $^-$ fixes F. Suppose that $\mathfrak{m} = (u_1, \dots, u_m)$.

Let $x \in R$. Define a power series in $F[[t_1, \dots, t_m]]$ associated to x as follows. Write

$$x = \alpha_0 + x_1, \quad \text{with} \quad \alpha_0 \in F,\ x_1 \in \mathfrak{m} \quad \text{and set} \quad \overline{x} = \alpha_0.$$

Then we have

$$x_1 = \sum_{i=1}^{m} \alpha_i u_i + x_2 \quad \text{with} \quad \alpha_i \in F,\ x_2 \in \mathfrak{m}^2.$$

Similarly, define

$$x_2 = \sum_{i,j=1}^{m} \alpha_{ij} u_i u_j + x_3 \text{ with } \alpha_{ij} \in F,\ x_3 \in \mathfrak{m}^3.$$

Continuing in this way we see that for each r, we can find homogeneous polynomials $f_i \in F[t_1, \ldots, t_m]$ of degree i, $i = 0, \ldots, r$, satisfying

$$x = \sum_{i=0}^{r} f_i(u_1, \ldots, u_m) + x_{r+1} \quad \text{with} \quad x_{r+1} \in \mathfrak{m}^{r+1}.$$

Thus we can assign to x a formal power series

$$\Phi(x) = f_0 + f_2 + f_3 + \cdots \quad \text{in} \quad F[[t_1, \ldots, t_m]].$$

Definition 98.20. Let $(R, \mathfrak{m})$ be an equicharacteristic local Noetherian ring with maximal ideal $\mathfrak{m} = (u_1, \ldots, u_m)$. If $x \in R$, then a formal power series $\Phi(x) = f_0 + f_1 + \cdots \in F[[t_1, \ldots, t_m]]$ is called a *(formal) Taylor series* of x *relative to* $u_1, \ldots, u_m$ if for all $r \geq 0$,

$$x - \sum_{i=0}^{r} f_i(u_1, \ldots, u_m) \quad \text{lies in} \quad \mathfrak{m}^{r+1}.$$

Example 98.21. Let F be a field and $R = F[t]_{(t)}$, so $\mathfrak{m} = (t)$. Let $f = p/q$ with $p, q \in F[t]$ and $q(0) \neq 0$. Then $f \in R$. The Taylor series relative to t is a formal power series $\sum_{i=0}^{\infty} \alpha_i t^i$ such that $(p/q) - \sum_{i=0}^{r} \alpha_i t^i \equiv 0 \mod t^{r+1}$ for all $r \geq 0$. For example, $1/(1-t)$ has a Taylor series $\sum_{i=0}^{\infty} t^i$, since

$$\frac{1}{1-t} - \sum_{i=0}^{r} t^i = \frac{t^{r+1}}{1-t} \equiv 0 \mod t^{r+1} \quad \text{for all } r \geq 0.$$

Proposition 98.22. *Let $(R, \mathfrak{m})$ be an equicharacteristic regular local ring with $x_1, \ldots, x_n$ a regular system of parameters. Then for all x in R, there exist a unique Taylor series for x relative to $x_1, \ldots, x_n$.*

PROOF. If $\Phi(x) = f_0 + f_1 + \cdots$ and $\Phi'(x) = f_0' + f_1' + \cdots$ are two Tayler series for x, $x \in R$, then 0 is a Taylor series for $\Phi(x) - \Phi'(x)$. So it suffices to show if 0 has a Taylor series $\Phi = f_0 + f_1 + \cdots$, then $\Phi = 0$.

Suppose that f_r is the first nonzero homogenous term in Φ. Then $f_r(x_1, \ldots, x_n)$ lies in $\mathfrak{m}^{r+1}$. As a regular local ring is a domain, all coefficients of f_r lie in $\mathfrak{m}$ by Lemma 98.14, so $f_r(x_1, \ldots, x_n) = 0$. □

The proposition implies:

Corollary 98.23. *Let $(R, \mathfrak{m})$ be an equicharacteristic regular local ring of dimension n with $F = R/\mathfrak{m}$. If $\mathfrak{m} = (x_1, \ldots, x_n)$, then there exists an F-algebra monomorphism $R \to F[[t_1, \ldots, t_n]]$.*

We now turn to proving that an equicharacteristic regular local ring is a UFD, We need some preliminaries. It is useful to set up the following notation.

Notation 98.24. Let $(R, \mathfrak{m})$ be an equicharacteristic regular local ring of dimension n with $F = R/\mathfrak{m}$. As before we view $F \subset R$ and as constructed before, we have an F-algebra monomorphism $R \to F[[t_, \ldots, t_n]]$ by $x \mapsto \Phi(x)$. We shall view this map as an inclusion. In particular, we shall identify

$$x \in R \quad \text{with its Taylor series, } \Phi(x) = \sum_{i=0}^{\infty} f_i \quad \text{in} \quad F[[t_1, \ldots, t_n]],$$

where $f_i \in F[t_1,\dots,t_n]]$ is homogeneous of degree i relative to $t_1,\dots,t_n$. We set $\widehat{R} = F[[t_1,\dots,t_n]]$ and if $\mathfrak{m} = (x_1.\dots,x_n)$, we identify x_i in R with t_i in $\widehat{R}$. We shall also write $\underline{x} = (x_1,\dots,x_n)$ when viewed as a point in $(R/\mathfrak{m})^n = F^n$.
We know that $\widehat{R}$ is a regular local ring with maximal ideal $\widehat{\mathfrak{m}} = (t_1,\dots,t_n)$ and R and $\widehat{R}$ are domains.

We begin by looking at a relationship between ideals in R and $\widehat{R}$.

Lemma 98.25. *In the Notation 98.24, for every $i \geq 1$, we have $\widehat{\mathfrak{m}}^i \cap R = \mathfrak{m}^i$.*

PROOF. Certainly, for every $i \geq 1$, we have $\mathfrak{m}^i \subset \widehat{\mathfrak{m}}^i \cap R$, so we need only show that $\widehat{\mathfrak{m}}^i \cap R \subset \mathfrak{m}^i$. Let $x \in R$ have Taylor series relative to $x_1,\dots,x_n$ given by $\Phi(x) = \sum_{i=0}^\infty f_i$ with each $f_j = 0$ for $j < i$. Then by definition,

$$x - \sum_{j=0}^{i-1} f_j(\underline{x}) \equiv 0 \mod \mathfrak{m}^i.$$

as needed. □

Lemma 98.26. *In the Notation 98.24, if $\mathfrak{A} \subset R$ is an ideal, then we have $\widehat{R}\mathfrak{A} \cap R = \mathfrak{A}$.*

PROOF. Again it suffices to show $\widehat{R}\mathfrak{A} \cap R \subset \mathfrak{A}$. As $\mathfrak{A}$ is Noetherian, $\mathfrak{A} = (a_1,\dots,a_m)$ for some $a_1,\dots,a_m \in \mathfrak{A}$. Let $x \in \widehat{R}\mathfrak{A} \cap R$. We can write

$$x = g^{(1)}(\underline{x})a_1 + \cdots + g^{(m)}(\underline{x})a_m \quad \text{with} \quad g^{(i)} \in \widehat{R}$$

and

$$g^{(i)} = g_0^{(i)} + \cdots + g_n^{(i)} + \cdots \quad \text{the corresponding power series.}$$

Each $g_n^{(i)}(\underline{x})$ can be approximated up to an element in $\widehat{m}^{n+1}$ for each n, so we can find $h_n^{(i)}$ in $F[t_1,\dots,t_m]$ with $h_n^{(i)}(\underline{x}) \in R$ and

$$h_n^{(i)}(\underline{x}) = g_0^{(i)}(\underline{x}) + \cdots + g_n^{(i)}(\underline{x}) \mod \widehat{\mathfrak{m}}^{n+1}.$$

Therefore, we have

$$\begin{aligned}\sum_{i=1}^m g^{(i)}(\underline{x})a_i &= \sum_{i=1}^m \big(h_n^{(i)}(\underline{x}) + g^{(i)}(\underline{x}) - h_n^{(i)}(\underline{x}))\big)a_i \\ &= \sum_{i=1}^m h_n^{(i)}(\underline{x})a_i + \sum_{i=1}^m \big(g^{(i)}(\underline{x}) - h_n^{(i)}(\underline{x})\big)a_i\end{aligned}$$

with $\sum_{i=1}^m \big(g^{(i)}(\underline{x}) - h^{(i)}(\underline{x})\big)a_i$ lying in $\widehat{\mathfrak{m}}^{n+1} \cap R = \mathfrak{m}^{n+1}$ by the last lemma. Consequently, $\sum_{i=1}^m g^{(i)}(\underline{x})a_i$ lies in $\mathfrak{A} + \mathfrak{m}^{n+1}$ for every $n \geq 1$. Since $\bigcap_{i=0}^\infty(\mathfrak{A} + \mathfrak{m}^n) = \mathfrak{A}$ by Corollary 98.5, the lemma is proven. □

To prove that an equicharacteristic regular local ring R is a UFD, we only need to show that every irreducible element in R is a prime element, as R is Noetherian. We want to use the fact that $\widehat{R}$ is a UFD that we showed in Theorem 37.9. To do so we must analyze division in R and how it relates to division in $\widehat{R}$. But Lemma 98.26 above has this as a consequence.

Corollary 98.27. *In the Notation 98.24, suppose that $a, b \in R$ satisfies $b \mid a$ in $\widehat{R}$. Then $b \mid a$ in R.*

Proof. If $b \mid a$ in $\widehat{R}$, i.e., $\widehat{R}\,a \subset \widehat{R}\,b$, then $\widehat{R}\,a \cap R \subset \widehat{R}\,b \cap R$. Hence $Ra \subset Rb$ by Lemma 98.26, i.e., $b \mid a$ in R. □

We use this to see that if x, y in our equicharacteristic regular local ring R are non-associates in $\widehat{R}$, then they are non-associates in R. This is the key to what we need to finish our goal that a equicharacteristic regular local ring is a UFD.

Proposition 98.28. *In the Notation 98.24, suppose that $x, y \in R$ have a nonunit factor in $\widehat{R}$. Then x and y have a nonunit factor in R.*

Proof. Since $\widehat{R}$ is a UFD, x and y have a greatest common divisor $\widehat{d}$ in $\widehat{R}$ with $\widehat{d} \notin \widehat{R}^{\times}$. Let

$$x = \widehat{d}\,\widehat{a} \quad\text{and}\quad y = \widehat{d}\,\widehat{b} \quad\text{with}\quad \widehat{a}, \widehat{b} \in \widehat{R} \quad\text{relatively prime.}$$

We can approximate

$$\begin{aligned} &\widehat{a} \quad\text{by}\quad \{a_n\} \subset R \quad\text{with}\quad \widehat{a} \equiv a_n \mod \widehat{\mathfrak{m}}^{n+1} \\ &\widehat{b} \quad\text{by}\quad \{b_n\} \subset R \quad\text{with}\quad \widehat{b} \equiv b_n \mod \widehat{\mathfrak{m}}^{n+1} \end{aligned}$$

for every $n \geq 1$. So we have

$$0 = x\widehat{b} - y\widehat{a} = x(\widehat{b} - b_n + b_n) - y(\widehat{a} - a_n + a_n),$$

hence

$$(*) \qquad xb_n - ya_n = y(\widehat{a} - a_n) - x(\widehat{b} - b_n).$$

Let $\mathfrak{A} = x\,\mathfrak{m}^{n+1} + y\,\mathfrak{m}^{n+1}$. The right hand side of $(*)$ lies in $\widehat{R}\mathfrak{A} \cap R = \mathfrak{A}$, using Lemma 98.26. Therefore, we have an equation

$$xb_n - ya_n = -xr_n + ys_n \quad\text{with}\quad r_n, s_n \in \mathfrak{m}^{n+1}.$$

Consequently,

$$x(b_n + r_n) = y(a_n + s_n) \quad\text{in}\quad R.$$

Therefore we see that in $\widehat{R}$, after cancelling $\widehat{d}$, we have

$$(**) \qquad \widehat{a}(b_n + r_n) = \widehat{b}(a_n + s_n) \quad\text{in}\quad \widehat{R}.$$

Since $\widehat{a}$ and $\widehat{b}$ are relatively prime in the UFD $\widehat{R}$, we must have $\widehat{a}$ divides every $a_n + s_n$, i.e.,

$$(\dagger) \qquad a_n + s_n = \widehat{a}\,\widehat{u} \quad\text{some}\quad \widehat{u} \in \widehat{R}.$$

Reading the equation $(\dagger)$ modulo $\widehat{\mathfrak{m}}^{n+1}$, we have $\widehat{a} \equiv \widehat{a}\,\widehat{u} \mod \widehat{\mathfrak{m}}^{n+1}$, since $s_n \in \mathfrak{m}^{n+1}$. We cannot have $\widehat{u} \in \widehat{\mathfrak{m}}$, lest $1 - \widehat{u} \in \widehat{R}^{\times}$. Therefore, $\widehat{u} \in \widehat{R}^{\times}$. Fix $n >> 0$ satisfying $\widehat{a} \notin \widehat{\mathfrak{m}}^{n+1}$. Then $a_n + s_n$ is nonzero in the domain $\widehat{R}$, hence equation $(\dagger)$ implies that

$$\widehat{a} = \widehat{u}^{-1}(a_n + s_n) \quad\text{and}\quad \widehat{b} = \widehat{u}^{-1}(b_n + r_n).$$

Therefore,

$$x = \widehat{d}\,\widehat{a} = (a_n + s_n)\widehat{u}^{-1}\widehat{d} \quad\text{and}\quad y = \widehat{d}\,\widehat{b} = (b_n + r_n)\widehat{u}^{-1}\widehat{d}.$$

Since $(a_n + s_n) \mid x$ in $\widehat{R}$, we conclude that $(a_n + s_n) \mid x$ in R by Corollary 98.27. It follows by $(**)$ that $d := \widehat{u}^{-1}\widehat{d}$ in R and so is a common factor of x and y in R. If $d \in R^\times$, then $\widehat{d} = \widehat{u}d \in \widehat{R}^\times$, which is impossible. This proves the proposition. $\square$

Theorem 98.29. *Let $(R, \mathfrak{m})$ be an equicharacteristic regular local ring. Then R is a UFD.*

PROOF. As R is a Noetherian domain, by Euclid's Argument (Proposition 30.9), it suffices to show if $x \in R$ is irreducible, then it is a prime element. So assume that $x \mid yz$ with $y, z \in R$ and x irreducible in R.

Case 1. Suppose that x and y have a nonunit common factor in $\widehat{R}$:

By Proposition 98.28 above, x and y have a nonunit common factor in R. As x is irreducible in R, we have $x \mid y$ in R.

Case 2. Suppose that x and y have no nonunit common factor in $\widehat{R}$:

Since $\widehat{R}$ is a UFD and x and y are relatively prime in $\widehat{R}$, we know that $x \mid z$ in $\widehat{R}$. Therefore, $x \mid z$ in R by Corollary 98.27. $\square$

As a UFD is normal, we have:

Corollary 98.30. *Let $(R, \mathfrak{m})$ be an equicharacteristic regular local ring. Then R is a normal domain.*

Remark 98.31. Being a regular local ring is much stronger than just being normal. Let F be a field and $f_1, \dots, f_m \in F[t_1, \dots, t_n]$ with $m \geq n$. Suppose that $\widetilde{F}$ is an algebraic closure of F. Set $\widetilde{R} = \widetilde{F}[t_1, \dots, t_n]/(f_1, \dots, f_m)$. A point $\underline{x} \in Z_{\widetilde{F}}(f_1, \dots, f_m)$ is called a *simple point* of $Z_{\widetilde{F}}(f_1, \dots, f_m)$ if the rank of the *Jacobian matrix* of $\underline{x}$, i.e., the $m \times n$ matrix $\big(\partial f_i/\partial t_j(\underline{x})\big)$, is n. Let $\mathfrak{m}_{\underline{x}}$ be the maximal ideal defined by $\underline{x}$. Then it is a fact that $\widetilde{R}_{\mathfrak{m}_{\underline{x}}}$ is a regular local ring if and only if $\underline{x}$ is a simple point. Suppose that F is a perfect field. Then the localization at any maximal ideal $\mathfrak{m}$ of $R = F[t_1, \dots, t_n]/(f_1, \dots, f_m)$ is a regular local ring if $\mathfrak{m} = R \cap \mathfrak{m}_{\underline{x}}$ when $\underline{x}$ is a simple point in $Z_{\widetilde{F}}(f_1, \dots, f_m)$. In general, this is not true for inseparable extensions.

Exercises 98.32.

1. Show if R is a Noetherian ring and $\mathfrak{A}$ is an ideal of R, then $G_{\mathfrak{A}}(R)$ is also Noetherian.
2. Show every discrete valuation ring is a regular local ring.
3. Let R be an equicharacteristic local ring with a, b nonzero elements in R. Let $\widehat{R}$ be as in Notation 98.24. Show that a gcd of a and b in R and in $\widehat{R}$ are associates in $\widehat{R}$.
4. Let R be a commutative ring and $\mathfrak{A} < R$ an ideal. For each pair of non-negative integers $i \leq j$ let $\theta_{i,j} : R/\mathfrak{A}^j \to R/\mathfrak{A}^i$ be the natural ring epimorphism. Show that there exist a ring $\widehat{R}$ and for each i a ring homomorphism $\psi_i : \widehat{R} \to R/\mathfrak{A}^i$ such that for all non-negative $i \leq j$

$$\begin{tikzcd} & R/\mathfrak{A}^i \\ \widehat{R} \arrow[ur, "\psi_i"] \arrow[dr, "\psi_j"'] & \\ & R/\mathfrak{A}^j \arrow[uu, "\theta_{ij}"'] \end{tikzcd}$$

commutes, and if there exist ring homomorphisms $\theta_i : S \to R/\mathfrak{A}^i$ such that for all non-negative $i \leq j$

$$\begin{tikzcd} & R/\mathfrak{A}^i \\ S \arrow[ur, "\varphi_i"] \arrow[dr, "\varphi_j"'] & \\ & R/\mathfrak{A}^j \arrow[uu, "\theta_{ij}"'] \end{tikzcd}$$

commutes, then there exists a unique ring homomorphism $\mu : S \to \widehat{R}$ such that

(98.33)
$$\begin{tikzcd} & & R/\mathfrak{A}^i \\ S \arrow[urr, "\varphi_i"] \arrow[r, "\mu"] \arrow[drr, "\varphi_j"'] & \widehat{R} \arrow[ur, "\psi_i"] \arrow[dr, "\psi_j"'] & \\ & & R/\mathfrak{A}^j \arrow[uu, "\theta_{ij}"'] \end{tikzcd}$$

commutes for all $i \leq j$. Such a $\widehat{R}$ is unique up to a unique isomorphism and is called the *completion* of R *relative to the ideal* $\mathfrak{A}$.

5. Let R be a Noetherian ring, $\mathfrak{A} < R$ an ideal. Let $\widehat{R}$ be the completion of R relative to $\mathfrak{A}$. Show that there exists ring homomorphism $\varphi : R \to \widehat{R}$ such that the diagram in 98.33 is valid with $\varphi_i : R \to R/\mathfrak{A}^i$ the canonical surjections and that this map is a monomorphism if R is either a domain or a local ring.
6. In the notation of Exercise 5, show that $\widehat{R}$ is the subring of $\prod_{i=0}^{\infty} R/\mathfrak{A}^i$ (a ring with componentwise operations) consisting of all sequences
$$(\ldots, x_n \ldots, x_1, x_0 \mid \varphi(x_{n+1} + \mathfrak{A}^{n+1}) = x_n + \mathfrak{A}^n \quad \text{for all} \quad n \geq 1\}$$
(a subring with component-wise operations).
7. Let R be a Noetherian ring, $\mathfrak{A} < R$ an ideal. Define a topology on R, called the $\mathfrak{A}$-*adic topology* by $\{r + \mathfrak{A}^n \mid r \in R,\ n \geq 0\}$ is a base of open sets for the topology of R. Let $\widehat{R}$

be the completion of R relative to $\mathfrak{A}$ and $\widehat{\mathfrak{A}} = \widehat{R}\mathfrak{A}$. Show that every Cauchy sequence in the $\widehat{\mathfrak{A}}$-adic topology of $\widehat{R}$ converges, i.e., $\widehat{R}$ is topologically complete.

8. Let F be a field. Show the completion of $F[t_1, \ldots, t_n]$ is the power series ring $F[[t_1, \ldots, t_n]]$.
9. Let p be a prime in $\mathbb{Z}$, then $\widehat{\mathbb{Z}}$, the completion of $\mathbb{Z}$ relative to the ideal (p) is called the *p-adic integers* [and is historically denoted by $\mathbb{Z}_p$.] Let $\psi_i : \widehat{\mathbb{Z}} \to \mathbb{Z}/(p^i)$ be the canonical surjections for each positive i. Show all of the following:
   (i) $\widehat{\mathbb{Z}}$ is a domain.
   (ii) For each $n > 0$
   $$0 \to \widehat{\mathbb{Z}} \xrightarrow{p^n} \widehat{\mathbb{Z}} \xrightarrow{\psi_n} \mathbb{Z}/(p^n) \to 0.$$
   is an exact sequence of additive groups where the map p^n is multiplication by p^n. In particular, $\widehat{\mathbb{Z}}/p^n\widehat{\mathbb{Z}} \cong \mathbb{Z}/(p^n)$ (and are usually identified).
   (iii) If $x \in \mathbb{Z}/\bigl(p^n\bigr)$, then x unit in $\mathbb{Z}/\bigl(p^n\bigr)$ and only if $p \not| x$.
   (iv) An element $x \in \widehat{\mathbb{Z}}$ is a unit if and only if $p \not| x$.
   (v) Every element in $\widehat{\mathbb{Z}}$ is of the form $p^r u$ for some $u \in \widehat{\mathbb{Z}}^\times$ and $r \geq 0$. In particular, $\widehat{\mathbb{Z}}$ is a regular local ring.
   (vi) Every element in $\widehat{\mathbb{Z}}$ can be uniquely written as a power series $\sum_{i=1}^\infty a_i p^i$ with $a_i \in \{0, 1, \ldots, p-1\}$ for all i.

99. Addendum: Fibers

One of the first memorable theorems in linear algebra is given vector spaces V and W over a field F with V finite dimensional and a linear transformation $T : V \to W$, then $\dim V = \dim \operatorname{im} T + \dim \ker T$. As $\ker T = T^{-1}(0)$, the fiber of the zero of W is a subspace of V and $\operatorname{im} T$ a subspace of W, this makes sense as a theorem about vector spaces. We wish to generalize the analogous result to maps of irreducible abstract F-affine varieties. One problem is that the image of an an affine variety need not be a variety, although its closure is. We write this generalization as a theorem about F-affine domains and afterward translate this to the result that we really want.

Theorem 99.1. *Let $\varphi : A \to B$ be an injective F-algebra homomorphism of F-affine domains A and B of dimensions m and n respectively. Then*

(1) For each $\mathfrak{m} \in \operatorname{Max}(A)$, each irreducible component of $V\bigl(\varphi(\mathfrak{m})\bigr)$ has dimension greater than or equal to $n - m$.

(2) There exists an open set U of $\operatorname{Spec}(A)$ such that for all $\mathfrak{m}$ in $U \cap \operatorname{Max}(A)$, every irreducible component of $V\bigl(\varphi(\mathfrak{m})\bigr)$ is of dimension $n - m$.

Proof. (1): There exist $a_1, \ldots, a_m$ in $\mathfrak{m}$ such that $\mathfrak{m}$ is minimal over $(a_1, \ldots, a_m)$ by Corollary 97.28. Then every irreducible component of the variety $V_B\bigl(\varphi(a_1), \ldots, \varphi(a_m)\bigr)$ has dimension greater than or equal to $n - m$ by Theorem 97.16 and the Principal Ideal Theorem 97.19.

(2): We may view $A \subset B$ hence $qf(A) \subset qf(B)$. The transcendence degree of $qf(B)$ over $qf(A)$ is $n - m$. Let $B = F[b_1, \ldots, b_N]$ with $b_1, \ldots, b_{n-m}$ a transcendence basis of $qf(B)$ over $qf(A)$ and $A = F[a_1, \ldots, a_M]$. Hence m of the a_i's are algebraically independent over F and the others polynomials in them (after appropriate clearing of denominators).

Fix i, $i = n - m + 1, \dots, N$. Since b_i is algebraically dependent on $b_1, \dots, b_{n-m}$ over $qf(A)$, there exist $f_i \in F(a_1, \dots, a_M, b_1, \dots, b_{n-m})[t_i]$ satisfying

$$(*) \qquad f_i(a_1, \dots, a_M, b_1, \dots, b_{n-m}, b_i) = 0$$

We may assume that f_i lies in $F[a_1, \dots a_M, b_1, \dots, b_{n-m}, t_i]$. This allows us to view as $F[a_1, \dots a_M][t_i, b_1, \dots, b_{n-m}]$ by clearing denominators. Let $g_i(a_1, \dots, a_M) \in F[a_1, \dots, a_M]$ be the coefficient of f_i of maximal total degree and maximal in the lexicographic order of $t_i, b_1, \dots, b_{n-m}$ (of nontrivial monomials of f_i). Then $V_A(g_i) < \operatorname{Spec}(A)$ is closed. Therefore, $V = \bigcup_{i=n-m+1}^{N} V_A(g_i) < \operatorname{Spec}(A)$ is also closed. In particular, $U = \operatorname{Spec}(A) \setminus V$ is a nonempty open subset. Let $\mathfrak{P} \in U$ and $X = V_B(\mathfrak{p})$ be an irreducible component of $V_B\big(\varphi(\mathfrak{P})\big)$. Let $^- : B \to B/\varphi(\mathfrak{P})$ by $b_i \mapsto \overline{b_i}$ for $i = 1, \dots, N$ and $^\sim : B/\varphi(\mathfrak{P}) \to B/\mathfrak{p}$ by $\overline{b_i} \mapsto \widetilde{b_i}$ for $i = 1, \dots, N$. By equation (*), each $\widetilde{b_i}$, $i = n - m + 1, \dots, N$, is algebraically dependent on $\widetilde{b_1}, \dots, \widetilde{b}_{n-m}$ over F. It follows that $\dim_A V(\mathfrak{p}) = \operatorname{tr\,deg}_{qf(A)} B/\mathfrak{p} \le n - m$ and hence by (1), we have equality. $\square$

Remark 99.2. We translate the above theorem in terms of irreducible abstract F-affine varieties. If $\varphi : X \to Y$ is a dominant map of irreducible abstract F-affine varieties (i.e., the associative map of an F-affine monomorphism of F-affine domains), then for all closed points $y \in Y$ and irreducible components Z of $\varphi^{-1}(y) := \{x \in X \mid \varphi(x) = y\}$, we have

$$\dim X \le \dim Y + \dim Z$$

and there exists an nonempty open set $U \subset X$ such that

$$\dim X = \dim Y + \dim Z$$

for all $y \in U$. For F-affine varieties over an algebraically closed field, we get the analogous result with the closed points being elements in F^M (in the notation of the proof of the theorem) by the Hilbert Nullstellensatz.

Remark 99.3. Let X be a topological space and $Y \subset X$. We say that Y is *locally closed* in X if Y is open in a closed subset of X. (Equivalently, Y is the intersection of an open and a closed subset of X.) We say that Y is *constructible* if Y is a finite union of locally closed subsets of X. Then it is a fact that if $\varphi : A \to B$ is an F-algebra homomorphism of affine F-algebras, it is always a constructible subset of $\operatorname{Spec}(A)$ although $\operatorname{im} {}^a\varphi$ may not be an affine F-algebra.

Exercises 99.4.

1. Let $\varphi : A \to B$ be an algebra homomorphism of domains. Using tensor products (cf. Exercises 39.12(15) and (23)), one defines the *scheme theoretic fiber* of a ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ over $\mathfrak{p}$ in $\operatorname{Spec}(A)$ to be $\operatorname{Spec}\big(B \otimes_A qf(A/\mathfrak{p})\big)$. Show that $B \otimes_A qf(A/\mathfrak{p}) = B_\mathfrak{p} \otimes_A A/\mathfrak{p} = B_\mathfrak{p}/\mathfrak{p}B_\mathfrak{p}$.
2. In the previous exercise, show that the map
$$({}^a\varphi)^{-1}(\mathfrak{P}) \to \operatorname{Spec}(B) \otimes_A qf(A/\mathfrak{p}) \;\text{ by }\; \mathfrak{P} \mapsto \mathfrak{P}B_\mathfrak{p}/\mathfrak{p}$$
is a homeomorphism of topological spaces.

100. Addendum: Japanese Rings

Let A be a Noetherian domain with quotient field F. We say that A is a *Japanese ring* if whenever K/F is a finite extension of fields, the integral closure A_K of A is a finitely generated A-module. We saw in Theorem 80.5 that if A is a Noetherian normal domain with quotient field F and K/F a finite separable extension, then the integral closure A_K of A in K is a finitely generated A-module. In particular, if $\operatorname{char}(F) = 0$ a normal domain A is Japanese. The hypothesis that K/F is separable is a nontrivial condition if $\operatorname{char}(F) = p > 0$. Indeed in the positive characteristic case, A may not be Japanese, even if A is normal. In fact, Schmidt-Nagata showed if F is satisfies $[F : F^p]$ is infinite and

$$\mathcal{S} := \{F_\alpha \mid F^p \subset F_\alpha \subset F \text{ are fields with } F_\alpha/F^p \text{ a finite extension}\}$$

partially ordered by $\subset$ with $\Gamma = \{\alpha \mid F_\alpha \in \mathcal{S}\}$ and $A_\alpha = F_\alpha[[t]]$, the formal power series ring with coefficients in A_α. then A_α is a DVR with maximal ideal A_α. Further let $A := \bigcup_\Gamma A_\alpha$, a ring with the obvious structure. Then it can be shown that A is a discrete valuation ring but is not Japanese, i.e., the result, in general, is false for dimension one noetherian rings. Of course fields are Japanese. We show even more is true. To investigate inseparable extensions, we need to use properties of purely inseparable extensions (cf. Exercise 53.10(14)). We begin with the following.

Lemma 100.1. *Let A be a Noetherian domain with quotient field F. Suppose that for all finite, purely inseparable field extensions E of F that A_E is a finitely generated A-module. Then A is Japanese.*

Proof. Let K/F be a finite field extension and L/K a normal closure of K/F. As $A_K \subset A_L$ and A is a Noetherian ring, it suffices to show that A_L is a finitely generated A-module by Theorem 40.7. So we may assume that K/F is a normal extension. Let $E = K^{G(K/F)}$. Then E/F is Galois, so separable and E/F is purely inseparable by Exercise 53.10(14). As A_E is a finitely generated A-module by hypothesis, it is a Noetherian normal domain, so A_K is a finitely generated A_E-module by Theorem 80.5. It follows that A_K is a finitely generated A-module. □

A Noetherian domain A is called *universally Japanese* if every domain B that is a finitely generated A-algebra B is Japanese.

Theorem 100.2. *Every field is universally Japanese.*

Proof. Let A be an affine F-algebra that is also a domain with quotient field K and L/K a finite field extension. We must show that A_L is a finitely generated A-module. By the Noether Normalization Theorem 97.1, there exist $x_1, \ldots, x_n$ in A algebraically independent over F with $F[x_1, \ldots, x_n] \subset A$ integral. As $K/F(x_1, \ldots, x_n)$ is a finite field extension. so is $L/F(x_1, \ldots, x_n)$. Therefore, it suffices to show that A_L is a finitely generated $F[x_1, \ldots, x_n]$-module, i.e., we may assume that $A = F[x_1, \ldots, x_n]$. Since $F[x_1, \ldots, x_n] \cong F[t_1, \ldots, t_n]$, we can assume further that $A = F[t_1, \ldots, t_n]$ and $K = F(t_1, \ldots, t_n)$. Moreover, by Lemma 100.1, we may assume that L/K is purely inseparable. As A is a normal Noetherian domain, we may also assume that $\operatorname{char}(F) = p > 0$.

Since L/K is finite and purely inseparable, there exists a positive integer m satisfying $L^{p^m} \subset K = qf(A)$. Let

$$\mathcal{C} := \text{the set of the } p^m\text{th roots of all the coefficients of all the } f_i, g_i$$

and

$$\begin{aligned} L &= K(a_1, \dots, a_n) \quad \text{with} \\ a_i^{p^m} &= \frac{f_i}{g_i} \quad \text{with} \quad f_i, g_i, \in A,\ g_i \neq 0 \text{ for all } i \\ \widetilde{F} &= F(\mathcal{C}) = F[\mathcal{C}] \\ E &= \widetilde{F}(t_1^{\frac{1}{p^m}}, \dots, t_n^{\frac{1}{p^m}}). \end{aligned}$$

We have $K \subset L \subset E$, using the Children's Binomial Theorem. As A is a Noetherian ring, it suffices to show that A_E is a finitely generated A-module. Let

$$B = \widetilde{F}[t_1^{\frac{1}{p^m}}, \dots, t_n^{\frac{1}{p^m}}] = F[\mathcal{C}, t_1^{\frac{1}{p^m}}, \dots, t_n^{\frac{1}{p^m}}].$$

Each of the finitely many elements, $c \in \mathcal{C}, t_1^{\frac{1}{p^m}}, \dots, t_n^{\frac{1}{p^m}}$, is integral over A, so B is a finitely generated A-module. We also have that $B \subset A_E$ with $qf(B) = E$ and $B = \widetilde{F}[t_1^{\frac{1}{p^m}}, \dots, t_n^{\frac{1}{p^m}}]$ a UFD, hence normal. Thus $A_E = B$ and the result follows $\square$

Corollary 100.3. *Let F be a field and A an affine F-algebra that is also a domain with $K = qf(A)$. If L/F be a finite field extension, then A_L is an affine F-algebra. In particular, the integral closure of A is a affine F-algebra.*

A noetherian ring A is called a *Nagata ring* if $A/\mathfrak{p}$ is Japanese for every prime ideal $\mathfrak{p}$ in $\operatorname{Spec}(A)$. It is a deep theorem that a noetherian domain A is a Nagata ring (a property internal to A) if and only if it is universally Japanese. In particular, if A is a Dedekind domain of characteristic zero, then A is a Nagata ring as it is Japanese and its quotients determined by maximal ideals are finite fields. Hence any Dedekind domain of characteristic zero is univerally Japanese. This is important in arithmetic algebraic geometry.

101. C_n-fields

We saw that over a finite field, every homogeneous polynomial of degree d in more than d variables has a nontrivial zero. We generalize this result, to classes of fields satisfying every homogeneous polynomial of degree d over the field in more than d^n variables for a fixed n has a nontrivial zero. Examples of fields satisfying this are fields of transcendence degree n over algebraically closed fields. This result will be applicable to the study of non-commutative F-algebras to be studied. We do give one example of this in this section, viz., an application to the generalization the Hamiltonian quaternions. Except for using the Principal Ideal Theorem to begin the first case of an induction proof, this section would have fit in the field theory part of the text.

Definition 101.1. Let $f \in F[t_1, \ldots, t_n]$ be a homogeneous form of degree d. We say that f is *normic* if f has only the trivial zero.

Example 101.2. Let K/F be a finite extensions of fields with $\mathfrak{B} = \{x_1, \ldots, x_n\}$ an F-basis. The *Norm Form of* $N_{K/F}$ is defined to be the homogeneous polynomial $N_{K/F}(t_1, \ldots, t_n) \in F[t_1, \ldots, t_n)$ of degree n satisfying $(a_1, \ldots, a_n) \mapsto N_{K/F}(a_1x_1 + \cdots + a_nx_n)$ for $a_1, \ldots, a_n \in F$ relative to the basis $\mathcal{B}$. If $x = a_1x_1 + \cdots + a_nx_n$ is not zero, then $N_{K/F}(x)$ is not zero, so $N_{K/F}$ is normic.
[If $\operatorname{char} F = p > 0$ and K/F is not separable, let $K/E/F$ with E the maximal separable extension of F in K, then $N_{K/F}(x)$ is defined to be $\big(N_{E/F}(x)\big)^{[K:E]}$.]

Lemma 101.3. *Let F be a field that is not algebraically closed. Then there exist normic forms over F of arbitrary large degree.*

Proof. Let K/F be of degree $n > 1$. By the example, there exists a normic form of degree n. Let $s > 1$, and suppose that f_s is a normic form of degree n^s. Then

$$f_{s+1}(t_1, \ldots, t_{n^2}) := \\ f_s\big(f_s(t_1, \ldots, t_n), f_s(f(t_{n+1}, \ldots, t_{2n}), \ldots, (f_s(t_{n^2-n+1}, \ldots .t_n^2)\big)$$

is a homogeneous form of degree n^{s+1} and is a normic form, since any zero of f_{s+1} would give a zero of f_s. □

Definition 101.4. A field F is called a *C_n-field* if for every positive integer d and every homogeneous polynomial f of degree d in more than d^n variables has a nontrivial zero, i.e., we have $|Z_F(f)| > 1$. [A C_1-field is also called a *quasi-algebraically closed field*.]

Example 101.5. Let F be a field.

(1) F is algebraically closed field it is clearly a C_0-field and conversely no non-algebraically closed field is a C_0-field by Example 101.2.
(2) If F is finite, then it is a C_1-field by Chevalley-Warning Theorem 67.5.

We give a simple application for C_1-fields.

Proposition 101.6. *Let F be a C_1-field and K/F a finite field extension. Then $N_{K/F} : K^\times \to L^\times$ is surjective.*

Proof. Suppose that $[K : F] = n$ with $\mathcal{B} = \{x_1, \ldots, x_n\}$ an F-basis. Let $N_{K/F}(t_1, \ldots, t_n) \in F[t_1, \ldots, t_n)$ denote the norm form based on $\mathcal{B}$. Let $0 \neq x \in F$ and $f = N(t_1, \ldots, t_n, t) - xt^n \in F[t_1, \ldots, t_n, t]$, a homogeneous form of degree n in $n+1$ variables, so has a solution $(a_1, \ldots, a_n, a) \in F^{n+1}$. As the norm form is normic, we cannot have $a = 0$. Let $b_i = a_i/a$ for $i = 1, \ldots, n$. Then $N(b_1x_1 + \cdots + b_nx_n) = N_{K/F}(b_1, \ldots, b_n) = x$. □

The definition of C_n-fields arose in Lang's thesis. One can also define $C_n(d)$-fields, i.e., fields for which all homogeneous polynomials over F of fixed degree d in greater than d^n-variables have a nontrivial zero. Let F be a local field, i.e., a field with a complete under a discrete valuation with finite residue class field. This includes all $\mathfrak{p}$-adic fields arising from completions of number fields under the $\mathfrak{p}$-adic completion, e.g., the field of p-adic numbers $\mathbb{Q}_p$, the quotient field of the p-adic integers (cf. Exercise 98.32(9)). Then F is a $C_2(2)$-field. In fact, if K is a number field in which -1 is a sum of squares, it is a $C_2(2)$-field (by the theorem of Hasse-Minkowski). Artin, who was Lang's advisor, conjectured that p-adic fields were C_2-fields. It was shown that they were $C_2(3)$-fields, but Terjanian showed Artin's conjecture to be false. However, using model theory, Ax and Kochen, showed that that for a fixed d only finitely many $\mathbb{Q}_p$ were not $C_2(d)$-fields.

In this section, we show that if F is an algebraically closed field, any field of transcendence degree n over F is a C_n-field. We apply this theorem to establish a special case of Tsen's theorem about algebraic extensions of $F(t)$ with F an algebraically closed field. The general case of Tsen's Theorem will be proven in Section 106.

Theorem 101.7. (Lang-Nagata) *Let F be a C_n-field and $f_1, \dots, f_r \in F[t_1, \dots, t_n]$ be homogeneous forms of degree d. If $N > rd^n$, then $|Z(f_1, \dots, f_r)| > 0$.*

PROOF. If $n = 0$, then F is algebraically closed. By Corollary 97.20 to the Principal Ideal Theorem, $\dim Z(f_1, \dots, f_r) \geq N - r > 0$, so the result follows. Therefore, we may assume that F is not algebraically closed, i.e., F is not a C_0-field so $n > 0$. In particular, by the lemma, there exists a normic form φ of degree $s > r$. Define

$$\begin{aligned}\varphi_1 :=& \\ &\varphi\big(f_1(t_1, \dots, t_N), \cdots, f_r(t_1, \dots, t_N), f_1(t_{N+1}, t_{2N}), \dots, f_r(t_{N+1}, t_{2N}),\\ &\quad \dots f_1(t_{kN+1}, t_{(k+1)N}), \dots, f_r(t_{kN+1}, t_{(k+1)N}), 0, \dots, 0\big)\end{aligned}$$

where $rk \leq s < r(k+1)$. Let $[x]$ denote the largest integer in $x \in \mathbb{R}$. So φ_1 is a homogeneous form in $N\big[\frac{s}{r}\big]$ variables of degree $ds \leq dr(\big[\frac{s}{r}\big] + 1)$ variables. If $n = 1$, i.e., F is a C_1-field, then φ_1 has a nontrivial zero if

$$(*) \qquad N\left[\frac{s}{r}\right] > dr(\left[\frac{s}{r}\right] + 1) \geq ds.$$

Since $N - dr > 0$, equation (*) will hold if s is chosen sufficiently large. As φ is normic, a nontrivial zero for φ_1 gives an element in $Z(f_1 \dots, f_r)$. If $n > 1$, we can iterate the above process, defining $\varphi_1, \dots, \varphi_m, \dots$. We must show that there exists an integer m satisfying $|Z(\varphi_m)| > 1$. Indeed, suppose such an m exists. Choose a minimal such m. As φ_{m-1} is normic by the choice of m, a nontrivial zero for φ_m gives a nontrivial element in $Z(f_1 \dots, f_r)$.
Consequently, we need only show that there exists a nonzero m satisfying $|Z(\varphi_m)| > 1$. Let

$$D_m = \deg \varphi_m = d^m s \,(= \text{total degree})$$

$$N_m = \text{ the number of variables of } \varphi_m = \left[\frac{N_{m-1}}{r}\right] N.$$

Claim. If $m >> 0$, then $N_m > (D_m)^n$:

If we prove the claim, then we are done by the definition of being a C_n-field. We shall, in fact, prove that $N_m D_m^{-n} \to \infty$ as $m \to \infty$.

[If we ignore [ ], we roughly have $N_m \approx \dfrac{N_{m-1}}{r} N \approx \dfrac{N^m}{r^m} s$, so

$$N_m D^{-n} \approx N_m r^{-m} s^{1-m} d^{-mn} \geq r^m d^{nm} r^{-m} s^{1-m} d^{-mn} + m + 1$$

which goes to infinity as $m \to \infty$. Hence the above is plausible.]
We have
$$N_{m+1} = \Big[\frac{N_m}{r}\Big]N > \frac{N_m}{r}N > \Big(\frac{N_m}{r} - 1\Big)N = N_m\frac{N}{r} - N.$$
Iterating this inequality yields
$$\begin{aligned} N_{m+1} &> \Big(N_{m-1}\frac{N}{r} - N\Big)\frac{N}{r} = N_{m-1}\frac{N^2}{r^2} - N\Big(1 + \frac{N}{r}\Big) \\ &> \cdots > N_1\Big(\frac{N}{r}\Big)^m - N\Big(1 + \frac{N}{r} + \frac{N^2}{r^2} + \cdots + \frac{N^{m-1}}{r^{m-1}}\Big) \\ &= N_1\Big(\frac{N}{r}\Big)^m - N\Big(\frac{(N/r)^m - 1)}{(N/r) - 1}\Big) \\ &= \Big(N_1 - \frac{N_r}{N-r}\Big)\Big(\frac{N}{r}\Big)^m + \frac{Nr}{N-r} \\ &> \Big(N_1 - \frac{Nr}{N-r}\Big)\Big(\frac{N}{r}\Big)^m. \end{aligned}$$
As $N_1 = N[\frac{s}{r}]$, we may choose $s >> 0$ to satisfying $N_1 > Nr/(N-r)$. Hence if we set $C = N_1 - Nr/(N-r)$, we have
$$\frac{N_{m+1}}{D_{m+1}^n} > \frac{C(N/r)^m}{D_{m+1}^n} = \frac{C(N/r)^m}{d^{m+1)n}s^n} = \frac{C}{(ds)^n}\Big(\frac{N}{rd^n}\Big)^m \to \infty$$
as $m \to \infty$, since $N > rd^n$. □

Theorem 101.8. Lang) *Let F be a C_n-field and K/F a field extension.*

(1) If K/F is algebraic, then so is K.
(2) $F(t)$ is a C_{n+1}-field.

In particular, if $m = \mathrm{tr}\ \deg_F K$, then K is a C_{n+m}-field.

Proof. (1): Let $f \in K[t_1, \dots, t_N]$ be a homogeneous form of degree d and $N > d^n$. Then there exists a field extension $K/E/F$ with E/F finite and $f \in E[t_1, \dots, t_N]$. Let $\mathcal{B} = \{x_1, \dots, x_m\}$, be an F-basis for E. Let t_{ij}, $1 \le i \le N$ and $1 \le j \le m$, be the algebraically independent elements over F satisfying $t_i = t_{i1}x_1 + \cdots + t_{im}x_m$ for $i = 1, \dots N$. Substituting these into f and writing the coefficients of F in the basis $\mathcal{B}$ yields
$$f(t_1, \dots, t_N) = f_1(t_{11}, \dots, t_{Nm})x_1 + \cdots + f_m(t_{11}, \dots, t_{Nm})x_m$$
with $f_i \in F[t_1, \dots, t_N) \in F(t)[t_{11}, \dots, t_{Mm}]$ homogeneous forms of degree d, $j = 1, \dots, m$. Since $Nm > md^n$, the f_i's have a common nontrivial zero by the Lang-Nagata Theorem 101.7. This gives a nontrivial zero for f over K.

(2): Let $f \in F(t)[t_1, \dots, t_N]$ be a homogeneous form of degree d with $N > d^{n+1}$. Clearing denominators does not change the condition that f has or does not have a nontrivial zero, so we may assume that $f \in F[t][t_1, \dots, t_N]$. Let $r = \deg_t f$, the degree of f in t. Let $m > 0$, to be chosen later. Let t_{ij}, $1 \le i \le N$, $0 \le j \le m$ be the algebraically independent over F satisfying $t_i = t_{i0} + t_{i1}t + \cdots + t_{im}t^m$ for $i = 1, \ldots N$. Substituting these into f yields

$$f(t_1, \ldots, t_N) = f_0(t_{i0}, \ldots, t_{Nm})t^0 + \cdots + f_m(t_{10}, \ldots, t_{Nm})t^{dm+r}.$$

Each f_i is a homogeneous form of degree d in $N(m+1)$ variables. Since $N > d^{n+1}$, we can choose $m >> 0$ to satisfy $N(m+1) > d^n(dm+r+1)$, i.e., $(N - d^{n+1})m > d^n(r+1) - N$. With this choice of m, the f_i have a nontrivial common zero that yields one for f over $F(t)$. □

Corollary 101.9. *Let F be an algebraically closed field and K/F satisfying* $\operatorname{tr}\deg_F K = n$. *Then K is a C_n-field.*

We give an application, which we shall generalize later. To do so, we generalize the definition of Hamiltonian quaternions, which in itself, is of mathematical interest.

Construction 101.10. We follow the steps in our construction of the Hamiltonian quaternions. Let F be a field of characteristic different than two (leaving some verifications to the reader). ([f the characteristic of F is two, a different generalization holds.] Let $a, b \in F^\times$. Let A be a four dimensional vector space over F on basis $\{1, i, j, k\}$ where $1 = 1_F \in V$. One checks that A becomes a (non-commutative) ring by linearly extending

$$i^2 = a, \quad j^2 = b, \quad k = ij = -ji \tag{101.11}$$

(so $k^2 = -ab$) with $0_A = 0_F$ and $1_A = 1_F$, called a *generalized quaternion algebra* .

A is an F-algebra, i.e., $rx = xr$ for all $r \in F, x \in A$ with center F (check). We denote this F-algebra by $\left(\dfrac{a,b}{F}\right)$. We also assume when this is written, that it comes with a basis $\{1, i, j, k\}$ satisfying equation (101.11). So the Hamiltonian quaternions can be written $\mathcal{H} = \left(\dfrac{-1,-1}{F}\right)$. Check that $\left(\dfrac{b,a}{F}\right) \cong \left(\dfrac{a,b}{F}\right)$ as F-algebras.

Let $A = \left(\dfrac{a,b}{F}\right)$. Define $^- : A \to A$ by

$$x = x_0 1 + x_1 i + x_2 j + x_3 k \mapsto \overline{x} = x_0 1 - x_1 i - x_2 j - x_3 k, \text{ with the } x_i \in F.$$

This map is an anti-isomorphism (just as for the Hamiltonian quaternions) and an involution.

We also have a norm map $N : A \to F$ defined by

$$z \mapsto z\overline{z} = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2$$

for all $z = x_0 1 + x_1 i + x_2 j + x_3 k$ in A. It satisfies for all $x, y \in A$:

1. $1 = \overline{1}$.
2. $\overline{x+y} = \overline{x} + \overline{y}$.
3. $\overline{xy} = \overline{y}\,\overline{x}$.
4. $N(xy) = N(x)N(y)(= N(y)N(x))$.

If A is a generalized quaternion algebra and $N(x) \neq 0$ for $x \in A$, then x has an inverse, viz., $\overline{x}/N(x)$, hence is a division ring if $N(x) \neq 0$ for all nonzero $x \in A$. The difference between the Hamiltonian quaternions and generalized quaternions, is that it is now possible that $N(x) = 0$ with x nonzero which means that A may not be a division ring.

Lemma 101.12. *Let F be a field of characteristic different from two. If $a, b, x, y \in F^{\times}$, then*
$$\Big(\frac{a,b}{F}\Big) \cong \Big(\frac{ax^2, by^2}{F}\Big):$$

PROOF. Let $\{1, i, j, k\}$ and $\{1, i', j', k'\}$ be the defining F-bases for $\Big(\frac{a,b}{F}\Big) \cong \Big(\frac{ax^2, by^2}{F}\Big)$, respectively. As $(xi)(yj) = xy(ij) = -(yj)(xi)$, we see that the F-linear map
$$\varphi : \Big(\frac{ax^2, by^2}{F}\Big) \to \Big(\frac{a,b}{F}\Big) \text{ defined by } i' \mapsto xi, \quad j' \mapsto yj', \text{ and } k' \mapsto xyk$$
is an F-linear isomorphism and is checked to be a ring isomorphism (hence an F-algebra isomorphism). $\square$

Corollary 101.13. *Let F be a field of characteristic different from two. Then $\Big(\frac{1,b}{F}\Big) \cong \mathbb{M}_2(F)$.*

PROOF. If $\{1, i, j, k\}$ is the F-basis for $\Big(\frac{1,b}{F}\Big)$, then
$$i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } j \mapsto \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix}$$
defines an F-algebra isomorphism. $\square$

Examples 101.14. 1. Over the complex numbers, every generalized quaternion $\mathbb{C}$-algebra is isomorphic to $\mathbb{M}_n(\mathbb{C})$.

2. Over the real numbers, if $a, b \in \mathbb{R}^{\times}$,
$$\Big(\frac{a,b}{\mathbb{R}}\Big) \cong \begin{cases} \mathcal{H}, & \text{if } a < 0 \text{ and } b < 0 \\ \mathbb{M}_2(\mathbb{R}), & \text{otherwise.} \end{cases}$$

More generally, we have

Proposition 101.15. *Let F be a field of characteristic different from two and $A = \Big(\frac{a,b}{F}\Big)$. Then the following are equivalent:*

(1) A is isomorphic to $\mathbb{M}_2(F)$.
(2) A is not a division ring.
(3) There exists a nonzero element in A with zero norm.
(4) The element b is a norm from the field extension $F(\sqrt{a})/F$.

PROOF. We need only prove (3) $\Rightarrow$ (4) and (4) $\Rightarrow$ (1). If a is a square, both implications follow, so we may assume not.

(3) $\Rightarrow$ (4): $F(\sqrt{a})/F$ is a quadratic extension. If $x = x_0 1 + x_1 i + x_2 j + x_3 k$ in A satisfies $N(x) = 0$, we have
$$N_{F(\sqrt{a})/F}(x_2 + \sqrt{a}x_3)b = (x_2^2 - ax_3^2)b = x_0^2 - ax_1^2.$$
If $x_2^2 - ax_3^2$ is zero, this equation implies that a is a square, so it is not zero. Hence
$$b = N_{F(\sqrt{a})/F}(x_0 + \sqrt{a}x_1)N_{F(\sqrt{a})/F}(x_2 + \sqrt{a}x_3)^{-1}$$

and (4) follows.
(4) $\Rightarrow$ (1): As b is a norm from $F(\sqrt{a})$, so is b^{-1}. Let $b = y^2 - az^2$, with $y, z \in F$. Set $j' = yj + zij$. Then $(j')^2 = by^2 - abz^2 = 1$. Check that $j'i = -ij'$. Let $i' := (1+a)i + (1-a)j'i$. Check that $i'j' = -j'i'$ and $(i')^2 = (1+a)^2a - (1-a^2)a = 4a^2$. Then the map $A \to \left(\dfrac{1, 4a^2}{F}\right)$ defined by $i \mapsto i'$, $j \mapsto j'$, $k \mapsto k' = i'j'$ defines an F-algebra isomorphism. Since we know that $\left(\dfrac{1, 4a^2}{F}\right) \cong \mathbb{M}_2(F)$, the result follows; $\square$

Corollary 101.16. *Let F be a field of characteristic not two. A generalized quaternion F-algebra is either a division F-algebra or isomorphic to $\mathbb{M}_2(F)$. In particular, it is an F-algebra with center F that is simple as ring and four dimensional as an F-vector space.*

We now give our application to the principal subject of this section. The generalization of this result is called Tsen's Theorem, which we shall come back to when we study noncommutative algebra.

Corollary 101.17. *Let F be a field of transcendence degree one over an algebraically closed field. The every generalized quaternion algebra over F is isomorphic to $\mathbb{M}_2(F)$.*

PROOF. Let $A = \left(\dfrac{a, b}{F}\right)$ be a generalized quaternion F-algebra on basis $\{i, j, k\}$. Define $N_{A/F)} \in F[t_1, \dots, t_4]$ by $N_{A/F}(x_0, x_1, x_3, x_4) = N(x_0 + x_1 i + x_2 j + x_3 k)$. Then $N_{A/F}$ is a homogeneous polynomial of degree four. As F is a C_1-field, it has a nontrivial zero. The result follows. $\square$

Exercises 101.18.

1. Fill in the details in Construction 101.10.
2. Fill in the details of Proposition 101.15
3. Show that the generalized quaternion algebra $\left(\dfrac{a, b}{F}\right)$ is isomorphic to $\left(\dfrac{a, -ab}{F}\right)$. In particular, $\left(\dfrac{a, a}{F}\right) \cong \left(\dfrac{a, -1}{F}\right)$.
4. Show if $A = \left(\dfrac{a, b}{F}\right)$ is a generalized quaternion algebra, then A is not a division F-algebra if and only if the quadratic polynomial $at_1^2 + bt_2^2 = 1$ has a nontrivial solution. In particular if $a \neq 0$ or 1, then $\left(\dfrac{a, 1-a}{F}\right)$ is not a division F-algebra.

Part 7

Semisimple Algebras and Representation Theory

CHAPTER XVIII

Division and Semisimple Rings

In this chapter, we study the simplest non-commutative rings, viz., division rings. More generally, we study matrix rings over division rings. In the first section, we show that simple left Artinian rings are up to isomorphism matrix rings over division rings and uniquely so. We then study products of such rings. This is important as it is the foundation of the theory of finite group representations that we shall study in the next chapter. We then investigate another way of obtaining division rings important in number theory. Finally, we generalize Wedderburn's Theorem that finite division rings are fields to a theorem of Jacobson, showing that a ring is commutative if for every element x in the ring, there exists an integer n satisfying $x^n = x$.

102. Wedderburn Theory

Modules over a division ring are just vector spaces with a line in a vector space a 'simple' module. As vector spaces are direct sums of lines, a module over a division ring is a direct sum of 'simple' submodules. modules over division rings, i.e., vector spaces are direct sums of lines. In this section, we shall generalizes this to a module over a simple left Artinian ring, i.e., any such is a direct sum of 'simple' submodules. We shall also prove Wedderburn's Theorem that any simple Artinian ring is, in fact, isomorphic to a matrix ring over a division ring. We begin with properties of modules.

Definition 102.1. Let R a nonzero ring. We say

1. A nonzero (left) R-module is *irreducible* (or *simple*) if it contains no proper submodules. An irreducible left ideal in R is also called a *minimal left ideal*.
2. A (left) R-module is called *completely reducible* if every submodule of it is a direct summand of it.
3. The ring R is *(left) semisimple* if it is a completely reducible module over itself.

Of course, we have the analogous definitions for right R-modules. When we wish to consider right modules we shall write the word right.

Examples 102.2. 1. Over a division ring, every vector space is completely reducible, since every vector space is free on a basis.

2. Let D be a division ring. Then the kth column space of $\mathbb{M}_n(D)$,

$$M_k := \{A = (a_{ij}) \in \mathbb{M}_n(D) \mid a_{ij} = 0 \text{ if } j \neq k\},$$

is a minimal left ideal of R and is an n-dimension vector space over D. Moreover, as $\mathbb{M}_n(D)$-modules, we have

$$\mathbb{M}_n(D) = M_1 \oplus \cdots \oplus M_n \text{ and } M_i \cong M_j \text{ for all } i, j.$$

Lemma 102.3. *Let $\widetilde{M}$ be a nonzero R-module and M_i irreducible submodules of $\widetilde{M}$ for i in I (not necessarily non-isomorphic). Let M be the R-module $\sum_I M_i$. Then there exists a subset J of I satisfying M is the direct sum $\bigoplus_J M_j$.*

Proof. Using Zorn's Lemma, we see that there exists a subset J of I maximal such that $N = \sum_J M_j = \bigoplus_J M_j$. Let i_o be an element in $I \setminus J$. Then the R-module M_{i_o} is irreducible and contains the submodule $N \cap M_{i_o}$, so either $N \cap M_{i_o} = M_{i_o}$, i.e., $M_{i_o} \subset N$ or $N \cap M_{i_o} = 0$, i.e., $\sum_{J \cup \{i_o\}} M_j = \bigoplus_{J \cup \{i_o\}} M_j$. By maximality, this second case does not occur, so $M = N$. □

Proposition 102.4. *Let M be a nonzero R-module. Then the following are equivalent:*

(1) M is a sum of irreducible R-submodules.
(2) M is a direct sum of irreducible R-modules.
(3) M is a completely reducible R-module.

Proof. (1) $\Rightarrow$ (2) follows from the lemma.

(2) $\Rightarrow$ (3): Let N be a submodule of M and $M = \bigoplus_I M_i$ with every M_i, $i \in I$, irreducible. Using Zorn's Lemma, there exists a maximal subset J of I such that $N + \bigoplus_J M_j = N \oplus \bigoplus_J M_j$. By the argument in the proof of the lemma, this must be M.

(3) $\Rightarrow$ (1): We prove the following:

Claim 102.5. If M is a completely reducible R-module and M_0 a nonzero submodule of M, then there exists an irreducible submodule of M_0:

Let m be a nonzero element of M_0. Then we have an R-epimorphism

$$\rho_m : R \to Rm \text{ given by } r \mapsto rm.$$

As the the annihilator $\operatorname{ann}_R m$ is a left ideal in R, the map ρ_m induces an R-isomorphism

$$\overline{\rho_m} : M/\operatorname{ann}_R m \to Rm.$$

Using Zorn's Lemma, we know that there is a maximal left ideal $\mathfrak{m}$ in R with $\operatorname{ann}_R m \subset \mathfrak{m}$. By the Correspondence Principle, $\mathfrak{m}m < Rm$ is maximal, hence $Rm/\mathfrak{m}m$ is irreducible. By hypothesis, M is completely reducible, so $M = M' \oplus \mathfrak{m}m$ for some submodule M' of M. Let x be an element of Rm. Then there exists an $r \in \mathfrak{m}$ and an $m' \in M'$ satisfying

$$x = m' + rm \text{ so } m' = x - rm \text{ lies in } M' \cap Rm.$$

It follows that

$$Rm = (M' \cap Rm) \oplus \mathfrak{m}m \subset M_0 \text{ and } M' \cap Rm \cong Rm/\mathfrak{m}m \text{ is irreducible.}$$

This establishes the claim. Now let M_0 be the sum of all the irreducible submodules of M. By the claim $M_0 \neq 0$. As M is completely reducible, $M = M_0 \oplus M_1$ for some submodule M_1. If $M_1 \neq 0$, then applying the claim again shows that M_1 contains an irreducible submodule of M. This contradicts the choice of M_0. □

Corollary 102.6. *Let D be a division ring. Then $\mathbb{M}_n(D)$ is semisimple.*

Remark 102.7. No nonzero ideal in $\mathbb{Z}$ is irreducible, so $\mathbb{Z}$ is not semisimple. If $p > 0$ is a prime in $\mathbb{Z}$, then $\mathbb{Z}/p\mathbb{Z}$ is semisimple as it is a field. More generally, $\mathbb{Z}/n\mathbb{Z}$ with $n > 0$ a product of distinct primes is semisimple. Note that $\mathbb{Z}/n\mathbb{Z}$, $n > 0$ a product of distinct primes, satisfies the descending chain condition, i.e., is an Artinian ring, and has no nonzero nilpotent (left) ideals $\mathfrak{A}$, i.e., (left) ideals $\mathfrak{A}$ such that $\mathfrak{A}^n = 0$ for some positive integer n. Note also that $\mathbb{Z}/n\mathbb{Z}$ with $n > 1$ not square free is Artinian but has nonzero nilpotent elements and is not semi-simple. (Cf. Exercise 103.7(4) below.)

Corollary 102.8. *Let M_i, $i \in I$, be completely reducible R-modules. Then $\coprod_I M_i$ is completely reducible.*

Corollary 102.9. *Let M be a completely reducible R-module, and N a submodule of M. Then N and M/N are completely reducible.*

Proof. We may assume that $N \neq 0$. By Claim 102.5, there exists a submodule N_0 of N that is the sum of all the irreducible submodules of N. As M is completely reducible, $M = N_0 \oplus N_1$ for some submodule N_1. As in the proof of the proposition, we see that $N = N_0 \oplus (N_1 \cap N)$. By Claim 102.5, we must have $N_1 \cap N = 0$, so $N = N_0$ and N is completely reducible. As $M = N \oplus N'$ for some submodule N' of M, N' is completely reducible by what we have just shown, hence so is $M/N \cong N'$. □

Definition 102.10. Let R be a ring. An element e in R is called an *idempotent* if $e^2 = e$. If $e_1, \dots, e_n$ are idempotents, they are called *orthogonal* if $e_i e_j = \delta_{ij} e_i$ for all i and j.

Examples 102.11. Let R be a ring.

1. The elements 0 and 1 of R are idempotents. They are called the *trivial idempotents* and are orthogonal.
2. If e is an idempotent so is $1 - e$, and then $e, 1 - e$ are orthogonal idempotents.
3. The sum of any orthogonal idempotents is an idempotent.
4. Let $M = M_1 \oplus M_2$ and $p_i : M \to M$ be the projection $m_1 + m_2 \mapsto m_i$, where $m_i \in M_i$, for $i = 1, 2$. Then p_1, p_2 are orthogonal idempotents in the ring $\operatorname{End}_R(M)$.
5. Let $S = \mathbb{M}_n(R)$ and e_{ij} the matrix with 1 in the ijth entry, zero elsewhere. Then $e_{11}, \dots, e_{nn}$ are orthogonal idempotents in S. We also note that
$$e_{ii} S e_{ii} \cong R \text{ as rings.}$$

Proposition 102.12. *Let R be a nonzero ring. Then the following are equivalent:*

(1) Every R-module is completely reducible.
(2) Every short exact sequence in R-modules splits.
(3) The ring R is semisimple.
(4) $R = \bigoplus_{i=1}^n \mathfrak{A}_i$ for some n and some left ideals $\mathfrak{A}_i$ with each $\mathfrak{A}_i$, $i = 1, \dots, n$, a minimal left ideal; and, furthermore, $\mathfrak{A}_i = Re_i$, $i = 1, \dots, n$, with $e_1, \dots, e_n$ orthogonal idempotents. Moreover if $\mathfrak{A}$ is a minimal left ideal, then $\mathfrak{A} = \mathfrak{A}_i$ for some i. [The $\mathfrak{A}_i$ need not be mutually non-isomorphic.]

Proof. The equivalence of (1) and (2) follows from the corollaries above and (4) $\Rightarrow$ (3) from the proposition. So we need only show (3) $\Rightarrow$ (4).

As R is completely reducible, $R = \bigoplus_I \mathfrak{A}_i$ for some minimal left ideals $\mathfrak{A}_i$. There exist nonzero $e_{i_1}, \dots, e_{i_n}$ in R for some n with $e_{i_j} \in \mathfrak{A}_{i_j}$ and $1 = e_{i_1} + \cdots + e_{i_n}$. It follows that $R \subset \mathfrak{A}_{i_1} \oplus \cdots \oplus \mathfrak{A}_{i_n}$, hence we must have equality. Changing notation, we may assume that $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$, with $e_i \in \mathfrak{A}_i$ for each i and $1 = e_1 + \cdots + e_n$. It follows that $e_j = \sum e_i e_j$ in $\mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$, hence $e_1, \dots, e_j$ are orthogonal idempotents. As $0 < Re_j \subset \mathfrak{A}_j$, we have $Re_j = \mathfrak{A}_j$, since $\mathfrak{A}_j$ is irreducible for each j. Finally, if $\mathfrak{A}$ is a minimal left ideal in R, then $\mathfrak{A} \cap \mathfrak{A}_j > 0$ for some j, so $\mathfrak{A} = \mathfrak{A}_j$. □

Remark 102.13. The proposition says that R is a semi-simple ring if and only if every R-module is R-projective (cf. Exercises 39.12(12), (12)) if and only if every R-module is R-injective (cf. Exercises 38.18(19), (21)).

Let D be a division ring and $R = \mathbb{M}_n(D)$. Then R is a simple ring, i.e., the only 2-sided ideals in R are 0 and R and R is left and right semisimple, left and right Noetherian (i.e, ACC on left and right ideals), left and right Artinian (i.e., DCC on left and right ideals), since R is a finite dimensional D-vector space. More generally, if $D_1, \dots, D_r$ are division rings and $A = \mathbb{M}_{n_1}(D_1) \times \cdots \times \mathbb{M}_{n_r}(D_r)$ then A is left and right semisimple, left and right Noetherian, and left and right Artinian.

One problem that arises when working with vector spaces over noncommutative division rings is that the composition of linear operators and the multiplication of their matrix representations (relative to some fixed bases) of these operators does not correspond, in fact, is reversed. One way around this is to have linear operators and scalars operate on vectors on different sides. Another way is indicated by the following lemma.

Lemma 102.14. *Let M be an R-module. Then the rings* $\mathrm{End}_R(\coprod_{i=1}^n M)$ *and* $\mathbb{M}_n\big(\mathrm{End}_R(M)\big)$ *are isomorphic.*

Proof. Let $N = \coprod_{i=1}^n M_i$ with $M = M_i$ for all i. Then we have the usual maps

$$\iota_i : M_i \to N \text{ given by } m \mapsto (0, \dots, \underbrace{m}_{i}, \dots, 0)$$

and

$$\pi_j : N \to M_j \text{ given by } (m_1, \dots, m_n) \mapsto m_j$$

are an R-monomorphism and R-epimorphism, respectively, satisfying

$$\pi_j \iota_i = \delta_{ij} 1_{M_i} \text{ and } \sum_{j=1}^n \iota_j \pi_j = 1_N.$$

Let $f \in \mathrm{End}_R(N)$ and $f_{ij} := \pi_i f \iota_j$ in $\mathrm{End}_R(M)$. Define

$$\varphi : \mathrm{End}_R(N) \to \mathbb{M}_n\big(\mathrm{End}_R(M)\big) \text{ by } f \mapsto (f_{ij}).$$

This map is clearly additive and it is a ring homomorphism, for if f, g lie in $\mathrm{End}_R(N)$, we have

$$(f_{ij})(g_{lk}) = (\pi_i f \iota_j)(\pi_l g \iota_k) = (\pi_i f g \iota_k) \text{ and } 1 \mapsto (\pi_l \iota_j) = I.$$

Now define

$$\psi : \mathbb{M}_n\big(\mathrm{End}_R(M)\big) \to \mathrm{End}_R(N) \text{ by } (f_{ij}) \mapsto \sum_{i=1}^n \sum_{j=1}^n \iota_i f_{ij} \pi_j,$$

an additive map that is a ring homomorphism, for if $(f_{ij}), (g_{ij})$ lie in $\mathbb{M}_n\big(\operatorname{End}_R(M)\big)$, we have $\psi\big(\operatorname{diag}(1_M, \dots 1_M)\big) = 1_N$ and

$$\begin{aligned}\psi\big((f_{il})(g_{kj})\big) = \psi\big((\sum_{l=1}^{n} f_{il}g_{lj})\big) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \iota_i f_{il} g_{lj} \pi_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \iota_i f_{il} \pi_l \iota_l g_{lj} \pi_j = \psi\big((f_{il})\big)\psi\big((g_{kj})\big).\end{aligned}$$

Clearly, $\varphi \circ \psi = 1_{\mathbb{M}_n\left(\operatorname{End}_R(M)\right)}$ and $\psi \circ \varphi = 1_{\operatorname{End}_R(N)}$, so φ and ψ are inverse isomorphisms. □

For the next lemma, we do write endomorphisms and scalars on different sides.

Lemma 102.15. *Let e be an idempotent in the ring R and define*

$$\rho : eRe \to \operatorname{End}_R(Re) \ \ by \ \ eae \mapsto \rho_{eae} : xe \mapsto xe \cdot eae.$$

View Re as a right $\big(\operatorname{End}_R(Re)\big)$-module. Then ρ is an isomorphism of rings.

Proof. Check that ρ_{eae} is an R-homomorphism of Re when we write endomorphisms on the right, i.e., $(re)\rho_{eae} = r(e\rho_{eae})$. [Notice when we write it in this way it looks like the associative law.] It is also easily checked that ρ is a ring homomorphism.

ρ is a $1-1$: If $0 = re \cdot eae = reae$ for all r in R, then, setting $r = e$, shows that $eae = 0$.

ρ is onto: Let f lie in $\operatorname{End}_R(Re)$. Then there exists an element a in R satisfying $(e)f = ae$ in $Re \subset R$. Therefore, for all r in R, we have

$$(re)f = (re \cdot e)f = re\big((e)f\big) = reae = re \cdot eae = (re)\rho_{eae},$$

so $f = \rho_{eae}$. □

The following is essentially an immediate observation, but of great use.

Lemma 102.16. (Schur's Lemma) *Let M be an irreducible R-module. Then $\operatorname{End}_R(M)$ is a division ring.*

Proof. Let f in $\operatorname{End}_R(M)$ be nonzero. Then $\ker f < M$ and $0 < \operatorname{im} f \subset M$. As M is irreducible, $\ker f = 0$ and $\operatorname{im} f = M$, hence f is an R-isomorphism. □

We can now classify nonzero simple, left Artinian rings.

Theorem 102.17. (Wedderburn's Theorem) *Let R be a nonzero ring. Then the following are equivalent:*

(1) R is simple and left Artinian.
(2) R is simple and semisimple.
(3) There exists a division ring D and a positive integer n such that $R \cong \mathbb{M}_n(D)$.

Suppose that R satisfies (3). Then in (3), D is unique up to a ring isomorphism and n is unique. More precisely, if R satisfies (2) and e is a nonzero idempotent in R, then the following are true:

(i) All minimal left ideals in R are isomorphic.

(ii) *If* $\mathfrak{A}$ *is a minimal left ideal in* R *and* $D = \mathrm{End}_R(\mathfrak{A})$*, then* $eRe \cong \mathbb{M}_m(D)$ *for some positive integer* m.

(iii) *In* (ii)*, the ring* D *is a division ring and the center* $Z(eRe)$ *of* eRe *is isomorphic to* $Z(D)$*, the center of* D.

(iv) *If* D *and* E *are division rings and* $\mathbb{M}_m(D) \cong \mathbb{M}_n(E)$*, with* m, n *positive integers, then* $D \cong E$ *and* $m = n$.

Proof. We have already shown that (3) $\Rightarrow$ (1) and (3) $\Rightarrow$ (2).

(1) $\Rightarrow$ (2): As R is left Artinian, there exists minimal left ideal $\mathfrak{A}$ of R by the Minimal Principle. Let

$$0 < \mathfrak{B} := \sum_{r\in R} \mathfrak{A}r \subset R,$$

a (2-sided) ideal. As R is simple, $\sum_{r\in R} \mathfrak{A}r = \mathfrak{B} = R$. Let $\rho_r : \mathfrak{A} \to \mathfrak{A}r$ be the R-epimorphism defined by $a \mapsto ar$. Since $\mathfrak{A}$ is irreducible, either $\rho_r = 0$ or ρ_r is an R-isomorphism, i.e., either $\mathfrak{A}r = 0$ or $\mathfrak{A}r \cong \mathfrak{A}$. It follows that R is completely reducible.

(2) $\Rightarrow$ (3): It suffices to show that (2) $\Rightarrow$ (i) — (iv), since 1 is an idempotent and $1R1 \cong R$. The argument to prove (1) $\Rightarrow$ (2) shows:

$R = \sum_{r\in R} \mathfrak{A}r$ with $\mathfrak{A}$ a minimal left ideal in R and
$\mathfrak{A} \cong \mathfrak{A}r$ for all r in R satisfying $\mathfrak{A}r \neq 0$,

as semisimple rings contain minimal left ideals. We now show conditions (i) — (iv) are satisfied.

(i): As argued before, we see that

There exists a minimal left ideal $\mathfrak{A}$ in R.
$R = \mathfrak{A}r_1 \oplus \cdots \oplus \mathfrak{A}r_n$ for some $r_1, \ldots, r_n$ in R.
Every minimal left ideal in R is $\mathfrak{A}r_i$ for some i, hence is isomorphic to $\mathfrak{A}$.

(ii). As R is completely reducible, there exists a nonzero idempotent e in R. As $0 < Re \subset R$, Re is also a completely reducible R-module. By Claim 102.5, Re contains an irreducible submodule, hence a minimal left ideal of R. Using the notation in the proof of (i), we have

$$Re = \bigoplus_{j=1}^{m} (\mathfrak{A}r_{i_j} \cap Re) \cong \coprod_{j=1}^{m} \mathfrak{A},$$

where the i_j, $1 \le j \le m$, are those integers satisfying $\mathfrak{A}r_{i_j} \cap Re \neq 0$. By Lemmas 102.14 and 102.15 (with $e = 1$ as a special case), we have

$$eRe \cong \mathrm{End}_R(Re) \cong \mathrm{End}_R(\coprod_{j=1}^{m} \mathfrak{A}) \cong \mathbb{M}_m\big(\mathrm{End}_R(\mathfrak{A})\big).$$

This establishes (ii).

(iii) follows by Schur's Lemma 102.16 and Exercise 102.21(4).

(iv): Let $A = \mathbb{M}_m(D)$, $B = \mathbb{M}_n(E)$, and $e_{11} = (\delta_{1i}\delta_{1j})$ in $\mathbb{M}_m(D)$. We have $Ae_{11} \subset \mathbb{M}_m(D)$ is a minimal left ideal. It is unique up to isomorphism by (i). Similarly, if $e'_{11} = (\delta_{1i}\delta_{1j})$

in $\mathbb{M}_n(E)$, then Be'_{11} is a minimal left ideal in B, unique up to isomorphism. By Lemma 102.15 and Example 102.11(5), we have

$$D \cong e_{11}Ae_{11} \cong \operatorname{End}_A(Ae_{11})$$
$$E \cong e'_{11}Be'_{11} \cong \operatorname{End}_B(Be'_{11}).$$

As $A \cong B$, their unique minimal left ideals (up to isomorphism) must correspond, so $D \cong E$ by the above. It follows that

$$m^2 = \dim_D A = \dim_E B = n^2,$$

so $m = n$ also. □

Corollary 102.18. *Let R be a simple ring. then R is left Artinian if and only if R is right Artinian if and only if R is left semisimple if and only if R is right semisimple.*

PROOF. This is true for $\mathbb{M}_n(D)$ with D a division ring. □

Corollary 102.19. *Let R be a simple, left Artinian ring. Then the following are equivalent:*

(1) R is a division ring.
(2) The only zero divisor of R is zero.
(3) The only idempotents of R are 0 and 1.
(4) The only nilpotent element in R is zero.

Let F be a field. Recall that a nonzero ring A is called an *F-algebra* if there exists a ring homomorphism $F \to Z(A)$ where $Z(A)$ is the center of A. This must be a monomorphism and we usually identify F with $F1_A$. This ring homomorphism makes A into an F-module, i.e., a vector space over F. We say that the F-algebra A is a *finite dimensional F-algebra* if $\dim_F A$ is finite.

Corollary 102.20. *Let F be a field and A a finite dimensional F-algebra. Suppose that A is also a simple ring. Then $A \cong \mathbb{M}_n(D)$ for some division ring D containing F in its center. The division ring D is a finite dimensional F-algebra, unique up to isomorphism and n is unique.*

PROOF. The ring A is left Artinian since a finite dimensional vector space over F. □

Let F be a field and A a finite dimensional F-algebra. If A is also simple and *central*, i.e., $Z(A) = F$, then A is called a *central simple F-algebra*. By Wedderburn Theorem, $A \cong \mathbb{M}_m(D)$ for some division ring D. More over $Z(D) = F$ and D is a finite dimensional vector space over F. If $D = F$, we say that A is *split*. If A and B are central simple F-algebras, we say that A and B are *similar* if $A \cong \mathbb{M}_m(D)$ and $B \cong \mathbb{M}_n(E)$ with D and E isomorphic division rings. This is an equivalence relation by Wedderburn's Theorem. The classes of central simple F-algebras under this equivalence relation can be given an abelian group structure and form what is called the *Brauer group of F*, an important group in algebra and number theory. The Brauer group of $\mathbb{R}$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ and the Brauer group of $\mathbb{C}$ is trivial, i.e., every central simple algebra over $\mathbb{C}$ splits. This follows from the results in Section 104. A theorem of Tsen shows that the Brauer group of a field of transcendence degree one over $\mathbb{C}$ is also trivial. This is harder to prove. The determination of the Brauer group of a number field is one of the crowning achievements of twentieth century number theory.

Exercises 102.21.

1. Let V be a nonzero finite dimensional vector space over a field (or division ring) F. Then V is is an $\mathrm{End}_F(V)$-module via evaluation, i.e., $fv := f(cv)$ for all f in $\mathrm{End}_F(V)$ and v in V. Show that V is $\mathrm{End}_F(V)$-irreducible.
2. Let R be a commutative ring with ideals $\mathfrak{A}_i$, $i = 1, 2$. Set
$$V_i := \{\mathfrak{p} \mid \mathfrak{p} \text{ a prime ideal with } \mathfrak{A}_i \subset \mathfrak{p}\}.$$
Suppose that $V_1 \cap V_2 = \emptyset$ and every prime ideal in R lies in V_1 or V_2, i.e., the set of prime ideals is the disjoint union of V_1 and V_2. Show that R contains a nontrivial idempotent.
3. Up to isomorphism, show that exists a unique irreducible R-module if R is simple and left Artinian.
4. Show if R is a ring, the map $R \to \mathbb{M}_n(R)$ given by $r \mapsto rI$ defines an isomorphism between the center $Z(R)$ of R and the center $Z\big(\mathbb{M}_n(R)\big)$ of $\mathbb{M}_n(R)$.

103. The Artin-Wedderburn Theorem

In the last section, we classified simple semisimple rings. In this section, we classify semisimple rings. The prototype for the simple case was a matrix ring over a division ring. The prototype for the general case is is a finite product of matrix rings over division rings. This generalization is quite useful, as it gives a foundation for that part of the Theory of Group Representations devoted to finite groups over fields of characteristic zero.

If we have a finitely generated completely reducible module over a ring R, then we would expect a nice decomposition for it. Indeed, if we look at the case of a finite dimensional vector space then it has a composition series. The analogue gives us the following:

Lemma 103.1. *Let $M_1, \dots, M_r$ and $N_1, \dots, N_s$ be two finite collections of non-isomorphic irreducible R-modules. If*
$$M_1^{m_1} \coprod \cdots \coprod M_r^{m_r} \cong N_1^{n_1} \coprod \cdots \coprod N_s^{n_s}$$
for some positive integers $m_1, \dots, m_r$, $n_1, \dots, n_s$, then $r = s$ and there exists a permutation $\sigma \in S_r$ such that $M_i \cong N_{\sigma(i)}$ for all i. In particular, if $M = N$, then $M_i = N_{\sigma(i)}$ for all i.

Proof. Let $M = M_1^{m_1} \coprod \cdots \coprod M_r^{m_r}$, $N = N_1^{n_1} \coprod \cdots \coprod N_s^{n_s}$, and $\varphi : M \to N$ an R-isomorphism. Clearly, φ and φ^{-1} sets up a bijection
$$\{M_1, \dots, M_r\} \longleftrightarrow \{N_1, \dots, N_s\},$$
so $r = s$. Changing notation, we may assume that $M_i \cong N_i$ for all i and we are reduced to showing the following

Claim. If N is an irreducible R-module and $N^m \cong N^n$, then $m = n$:

Since $D = \mathrm{End}_R(N)$ is a division ring by Schur's Lemma, and

$$\mathbb{M}_m\big(\mathrm{End}_R(N)\big) \cong \mathrm{End}_R(N^m) \cong \mathrm{End}_R(N^n) \cong \mathbb{M}_n\big(\mathrm{End}_R(N)\big)$$

by Lemma 102.14, we have $m = n$ by Wedderburn's Theorem 102.17. The last statement is immediate using $\varphi = 1_M$ □

An important generalization of this lemma, the Krull-Schmidt Theorem, occurs when we replace the word irreducible by the word indecomposable, where an R-module is called *indecomposable* if it cannot be written as a nontrivial direct sum of two submodules. We shall not prove this generalization.

Lemma 103.2. *Let R be a semisimple ring. Then R is left Artinian. More precisely, if $\mathfrak{B}$ is a nontrivial left ideal of R, then there exist unique minimal left ideals $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ in R satisfying $\mathfrak{B} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_m$.*

Proof. We have shown that $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$ for minimal left ideals $\mathfrak{A}_i$ and if $\mathfrak{A}$ is a minimal left ideal, then $\mathfrak{A} = \mathfrak{A}_i$ for some i. It follows, as before, that

$$\mathfrak{B} = \bigoplus_{i=1}^{m} (\mathfrak{A}_{j_i} \cap \mathfrak{B}) \text{ with the } j_i \text{ satisfying } \mathfrak{A}_{j_i} \cap \mathfrak{B} \neq 0.$$

The last lemma says that this is unique. As any submodule of $\mathfrak{B}$ is a direct sum of some of the $\mathfrak{A}_{j_1}, \ldots, \mathfrak{A}_{j_m}$, the result follows. □

Our first formulation of the generalization of Wedderburn's Theorem is the next result. However, in the proof we shall show much more. Since notation will be used in the proof, we shall postpone this more precise statement until after the proof.

Theorem 103.3. (Artin-Wedderburn Theorem) *Let R be a semisimple ring. Then R is an (internal) direct sum of finitely many simple left Artinian rings, all unique (up to order). In particular, any semisimple right is both left and right Artinian (and left and right semisimple).*

Proof. We prove this in a number of steps.

Step 1. Let $\mathfrak{A}$ be a minimal left ideal in R. Set

$$S_{\mathfrak{A}} := \{\mathfrak{B} \mid \mathfrak{B} \subset R \text{ a minimal left ideal with } \mathfrak{B} \cong \mathfrak{A}\}$$
$$B_{\mathfrak{A}} := \sum_{S_{\mathfrak{A}}} \mathfrak{B}.$$

Then the following are true:

(i) $B_{\mathfrak{A}} \subset R$ is a 2-sided ideal.

(ii) If $\mathfrak{A}'$ is a minimal left ideal in R, then

$$\mathfrak{A} \cong \mathfrak{A}' \text{ if and only if } B_{\mathfrak{A}} B_{\mathfrak{A}'} \text{ is not zero:}$$

(i): Let $\mathfrak{B} \in S_{\mathfrak{A}}$ and $x \in R$. As $\rho_x : \mathfrak{B} \to \mathfrak{B}x$ by $y \mapsto yx$ is an R-epimorphism, $\mathfrak{B}x = 0$ or $\mathfrak{B}x \cong \mathfrak{B} \cong \mathfrak{A}$. In either case, $\mathfrak{B}x \subset B_{\mathfrak{A}}$, so $B_{\mathfrak{A}}$ is a 2-sided ideal.

(ii): ($\Rightarrow$): We know that $\mathfrak{A} = Re > 0$ with e an idempotent. Then

$$0 \neq e^2 = e \cdot e \in B_{\mathfrak{A}} B_{\mathfrak{A}} = B_{\mathfrak{A}} B_{\mathfrak{A}'}.$$

($\Leftarrow$): Suppose that $B_{\mathfrak{A}}B_{\mathfrak{A}'}$ is not zero. Then there exist a $\mathfrak{B} \in S_{\mathfrak{A}}$ and a nonzero b' in $\mathfrak{B}' \in S_{\mathfrak{A}'}$ with $\mathfrak{B}b' \neq 0$. As $\mathfrak{B}$ is irreducible, $\mathfrak{B} \to \mathfrak{B}b'$ by $x \mapsto xb'$ is an R-isomorphism. As $0 < \mathfrak{B}b' \subset \mathfrak{B}'$ with $\mathfrak{B}'$ irreducible, we have

$$\mathfrak{A} \cong \mathfrak{B} \cong \mathfrak{B}b' \cong \mathfrak{B}' \cong \mathfrak{A}'.$$

Step 2. Let $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_q$ with $\mathfrak{A}_i$, $i = 1, \ldots, q$, the minimal left ideals in R, arranged such that $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$, $m \leq q$, are all the mutually non-isomorphic ones. [So if $i > m$, there exists a $j \leq m$ with $\mathfrak{A}_i \cong \mathfrak{A}_j$ and every minimal left ideal in $\mathfrak{A}$ is isomorphic to some $\mathfrak{A}_i$ with $i \leq m$.] Set $B_i = B_{\mathfrak{A}_i}$ for $1 \leq i \leq m$. Then $R = \sum_{i=1}^m B_i$ and $B_iB_j = 0$ if $i \neq j$ (by Step 1). Moreover,

(i) $R = \bigoplus_{i=1}^m B_i$.

(ii) If $\mathfrak{B}$ is a nonzero 2-sided ideal in R, then $\mathfrak{B} = B_{i_1} \oplus \cdots \oplus B_{i_n}$, some i_j, $1 \leq i_j \leq m$:

(i): Let $C := B_i \cap \sum_{i \neq j} B_j$. Then

$$B_iB_j = 0 \text{ so } CB_j = 0 = B_jC \text{ as } C \subset B_i,$$
$$B_jB_i = 0 \text{ so } CB_i = 0 = B_iC \text{ as } C \subset \sum_{j \neq i} B_j,$$

so $C = RC = 0$.

(ii): As R is completely reducible, there exists a minimal left ideal $\mathfrak{A}$ in R contained in $\mathfrak{B}$, and, in fact, $\mathfrak{B}$ is completely reducible so a sum of minimal left ideals of R. As $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ are all the minimal left ideals of R up to isomorphism, we may assume that $\mathfrak{A} \cong \mathfrak{A}_i$ lies in $\mathfrak{B}$ and need only show

Claim. $B_i \subset \mathfrak{B}$:

Write $\mathfrak{A} = Re$ with e a nonzero idempotent. We must show if $\mathfrak{A}' \in S_{\mathfrak{A}}$, then $\mathfrak{A}' \subset \mathfrak{B}$. Let $a \in \mathfrak{A} = Re = Re \cdot e = \mathfrak{A}e$. Therefore, there exists an element b in $\mathfrak{A}$ such that $a = be = be \cdot e = a \cdot e$, i.e., if $a \in \mathfrak{A}$ then $a = ae$. Let $\varphi : \mathfrak{A} \to \mathfrak{A}'$ be an R-isomorphism. Then for every a in $\mathfrak{A}$, we have $\varphi(a) = \varphi(ae) = a\varphi(e)$ in $\mathfrak{A}'$. Since $\mathfrak{B}$ is a 2-sided ideal, $\mathfrak{A}' = \mathfrak{A}\varphi(e) \subset \mathfrak{B}$, proving the claim.

Step 3. Let B_i be as in Step 2. Then B_i is a simple, left Artinian ring:

We know $R = B_1 \oplus \cdots \oplus B_m$, so $1 = f_1 + \cdots + f_m$ for some $f_i \in B_i$.

Claim. The elements $f_1, \ldots, f_m$ are *central* orthogonal idempotents, i.e., in addition to the f_i being orthogonal idempotents, they lie in the center of R:

We know that $f_1, \ldots, f_m$ are orthogonal idempotents, so we need only show that they are central. Let x be an element of R. Then we have $x = \sum_i x_i$ for some $x_i \in B_i$ for each i. As $1_Rx = x = x1_R$, we must have $f_ix_i = x_i = x_if_i$ for all i and $f_jx_i = 0 = x_if_j$ for all $j \neq i$. It follows that $f_ix = xf_i$ as needed.

It follows by the claim that B_i is a ring with $1_{B_i} = f_i$. Let $\mathfrak{B}$ be a 2-sided ideal in B_i. Since $B_iB_j = 0 = B_jB_i$ for all $j \neq i$, we have

$$\mathfrak{B} = B_i\mathfrak{B} = (\sum_j B_j)B_i\mathfrak{B} = \sum_j B_j\mathfrak{B} = R\mathfrak{B},$$

and similarly $\mathfrak{B} = \mathfrak{B}R$. Therefore, $\mathfrak{B}$ is a nonzero 2-sided ideal in R. By Step 2, $\mathfrak{B}$ is a sum of some of the B_j's, hence must be B_i, so B_i is simple. As R is completely reducible, there exists a minimal left ideal $\mathfrak{A}$ of R lying in B_i. If $\mathfrak{A}' < \mathfrak{A}$ is a left ideal in B_i, then $B_j\mathfrak{A}' \subset B_jB_i = 0$ for all $j \neq i$, so $\mathfrak{A}'$ is a left ideal in R, hence $\mathfrak{A}' = 0$. Therefore, $\mathfrak{A}$ is a minimal left ideal in B_i. It follows easily that B_i is semisimple, hence left Artinian.

Step 4. Suppose that $R = \bigoplus_{i=1}^m B_i = \bigoplus_{j=1}^n C_j$ with the B_i's as in Step 2 and each C_j a 2-sided ideal in R that is also a simple, left Artinian ring. Then $m = n$ and there exists a permutation $\sigma \in S_m$ such that $B_i = C_{\sigma(i)}$ for every i. The $B_1, \dots, B_m$ are called the *simple components* of R and the decomposition $R = B_1 \oplus \cdots \oplus B_m$ is called a *Wedderburn decomposition* of R:

Each $B_i \cap C_j$ is a 2-sided ideal in both simple rings B_i and C_j. Therefore, for each i, there exists a k with $B_i \cap C_k$ nonzero, hence $B_i = C_k$. Similarly, for each j, there exists an l such that $B_l \cap C_j$ is nonzero, hence $B_l = C_j$. The result follows.

Step 5. Finish:

Each B_i is a simple, left Artinian ring, hence also right Artinian (and right semisimple). It follows easily that R is also right Artinian. □

The proof of the theorem also shows the following:

Theorem 103.4. (Artin-Wedderburn) *Let R be a nonzero ring. Then the following are equivalent:*

(1) R is semisimple.

(2) R is a finite direct product of simple, left Artinian rings.

(3) There exist (possibly isomorphic) division rings $D_1, \dots, D_m$, some m, unique up to isomorphism (and order), and unique positive integers $n_1, \dots, n_m$ and a ring isomorphism

$$R \cong \mathbb{M}_{n_1}(D_1) \times \cdots \times \mathbb{M}_{n_m}(D_m) \ \textit{ up to order.}$$

If R is a semisimple ring and $\mathfrak{A}_1, \dots, \mathfrak{A}_m$ are all the non-isomorphic minimal left ideals in R, then $\{\mathfrak{A}_1, \dots, \mathfrak{A}_m\}$ is called a *basic set* for R.

We leave the proof of the following two corollaries as exercises:

Corollary 103.5. *Let R be a semisimple ring and $\{\mathfrak{A}_1, \dots, \mathfrak{A}_m\}$ a basic set. If M is an irreducible R-module then $M \cong \mathfrak{A}_i$ for some i. In particular, if R is also simple, then, up to isomorphism, there exists a unique irreducible R-module.*

Corollary 103.6. *Let R be a semisimple ring and $\{\mathfrak{A}_1 \dots, \mathfrak{A}_m\}$ a basic sets for R. Let $B_i = \sum_{S_{\mathfrak{A}_i}} \mathfrak{A}$ for $i = 1, \dots, r$ and M a nonzero R-module. Then B_iM is a submodule of M and is a sum of irreducible submodules each isomorphic to $\mathfrak{A}_i$. Further, $M = \bigoplus_{i=1}^m B_iM$.*

Exercises 103.7.

1. If R is a semisimple ring and M a nontrivial R-module, then M is a direct sum of irreducible modules, unique up to isomorphism and order.
2. Prove Corollary 103.5.
3. Prove Corollary 103.6.
4. Show that a ring is semisimple if and only if it is left artinian and has no nonzero nilpotent left ideals.

104. Finite Dimensional Real Division Algebras

We have seen that the (Hamiltonian) quaternion algebra $\mathcal{H}$, the four dimensional real vector space with basis $\{1, i, j, k\}$, made into a ring by defining a multiplication on this basis by

$$i^2 = -1,\ j^2 = -1,\ ij = -ji = k$$

and extending this linearly to the whole space defines a division ring. The center $\{z \in \mathcal{H} \mid yz = zy \text{ for all } y \text{ in } \mathcal{H}\}$ of $\mathcal{H}$ is $\mathbb{R}$. Frobenius solved the problem of finding all division algebras D that are finite dimensional real vector spaces with $\mathbb{R}$ in its center, i.e.,

$$\mathbb{R} = \{x \mid xy = yx \text{ for all } y \text{ in } D\}.$$

We call such algebras *finite dimensional real division algebras*. If D is such a division algebra, then $\mathbb{R} \to D$ given by $r \mapsto r1_D$ is a ring monomorphism that we view as an inclusion. The result is the following:

Theorem 104.1. (Frobenius) *Let D be a finite dimensional real division algebra. Then D is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathcal{H}$.*

Proof. We may assume that $\mathbb{R} < D$. As D is a finite dimensional real vector space, for any element x in D, we must have x is a root of a nonzero polynomial in $\mathbb{R}[t]$ and $\mathbb{R}[x]$ is a commutative ring. As $\mathbb{R}[x] \subset D$ with D a division ring, $\mathbb{R}[x]$ must be a domain. It follows, just as in the proof of Theorem 48.13, that $\mathbb{R}[x] = \mathbb{R}(x)$ is a field and a finite extension of $\mathbb{R}$. (Cf. Exercise 48.25(1).) By the Fundamental Theorem of Algebra, we must have $\dim_{\mathbb{R}} \mathbb{R}(x) \leq 2$ and if x is not a real number, $\{1, x\}$ is a real basis for $\mathbb{R}(x)$ and $\mathbb{R}(x) \cong \mathbb{C}$. In particular, there exists an i in D such that $i^2 = -1$. Then $\mathbb{R}(i) \cong \mathbb{C}$, so we can identify $\mathbb{R}(i)$ and $\mathbb{C}$, which we do, i.e., we may view $\mathbb{R} \subset \mathbb{C} \subset D$. Therefore, D is a (left) finite dimensional complex vector space. Let $T : D \to D$ be the $\mathbb{C}$-linear transformation given by $x \mapsto xi$. Then T satisfies $T^2 = -1_D$, so the minimal polynomial q_T of T satisfies $q_T \mid t^2 + 1$ in $\mathbb{C}[t]$. It follows that the only possible eigenvalues of T are $\pm i$ and at least one of them most be an eigenvalue. Let $E_T(\alpha) = \{v \in D \mid Tv = \alpha v\}$ for α in $\mathbb{C}$, so either $E_T(i) \neq 0$ or $E_T(-i) \neq 0$. Since $E_T(i) \cap E_T(-i) = 0$, we have $E_T(i) \oplus E_T(-i) \subset D$. Let x be an element in D. Then we have

$$T(x - ixi) = xi + ix = i(x - ixi), \text{ so } x - ixi \text{ lies in } E_T(i)$$
$$T(x + ixi) = xi - ix = -i(x - ixi), \text{ so } x + ixi \text{ lies in } E_T(-i).$$

As

$$x = \frac{1}{2}(x - ixi) + \frac{1}{2}(x + ixi) \text{ lies in } E_T(i) + E_T(-i),$$

we have $D = E_T(i) \oplus E_T(-i)$ as a complex vector space.

If $x \in E_T(i) = \{x \in D \mid ix = xi\}$, then $\mathbb{C}[x]$ is a commutative division ring, i.e., a field. Therefore, by the Fundamental Theorem of Algebra, $E_T(i) = \mathbb{C}$, so $\dim_{\mathbb{C}} E_T(i) = 1$. Consequently, if $E_T(-i) = 0$, we must have $D = \mathbb{C}$. So we may assume that $E_T(-i)$ is nonzero. If x and y lie in $E_T(-i)$, then

$$(*) \qquad xyi = x(-i)y = ixy,$$

so xy lies in $E_T(i)$. Let y be a nonzero element in $E_T(-i)$. Define a $\mathbb{C}$-linear transformation

$$\rho_y : D \to D \text{ by } x \mapsto xy.$$

As D is a division ring, ρ_y must be a monomorphism (why?). By (*), we have

$$\rho_y|_{E_T(-i)} : E_T(-i) \to E_T(i).$$

As the linear transformation $\rho_y|_{E_T(-i)}$ is injective and $\dim_{\mathbb{C}} E_T(i) = 1$, we must also have $\dim_{\mathbb{C}} E_T(-i) = 1$.

Claim: If x is a nonzero vector in $E_T(-i)$, then x^2 is real and $x^2 < 0$:

We know that $x^2 \in \mathbb{R}(x) \cap E_T(i) = \mathbb{R}(x) \cap \mathbb{C}$ by (*). As $\mathbb{R}[x]$ is a two dimensional real vector space on basis $\{1, x\}$ with x not lying in $\mathbb{C}$, if $a + bx$, a, b real, lies in $\mathbb{C}$, we must have $b = 0$. It follows that x^2 is real. If $x^2 > 0$, then $x^2 = \theta^2$ for some real number θ. This means that $\pm\theta, x$ would be three distinct roots of $t^2 - x^2$ in $\mathbb{R}[t]$, which is impossible. Therefore, $x^2 < 0$ and the claim is established.

Now let $j = x/\sqrt{|x^2|}$. Then $j^2 = -1$. Set $k = ij$, Then $\{j, k\}$ is a real basis for the vector space $E_T(-i)$ — check — and $\{1, i, j, k\}$ is a basis for the real vector space D satisfying $i^2 = -1 = -j^2$ and $k = ij = -ji$. If follows that $D \cong \mathcal{H}$. $\square$

105. Cyclic Algebras

Wedderburn's Theorem shows that any central simple algebra is a matrix ring over a division ring. However, it does not give any indication on how to find division rings. Wedderburn was also interested in discovering new division rings. The simplest construction of new central simple algebras became important in Number Theory. In this section, we introduce this construction. Behind its foundation is Hilbert's Theorem 90 (and its cohomological formulation that we shall not discuss).

We begin with the construction of new rings.

Definition 105.1. Let R be any nonzero ring and σ a ring automorphism of R. Define the *twisted polynomial ring* $R[t, \sigma]$ to be the ring with the usual addition of polynomials with multiplication given by

$$(\sum_i a_i t^i)(\sum_j b_j t^j) := \sum_{i,j} a_i \sigma^i(b_j) t^{i+j}$$

where the a_i, b_j lie in R. So, in general, t does not commute with elements in R, rather we have $tb = \sigma(b)t$ for all b in R.

Remarks 105.2. Let R be a nonzero ring and σ a ring automorphism of R.

1. If for any a and b in R, satisfying $ab = 0$, we have $a = 0$ or $b = 0$, i.e., R is a *non-commutative domain*, then so is the twisted polynomial ring $R[t, \sigma]$.
2. Let $R = K$ be a field. Then the left division algorithm holds in $K[t, \sigma]$. In particular, every left ideal is principal (i.e., $K[t, \sigma]$ is a *left PID*). If $F \subset K^{\langle\sigma\rangle}$ is a subfield and f is a polynomial in $F[t]$, then $K[t, \sigma]f$ is a two sided ideal, so $K[t, \sigma]/K[t, \sigma]f$ is a ring.

Definition 105.3. Let K/F be a (finite) cyclic field extension of degree n, i.e., Galois with cyclic Galois group, with $G(K/F) = \langle\sigma\rangle$ and $f = t^n - a$ a polynomial in $F[t]$ with a nonzero. Then $(K/F, \sigma, a) := K[t, \sigma]/K[t, \sigma]f$ is called a *cyclic algebra of degree* n over

F. Equivalently, using the canonical epimorphism $^- : K[t,\sigma] \to K[t,\sigma]/K[t,\sigma]f$, we have $(K/F,\sigma,a)$ is a vector space over K on basis $\{1, x, \dots, x^{n-1}\}$ with $x = \bar{t}$ and satisfying

$$x^n = a \quad \text{and} \quad x\beta = \sigma(\beta)x \text{ for all } \beta \text{ in } K.$$

It is convenient to use both of these formulations, so we switch between them with this notation without further comment.

Note that a cyclic algebra $(K/F,\sigma,a)$ of degree n over F is a finite dimensional F-algebra of F-dimension n^2.

Examples 105.4. 1. The Hamiltonian quaternions $\mathcal{H}$ is the cyclic algebra $(\mathbb{C}/\mathbb{R}, ^-, -1)$ over $\mathbb{R}$ of degree 2. Indeed, viewing $\mathbb{C} = \mathbb{R}(\sqrt{-1})$, then j is the unique element in $\mathcal{H}$ satisfying $j\alpha j^{-1} = \overline{\alpha}$ for all, i.e., $j\alpha = \overline{\alpha}j$ for all $\alpha \in \mathbb{C}$ (as it is true for i) and $j^2 = -1$.

2. More generally, the generalized quaternions in Construction 101.10 are cyclic algebras over a field F of characteristic different from two (and in characteristic two, a different construction produces an F-algebra of generalized quaternions which we have not constructed). Indeed, let $A = \left(\dfrac{a,b}{F}\right)$ be a generalized quaternion algebra with a not a square in F with F-basis $\{1, i, j, k\}$. Let $K = F(\sqrt{a}) = F(i)$ and $G(K/F) = \langle\sigma\rangle$. Then $j \in A^\times$ satisfies $j\alpha j^{-1} = \alpha$ for all $\alpha \in K$, i.e., $j\alpha = \overline{\alpha}j$, just as above, with $j^2 = b$.

3. We now generalize the construction of generalized quaternions. Let F be a field of characteristic zero or of prime degree not dividing n and containing a primitive nth root of unity. Suppose that K/F is a cyclic extension of degree n with $G(K/F) = \langle\sigma\rangle$. Then $K = F(\alpha)$ with irreducible $m_F(\alpha) = t^n - a$ in $F[t]$, so $\alpha^n = a$ and $\sigma(\alpha) = \zeta\alpha$ for some primitive nth root of unity ζ by Theorem 60.20 and its proof. Let A be an F-algebra of dimension n^2 on basis $\{\alpha^i\beta^j \mid 0 \le i, j \le n-1\}$ with $K \subset A$ satisfying

$$\begin{aligned} A &= \coprod_{i,j=0}^{n-1} F\alpha^i\beta^j \\ yz &= \zeta zy \\ \alpha^n &= a, \quad \beta^n = b \end{aligned}$$

Then $\beta\theta\beta^{-1} = \sigma(\theta)$, i.e., $\beta\theta = \sigma(\theta)\beta$, for all $\theta \in K$, so A is a cyclic F-algebra. Conversely, if A is a cyclic F-algebra with $K \subset A$, then we shall see in Section 106 that the Skolem-Noether Theorem 106.18 will imply that there exists $\beta \in A$ satisfying $\beta^{-1}\theta\beta^{-1} = \sigma(\theta)$, i.e., $\beta\theta = \sigma(\theta)\beta$, for all $\theta \in K$. In particular, $\beta^n = b$, for some $b \in F$, and $A = (K/F,\sigma,a)$. This cyclic algebra is usually written $\left(\dfrac{a,b}{F,\zeta}\right)$.

We shall also see that if F has a cyclic extension of degree n, then $\mathbb{M}_n(F)$ is also an example of a cyclic algebra. To do this, we show that cyclic algebras are central simple algebras.

Proposition 105.5. *Let $R = (K/F,\sigma,a)$ be a cyclic algebra over F of degree n. Then R is a central simple algebra over F. Moreover, the centralizer $Z_R(K) := \{x \in R \mid xb = bx \text{ for all } b \in K\}$ is K and the only field L satisfying $K \subset L \subset R$ is K.*

PROOF. We first show that R is simple. Write

$$R = K1_R \oplus \cdots \oplus Kx^{n-1}$$
$$x^n = a \text{ with } a \in F \text{ and } x\alpha = \sigma(\alpha)x \text{ for all } \alpha \in K.$$

Let $0 < \mathfrak{A} \subset R$ be a 2-sided ideal. Choose a nonzero element $z \in \mathfrak{A}$ satisfying

$$z = \alpha_{i_1}x^{i_1} + \cdots + \alpha_{i_r}x^{i_r} \text{ with } \alpha_{i_1}, \dots, \alpha_{i_r} \in K^\times,\ 0 \le i_1 < \cdots < i_r < n$$

and r minimal. If $r = 1$, then z is a unit in R as α_{i_1}, x^{i_1} are units, so we may assume that $r > 1$. As $\sigma^{i_1} \neq \sigma^{i_r}$, there exists and element β in K satisfying $\sigma^{i_1}(\beta) \neq \sigma^{i_r}(\beta)$. In the ideal $\mathfrak{A}$, we have

$$\begin{aligned}(1) \qquad z\beta &= (\alpha_{i_1}x^{i_1} + \cdots + \alpha_{i_r}x^{i_r})\beta \\ &= \alpha_{i_1}\sigma^{i_1}(\beta)x^{i_1} + \cdots + \alpha_{i_r}\sigma^{i_r}(\beta)x^{i_r} \\ (2) \quad \sigma^{i_1}(\beta)z &= \alpha_{i_1}\sigma^{i_1}(\beta)x^{i_1} + \cdots + \alpha_{i_r}\sigma^{i_1}(\beta)x^{i_r}.\end{aligned}$$

Subtracting (2) from (1) yields the nonzero element

$$z\beta - \sigma(\beta)z = \alpha_{i_2}\big(\sigma^{i_2}(\beta) - \sigma^{i_1}(\beta)\big)x^{i_2} + \cdots + \alpha_{i_r}\big(\sigma^{i_r}(\beta) - \sigma^{i_1}(\beta)\big)x^{i_r},$$

contradicting the minimality of r. Therefore, R is simple.

Suppose that $z = \sum_{i=0}^{n-1} \alpha_i x^i$ is nonzero in R and commutes with all β in K. Then $\alpha_i\sigma(\beta)x^i = \alpha_i\beta x^i$ for $i = 0, \dots, n-1$. If β does not lie in F, then $\alpha_i = 0$ for $i = 1, \dots, n-1$ and β must lie in K. If $K \subset L \subset R$ is a subfield, then $L \subset Z_R(K) = K$.

Finally, if $z \in Z(R)$, then $z \in Z_R(K) = K$. As $z\beta = \sigma(\beta)z$ for all $\beta \in K$, we have $z \in K^{\langle\sigma\rangle} = F$. $\square$

A field L lying in the cyclic algebra $R = (K/F, \sigma, a)$ not properly contained in any larger in R is called *a maximal subfield* of R. So K is one such. There can be many non-isomorphic ones.

The proposition allows us to show:

Computation 105.6. Let K/F be a cyclic extension of fields of degree n with Galois group $\langle\sigma\rangle$ and $R = (K/F, \sigma, 1)$. We show that $R \cong \mathbb{M}_n(F)$:
Since $t^n - 1 = (t-1)(t^{n-1} + \cdots + t + 1)$ in $K[t, \sigma]$ and $R = K[t, \sigma]/(t^n - 1)$, we have $t^n - 1$ lies in the principal ideal $R(t-1)$, a maximal left ideal in R by Remark 105.2(2). It follows that $M := K[t, \sigma]/K[t, \sigma](t-1) \cong K$ is a simple R-module of F-dimension n. Since R is a simple ring, the F-algebra homomorphism $\rho : R \to \operatorname{End}_F(M)$ defined by $r \mapsto \rho_r : m \mapsto mr$ must be a monomorphism. Since $\dim_F R = \dim_F \operatorname{End}_F(M)$ and $\mathbb{M}_n(F) \cong \operatorname{End}_F(M)$, the result is established.

Corollary 105.7. *Let $R = (K/F, \sigma, a)$ be a cyclic algebra over F of prime degree p. Then R is either a division ring or $R \cong \mathbb{M}_p(F)$.*

Recall that a central simple F algebra is called *split* if it is isomorphic to a matrix ring over F.

PROOF. By Wedderburn's Theorem 102.17, $R \cong \mathbb{M}_r(D)$ for some division ring D, so $p^2 = r^2 \dim_F(D)$. As p is a prime number, $p = r$ and $\dim_F(D) = 1$ splits, (i.e., R splits) or R is a division algebra (i.e., $r = 1$ and $R = D$). $\square$

Theorem 105.8. *Let $R = (K/F, \sigma, a)$ be a cyclic algebra of degree n over F. Then $R \cong \mathbb{M}_n(F)$ if and only if $a \in \mathrm{N}_{K/F}(K^\times)$.*

Proof. ($\Leftarrow$): Suppose that $a \in \mathrm{N}_{K/F}(K^\times)$. Then there exists an $\alpha \in K^\times$ such that $aN_{K/F}(\alpha) = 1$. Let $y = \alpha x$ where x is the image of t in $K[t, \sigma]$ under the canonical epimorphism. Set $y = \alpha x$ in R. As $x\alpha = \sigma(\alpha)x$ and $x^n = a$, we have

$$y^n = (\alpha x)^n = \sigma^{n-1}(\alpha)\cdots\sigma(\alpha)\alpha x^n = \mathrm{N}_{K/F}(\alpha)a = 1,$$

so, if $\beta \in K$,

$$y\beta = \alpha x\beta = \alpha\sigma(\beta)x = \sigma(\beta)\alpha x = \sigma(\beta)y.$$

The F-algebra map $R \to (K/F, \sigma, 1)$ sending x to y and fixing K is therefore an isomorphism. It follows by the computation that $R \cong \mathbb{M}_n(F)$.

($\Rightarrow$): Suppose that $R \cong \mathbb{M}_n(F)$. Then R has a simple left R-module M of dimension n over F. As $K[t, \sigma]$ is a left PID, we have $M \cong K[t, \sigma]/K[t, \sigma]f$ for some $f \in K[t, \sigma]$ with $K[t, \sigma]f \supset K[t, \sigma](t^n - a)$. Since

$$n = \dim_K M = (\dim_F K)(\deg f),$$

we must have $\deg f = 1$, i.e., $f = t - c$ for some c in K. As $K[t, \sigma]f \supset K[t, \sigma](t^n - a)$,

$$t^n - a = (b_{n-1}t^{n-1} + \cdots b_1 t + b_0)(t - c),$$

for some $b_0, \ldots, b_{n-1}$ in K. Multiplying out and comparing coefficients yields

$$\begin{aligned} b_{n-1} &= 1 \\ b_{n-2} &= \sigma^{n-1}(c) \\ b_{n-3} &= \sigma^{n-1}(c)\sigma^{n-2}(c) \\ &\vdots \\ b_0 &= \sigma^{n-1}(c)\cdots\sigma(c). \end{aligned}$$

So $a = b_0 c = \sigma^{n-1}(c)\cdots\sigma(c)c = \mathrm{N}_{K/F}(c)$ as needed. [Cf. this proof with that of Hilbert Theorem 90.] □

Let K/F be a cyclic Galois field extension of degree n with Galois group $\langle\sigma\rangle$ and $\bar{\ } : F^\times \to F^\times/\mathrm{N}_{K/F}(F^\times)$ be the canonical map. Wedderburn showed that $(K/F, \sigma, a)$ is a division ring if $\bar{a}$ has order n in $F^\times/\mathrm{N}_{K/F}(K^\times)$. This generalizes Corollary 105.7. A deep theorem in Algebraic Number Theory shows that every division algebra over a number field is a cyclic algebra. This is not true in general.

There is a more general construction for an arbitrary finite Galois extension K/F called a *crossed product algebra* that also produces central simple algebras that we shall study in §107. It is constructed as follows. Let K/F be a finite Galois extension of fields. Let R be a vector space over K of dimension $[K : F]$ on basis $\mathcal{B} := \{u_\sigma \mid \sigma \in G(K/F)\}$. We define multiplication on R as follows: Let σ, τ, η be arbitrary elements in $G(K/F)$ and set

$$\begin{aligned} u_\sigma\beta &= \sigma(\beta)u_\sigma && \text{for all } \beta \in K \\ u_\sigma u_\tau &= \alpha_{\sigma,\tau}u_{\sigma\tau} && \text{for some } \alpha_{\sigma,\tau} \in K \qquad (105.9) \\ \alpha_{\sigma,\tau}\alpha_{\sigma\tau,\eta} &= \sigma(\alpha_{\tau,\eta})\alpha_{\sigma,\tau\eta} && . \end{aligned}$$

The last relation is to guarantee associativity. It also shows that $\alpha_{1.\sigma} = \alpha_{1,1}$ and $\alpha_{\sigma,1} = \sigma(\alpha_{1,1})$ so the one of R is $\alpha_{1,1}^{-1}u_1$. One also checks that R is an F-algebra, i.e., $a(\alpha\beta) = (a\alpha)\beta = \alpha(a\beta)$ for all $a \in F$ and $\alpha, \beta \in R$. It is a fact that, although every equivalence class of central simple algebras contains a crossed product algebra, not every central simple algebra is one. So the theory is still not finished.

Exercises 105.10.

1. Let R be a ring and σ a ring automorphism of R. Define the *twisted power series ring* $R[[t,\sigma]]$ over R, to be the ring with the usual addition of (formal) power series and multiplication induced by $tx = \sigma(x)t$ for all x in F. Show if R is a non-commutative domain, so is $R[[t,\sigma]]$.
2. Let R be a field and σ a ring automorphism of R. Define the *twisted Laurent series ring*
$$R((t,\sigma)) := \{\sum_{-\infty}^{\infty} a_i t^i \mid a_i \in R \text{ with } a_i = 0 \text{ for almost all negative } i\}$$
over R, to be the ring with the usual addition of (formal) Laurent series (i.e., $\sum_{-\infty}^{\infty} a_i t^i + \sum_{-\infty}^{\infty} b_i t^i = \sum_{-\infty}^{\infty} a_i b_i t^i$) and multiplication induced by $tx = \sigma(x)t$ for all x in R. If R is a division ring, show that $R((t,\sigma))$ is also a division ring and never a field if σ is not the identity.
3. Let R be a field and σ a ring automorphism of R of finite order n. Show that the center of $R((t,\sigma))$ is $R((t^n))[= R((t^n, 1_R))]$. In particular, $R((t,\sigma))$ is not finite dimensional over its center.
4. Let F be a field of characteristic zero of of positive degree not dividing n containing a primitive nth root of unity. Suppose that K/F is a cyclic extension of degree n and $t^n - a$ and $t^n - b$ are irreducible polynomials in $F[t]$. Show that $\left(\dfrac{b,a}{F,\zeta^{-1}}\right) \cong \left(\dfrac{a,b}{F,\zeta}\right)$
5. Show equations (105.9(define a ring.
6. Show the ring defined by equations (107.7) is a central simple F-algebra.

106. Central Simple Algebras

In this section we shall study simple F-algebras more carefully, with F a field, under the restriction that the center $Z(A) := \{x \in A \mid yx = xy \text{ for all } y \in A\} \subset A$ of A is F and, furthermore, that A is a finite dimensional F-algebra (as an F-vector space) that we studied in §90. As such an A is left Artinian over F, Wedderburn's Theorem 102.17 applies, so $A \cong \mathbb{M}_n(D)$ for some unique integer n and division F-algebra D, unique up to isomorphism. We use this to study the collection of such algebras and see that it gives an appropriate generalization of algebraic field theory

In this section we shall need to use the definition and properties of tensor products of modules and algebras. Some of these were given as exercises in Exercises 39.12(15)–(23). A full discussion can be found later in the book in Section 119 below. In particular, let A be an R-algebra with R a commutative ring (every ring is an algebra over some commutative ring) and M, N (left) A-modules. Then the tensor product $M \otimes_R N$ is a (left) R-module generated by $\{x \otimes y \mid x \in A,\, y \in B\}$ with the R-action induced by $r(x \otimes y) = rx \otimes y = x \otimes ry$ for all $x \in M$, $y \in N$, and $r \in A$. If M is a left A-module, then $A \otimes_R M$ is canonically isomorphic to M as a left A-module and $M \otimes_R N$ is an A-module. In particular, if B is another R-algebra, then the R-module $B \otimes_R M$ becomes a B-module by $b(b_1 \otimes m) = bb_1 \otimes m$ for all $b, b_1 \in B$, $m \in M$. In addition, $B \otimes_R A$ is an R-algebra with multiplication induced by $(b_1 \otimes a_1)(b_2 \otimes a_2) = b_1 b_2 \otimes a_1 a_2$ for all $a_1, a_2 \in A$, $b_1, b_2 \in B$. For example, if K/F is an extension of fields an F-vector space V induces a K-vector space $K \otimes_F V$ (extension of scalars).

Let A be an R-algebra. Define the *opposite algebra* A^{op} of A as follows: Let $A^{\mathrm{op}} := \{a^{\mathrm{op}} \mid a \in A\}$ and $(\ )^{\mathrm{op}} : A \to A^{\mathrm{op}}$ be the bijection given by $a \mapsto a^{\mathrm{op}}$. Then A^{op} is an R-algebra with the same addition and R-action as A and with multiplication given by $(xy)^{\mathrm{op}} = y^{\mathrm{op}} x^{\mathrm{op}}$. In particular, if M is an left A-module and a right B-module then M is a left $A \otimes_F B^{\mathrm{op}}$-module via $(a \otimes b^{\mathrm{op}})m := amb$ with $r(amb) = r(a \otimes b^{\mathrm{op}})m = (ra \otimes b^{\mathrm{op}})m = (a \otimes rb^{\mathrm{op}})m = a(rm)b$, for all $r \in R$, $a \in A$, $b \in B$, and $m \in M$.

We also make the observation that if A is a simple R-algebra, then the *center* of $Z(A) := \{x \in A \mid yx = xy \text{ for all } y \in A\}$ is a field as $Ax = xA = AxA = A$, for all nonzero $x \in Z(A)$.

Throughout this section F will denote a field.

Definition 106.1. Let F be a field and A a simple F-algebra. We say F is a *central F-algebra* if $F = Z(A)$. We say A is a *finite dimensional F-algebra* if it is a finite dimensional F-vector space.

As mentioned above, the study of finite dimensional central simple F-algebras is bound to the study of finite dimensional central simple division F-algebras by Wedderburns's Theorem 102.17. For example, we used this to investigate cyclic F-algebras, which are finite dimensional central simple F-algebras. We will also be interested when we extend the field F. If K/F is a field extension and A an F-algebra, then the K-algebra $K \otimes_F A$ will be denoted by A^K. So if A is a finite dimensional F-algebra, $\dim_K A^K = \dim_F A$. We begin with the following:

Proposition 106.2. *Let F be a field, A and B two F-algebras. Then*

(1) $Z(A \otimes_F B) = Z(A) \otimes_F Z(B)$.
(2) If A is also central and simple and B is simple, then $A \otimes_F B$ is a simple F-algebra.
(3) If both A and B are central and simple F-algebras, then so is $A \otimes_F B$
(4) If K/F is a field extension and A is a finite dimensional central simple F-algebra, then so is A^K.

Proof. (1): Certainly $Z(A) \otimes_F Z(B) \subset Z(A \otimes_F B)$. Conversely, let $0 \neq z \in Z(A) \otimes_F Z(B)$. Write

$$(*) \qquad z = \sum_{i=1}^{r} a_i \otimes b_i \text{ with } a_i \in A,\, b_i \in B$$

and $\mathcal{A} = \{a_1, \ldots, a_r\}$, $\mathcal{B} = \{b_1, \ldots, b_r\}$.

We may assume that $\mathcal{B}$ is F-linearly independent. In particular, if $x \in A$, then

$$(x \otimes 1)z - z(x \otimes 1) = \sum_{i=1}^{r} (xa_i - a_i x) \otimes b_i.$$

So $xa_i = a_i x$ for all $x \in B$, since $\mathcal{B}$ is linearly independent. Hence, $z \in Z(A) \otimes B$.

We may now assume in (*) that we have chosen the $a_i \in Z(A)$ and $b_i \in \mathcal{B}$ with $\mathcal{A}$ an F-linearly independent set. The same argument then shows that all the b_i lie in $Z(B)$.

(2): Let $0 \neq \mathfrak{A} \subset A \otimes_F B$ be an ideal. Choose

$$0 \neq x = \sum_{i=1}^{r} a_i \otimes b_i \in \mathfrak{A}, \text{ with } a_i \in A,\, b_i \in B \text{ and } r \text{ minimal.}$$

It follows that $\mathcal{A} = \{a_1, \ldots, a_r\}$ and $\mathcal{B} = \{b_1, \ldots, b_r\}$ are F-linearly independent. Since A is simple, $A = Aa_1A$, so there exists an equation $1 = \sum_{j=1}^{s} c_j a_1 d_j$ for some $c_j, d_j \in A$ and s. Set $a_i' = \sum_{j=1}^{s} c_j a_i d_j$ for $i = 2, \ldots, r$. Then

$$x_1 := \sum_{j=1}^{s} c_j x d_j = 1 \otimes b_1 + \sum_{i=2}^{r} a_i' \otimes b_i \text{ lies in } \mathfrak{A}.$$

Since $\mathcal{B}$ is linearly independent, x_1 is not zero. Applying the same argument to Bb_1B produces b_i' with $b_1' = 1$ and

$$x_2 := 1 \otimes 1 + \sum_{i=2}^{r} a_i' \otimes b_i' \text{ lies in } \mathfrak{A},$$

By the minimality of r, $\mathcal{A}' = \{a_1', \ldots, a_r'\}$ and $\mathcal{B}' = \{b_1', \ldots, b_r'\}$ are F-linearly independent. But if $a \in A$,

$$ax_2 - x_2 a = \sum_{i=2}^{r} (aa_i' - a_i' a) \otimes b_i' \text{ lies in } \mathfrak{A}.$$

It follows that $aa_i' = a_i' a$ for all $a \in A$ by the minimallity of r. In particular, $a \in Z(A)$. As $\mathcal{A}$ is linearly independent, we must have $r = 1$.

Statements (3) and (4) follow from (1) and (2). □

Notation 106.3. It is convenient to extend our notation $[K : F] = \dim_F K$ for field extensions K/F. In particular, if A is a finite dimensional central F-algebra and $F \subset D \subset A$ a subalgebra with D a division F-algebra, let $[A : D] := \dim_D A$, the D-dimension of the (left) D-vector space A. Of course, we have $[A : F] = [A : D][D : F]$.

Proposition 106.4. *Let D be a finite dimensional central division F-algebra. Then $[D : F]$ is a square.*

Proof. Let $\widetilde{F}$ be an algebraic closure of F. Then $D^{\widetilde{F}}$ is a finite dimensional central $\widetilde{F}$-algebra by Proposition 106.2. Since $\widetilde{F}$ is algebraically closed, $\widetilde{F}$ is the only finite dimensional division $\widetilde{F}$-algebra. Hence $D^{\widetilde{F}} \cong \mathbb{M}_n(\widetilde{F})$, some n by Wedderburn's Theorem 102.17. Thus $[D : F] = [D^{\widetilde{F}} : \widetilde{F}] = n^2$ □

Corollary 106.5. *Let A be a finite dimensional central simple F-algebra. Then $[A : F]$ is a square.*

Proof. By Wedderburn's Theorem 102.17, $A \cong M_n(D)$ for some finite dimensional division F-algebra D. So $[A : D]$ is a square. As $Z\big(\mathbb{M}_n(D)\big) = Z(D)$, we must have $Z(D) = F$, i.e., D is central. Since $[A : D]$ is a square by the previous result and $[A : F] = [A : D][D : F]$, the result follows. $\square$

The corollary allows us to define an important invariant of finite dimensional central simple F-algebras.

Definition 106.6. Let A be a finite dimensional central simple F-algebra, so $[A : F]$ is a square. The *degree* of A is defined by $\deg A := \sqrt{[A : F]}$.

Definition 106.7. Let A be an R-algebra and $B \subset A$ a subalgebra. We let $Z_A(B) := \{x \in A \mid xb = bx \text{ for all } b \in B\}$, the *centralizer* of B in A. This is an R-algebra.

We can now prove the following basic theorem:

Theorem 106.8. (Double Centralizer Theorem) *Let A be a finite dimensional central simple F-algebra and B a simple subalgebra of A. Then the following are true:*

(1) $Z_A(B)$ is simple.
(2) $B = Z_A\big(Z_A(B)\big)$.
(3) $[A : F] = [B : F][Z_A(B) : F]$.
(4) If B is a finite dimensional central simple F-algebra, then $Z_A(B)$ is a finite dimensional central simple F-algebra and $A = B \otimes_F Z_A(B)$.

Proof. Let $C = Z_A(B)$ and $T = B \otimes_F A^{\mathrm{op}}$. By Proposition 106.2, T is a simple algebra, as A^{op} is clearly one. Let $\mathfrak{A}$ be a minimal left ideal of T and $D = \mathrm{End}_T(\mathfrak{A})$, a division ring by Schur's Lemma 102.16. The F-algebra A becomes a T-module by $(b \otimes a^{\mathrm{op}})x = bxa$ for all $a, x \in A$, $b \in B$. We determine $\mathrm{End}_T(A)$. If $c \in C = Z_A(B)$, then $\lambda_c : A \to A$ given by $a \mapsto ca$ is F-linear as

$$\lambda_c\big((b \otimes a^{\mathrm{op}})(x)\big) = cbxa = bcxa = (b \otimes a^{\mathrm{op}})(cx) = (b \otimes a^{\mathrm{op}})\lambda_c(x),$$

so lies in $\mathrm{End}_T(A)$. Next suppose that $f \in \mathrm{End}_T(A)$. Then

$$(*) \qquad f(a) = f((1 \otimes a^{\mathrm{op}})(1) = (1 \otimes a^{\mathrm{op}})f(1) = \lambda_{f(1)(a)}.$$

It follows that $\mathrm{End}_T(A) = \{\lambda_{f(1)} \mid f \in \mathrm{End}_T(A)\}$. Let $f \in \mathrm{End}_T(A)$. We show $f(1) \in C$. Let $b \in B$. As $B \subset A$, we have

$$bf(1) = (b \otimes 1^{\mathrm{op}})f(1) = f\big((b \otimes 1^{\mathrm{op}})1\big) = f(b) = f(1)b.$$

Therefore, $f(1) \in Z_A(B) = C$, and $\mathrm{End}_T(A) = \{\lambda_c \mid c \in C\}$.

Now A is a finite dimensional F-algebra, so A is a finitely generated T-module. As T is a simple algebra, $A \cong \mathfrak{A}^n$ for some n as a T-module, hence

$$C \cong \mathrm{End}_T(A) \cong \mathrm{End}_T(\mathfrak{A}^n) \cong \mathbb{M}_n(D)$$

is simple. This proves (1).

We also have that $T \cong \mathfrak{A}^r$ some r, so $T \cong \mathbb{M}_r(D)$ and $\mathfrak{A} \subset T$ corresponds to a column space of $\mathbb{M}_r(D)$, i.e., $[\mathfrak{A} : D] = r$. Thus

$$\begin{aligned}
&[A : F] = [A : D][D : F] = nr[D : F] \\
&[A : F][B : F] = [T : F] = [\mathbb{M}_r(D) : F] = r^2[D : F] \\
&[C : F] = n^2[D : F].
\end{aligned}$$

Therefore,

$$[A : F] = \frac{(nr[D : F])^2}{nr[D : F]} = \frac{[A : F][B : F][C : F]}{[A : F]} = [B : F][C : F].$$

Statement (3) now follows.

Finally, we show (4). Clearly, $B \subset Z_A\big(Z_A(B)\big) = Z_A(C)$. By (1), C is simple, so applying (3) to $C \subset A$ yields

$$[A : F] = [C : F][Z_A(C) : F] = [C : F][B : F].$$

So $[Z_A(C) : F] = [B : F]$. It follows that $B = Z_A(C)$. If B is a finite dimensional central simple F-algebra, then $B \otimes_F C$ is simple by Proposition 106.2 and $[B \otimes_F C : F] = [B : F][C : F] = [A : F]$, so the map $B \otimes_F A \to A$ induced by $b \otimes c \mapsto bc$ is an F-algebra isomorphism with

$$F = Z(A) \cong Z(B \otimes_F C) = Z(B) \otimes_F Z(C) = F \otimes_R Z(C) \cong Z(C).$$

Statement (4) now follows. $\square$

We leave as an easy exercise:

Remarks 106.9. Let A be an R-algebra. Then we have R-algebra isomorphisms:
1. $\mathbb{M}_n(R) \otimes_R A \cong \mathbb{M}_n(A) \cong A \otimes \mathbb{M}_n(R)$.
2. $\big(\mathbb{M}_n(A)\big)^{\mathrm{op}} \cong \mathbb{M}_n(A^{\mathrm{op}})$.
3. $\mathbb{M}_n(R) \otimes_R \mathbb{M}_m(R) \cong \mathbb{M}_{nm}(R)$.

Corollary 106.10. *Let A be a finite dimensional central simple F-algebra and $B \subset A$ a simple subalgebra. Then*

(1) $B \otimes_F A^{\mathrm{op}} \cong \mathbb{M}_s\big(Z_A(B)\big)$ with $s = [B : F]$.
(2) $A \otimes_F A^{\mathrm{op}} \cong \mathbb{M}_m(F)$ with $m = [A : F]$.

Proof. In the notation set up in the proof of the Double Centralizer Theorem, we have

$$T = B \otimes_R A^{\mathrm{op}} \cong \mathbb{M}_r(D) \text{ and } Z_A(B) \cong \mathbb{M}_n(D).$$

We also know that $[A : F] = [B : F][Z_A(B) : F]$. so if $s = [B : F]$, we have

$$\begin{aligned}
r^2[D : F] &= [T : F] = [B : F][A : F] \\
&= [B : F]^2[Z_A(B) : F] = s^2n^2[D : F].
\end{aligned}$$

If follows that $r = sn$ and

$$\begin{aligned}
B \otimes_R A^{\mathrm{op}} &\cong \mathbb{M}_r(D) \cong \mathbb{M}_{sn}(D) \cong \mathbb{M}_s(F) \otimes_F \mathbb{M}_n(F) \otimes_F D \\
&\cong \mathbb{M}_s(F) \otimes_F \mathbb{M}_n(D) \cong \mathbb{M}_s(F) \otimes_F Z_A(B) \cong \mathbb{M}_s\big(\mathbb{Z}_A(B)\big).
\end{aligned}$$

This establishes (1) and (2) follows from (1). $\square$

We next turn to subfields in a division F-algebra. We begin with some definitions.

Definition 106.11. Let A be an F-algebra and K/F a field extension. We say

(1) We say that A is *split* over F if $A \cong \mathbb{M}_n(F)$ some integer n.
(2) We say that A is *K-split* or *splits over* K if A^K is split. If this is the case, we call K a *splitting field* for A.
(3) We say K is a *maximal subfield* of A if $K \subset A$ and is a maximal such.
(4) If $A \cong \mathbb{M}_n(D)$, with D a finite dimensional division F-algebra, we let $\operatorname{ind} A := \deg D = \sqrt{[D : K]}$ called the *(Schur) index* of A.

We now show that there exist subfields of a finite dimensional division F-algebra that split A. We first note the following:

Remarks 106.12. 1. Let A be a finite dimensional central simple F-algebra. Then $\operatorname{ind} A \mid \deg A$.

2. If $A \cong \mathbb{M}_n(D)$ with D a finite dimensional division F-algebra and K/F a field extension, then K splits A if and only if it splits D.

Corollary 106.13. *Let D be a finite dimensional division F-algebra and $K \subset D$ be a maximal subfield. Then $[K : F] = [D : K] = \deg D$.*

Proof. Since K is simple, $[D : F] = [K : F][Z_D(K) : F]$ by the Double Centralizer Theorem. We observe that the maximal subfield $K \subset D$ must satisfy $K = Z_D(K)$. For if not, there exists an $a \in Z_D(K) \setminus K$ with $K < K(a)$ a subfield of D, a contradiction. Therefore, $[K : F][K : F] = [D : F] = [D : K][K : F]$, by Proposition 106.4. The result follows. $\square$

Corollary 106.14. *Let A be a finite dimensional central simple F-algebra. If $A \cong \mathbb{M}_n(D)$, D a division F-algebra, K a maximal subfield of D, then* $\operatorname{ind} A = [K : F]$.

If A is a finite dimensional central simple F-algebra, it is possible for a maximal subfield $K \subset A$ to satisfy $\deg A > [K : F]$.

Corollary 106.15. *Every maximal subfield K of a finite dimensional division F-algebra D splits D.*

Proof. We know that $Z_D(K) = K$, so for a unique integer s, $K \otimes_F D^{\mathrm{op}} \cong \mathbb{M}_s(K)$ some s by Corollary 106.10. Hence

$$K \otimes_F D = (K \otimes_F D^{\mathrm{op}})^{\mathrm{op}} \cong \big(M_s(K)\big)^{\mathrm{op}} \cong \mathbb{M}_s(K)$$

as needed. $\square$

In order to use Galois Theory in the study of finite dimensional central simple F-algebras, we need the following crucial proposition.

Proposition 106.16. *Let D be a finite dimensional division F-algebra. Then D contains a maximal subfield that is finite and separable over F. In particular, D contains a separable splitting field.*

PROOF. We may assume that $\operatorname{char} F = p > 0$. We induct on $n = \operatorname{ind} D = \deg D = \sqrt{[D : F]}$. We may assume that $n > 1$.

Claim. There exists $F < E \subset D$ with E/F a separable field extension:

Suppose that this is false. We first show that every element $x \in D$ is purely inseparable over F, i.e., x^{p^e} lies in F some integer e. Indeed if $x \in D$, then the minimal polynomial $m_F(x) = f(t^{p^e})$ in $F[t]$ for some separable polynomial $f \in F[t]$. Our assumption implies that f splits over F, hence x^{p^e} lies in F, and every element in D is purely inseparable over F. In particular, by Proposition 106.4, this means that for all $x \in D \setminus F$, we have $n^2 = [D : F] = [D : F(x)][F(x) : F]$. Hence $p \mid n$.

Let K be a maximal subfield of D. By the last corollary, K splits D, so there exists a K-algebra isomorphism $\varphi : K \otimes_F D \to \mathbb{M}_n(K)$. Let $x \in D \setminus F$. As x is purely inseparable over F, $a = x^{p^e}$ lies in F for some e. In particular,

$$\varphi\big(1 \otimes x)\big)^{p^e} = \varphi(1 \otimes x^{p^e}) = \varphi(1 \otimes a) = a\mathrm{I},$$

(where I is the identity matrix). Let $\widetilde{K}$ be an algebraic closure of K. Then $\big(\varphi(1 \otimes x) - a^{1/p^e} I\big)^{p^e} = 0$. In particular, all the eigenvalues of $\varphi(1 \otimes x)$ must be equal. Since $p \mid n$, the trace $\operatorname{tr}\big(\varphi(1 \otimes x)\big) = 0$ in K (since in $\widetilde{K}$). But the set $\{\alpha \in M_n(K) \mid \operatorname{tr}(\alpha) = 0\}$ cannot span $\mathbb{M}_n(K)$ as the trace is K-linear. This is a contradiction and establishes the claim.

Let $F < E \subset D$ with E/F separable and $D' = Z_D(E)$ the subdivision F-algebra of D. We next show that $E = Z(D')$. By the Double Centralizer Theorem 106.8, $E = Z_D(D')$, so $Z(D') \subset E$. Conversely, if $x \in E$, then $D' = Z_D(E)$ means that $xd' = d'x$ for all $d' \in D'$. Thus $E = Z_D(D')$. It follows that D is a finite dimensional division E-algebra. By the Double Centralizer Theorem 106.8, we have

$$\begin{aligned}[D : F] &= [D' : F][Z_D(D') : F] = [D' : F][E : F] \\ &= [D' : E][E : F]^2 > [D' : E]\end{aligned}$$

by the claim. Hence by induction, there exists a maximal subfield L of D' with L/E separable. In particular, L/F is also separable. To finish it suffices to show that L is a maximal subfield of D. If this is not the case, then $L < Z_D(L)$. But $Z_D(L) \subset Z_D(E) = D'$. This contradicts the maximality of L. □

Corollary 106.17. *Let A be a finite dimensional central simple F-algebra. Then A has a separable splitting field finite over F. In particular, A also has a Galois splitting field finite over F.*

PROOF. Let $A \cong \mathbb{M}_n(D)$ with D a finite dimensional division F-algebra. Then D contains a maximimal subfield K that is separable over F. Thus K splits D, hence A. Let L/K be the normal closure of K/F. Then L/F is Galois and L splits A. □

Perhaps the most important basic theorem is the following:

Theorem 106.18. (Skolem-Noether Theorem) *Let A be a simple, left Artinian central F-algebra and $B \subset A$ a finite dimensional simple F-subalgebra. If $f, g : B \to A$ are F-algebra maps, then there exists a unit $u \in A^\times$ satisfying*

$$f(b) = u^{-1} g(b) u \quad \textit{for all } b \in B.$$

PROOF. By Proposition 106.2, $C = B \otimes_F A^{\mathrm{op}}$ is a simple F-algebra, and by Wedderburn's Theorem 102.17, there exists a division F-algebra D such that $A \cong \mathbb{M}_n(D)$ for some n. Then

$$C \cong B \otimes_F \big(\mathbb{M}_n(D)\big)^{\mathrm{op}} \cong B \otimes_F \mathbb{M}_n(D^{\mathrm{op}}) \cong \mathbb{M}_n(B \otimes_F D^{\mathrm{op}})$$

is left Artinian, since it is a finite dimensional D^{op}-algebra with D^{op} a division ring. Let M be an irreducible C-module, so unique up to isomorphism with C-action induced by $(b \otimes a^{\mathrm{op}})(m) = bma$ for all $b \in B$, all $a \in A$, and all $m \in M$. Define C-modules A_1 and A_2 as follows: $A_1 = A$ with C-action induced by

$$(b \otimes a^{\mathrm{op}})(x) = f(b)xa \text{ for all } b \in B \text{ and all } a, x \in A_1$$

and $A_2 = A$ with C-action induced by

$$(b \otimes a^{\mathrm{op}})(x) = g(b)xa \text{ for all } b \in B \text{ and all } a, x \in A_2.$$

Then there exist unique integers n_i such that $A_i \cong M^{n_i}$ as C-modules, since A_i is a finitely generated C-module for $i = 1, 2$. Since A is a left D-vector space, M is a right D-vector space. Therefore,

$$n_1 = \frac{[A : D]}{[M : D]} = n_2$$

as $[A_i] = [A : D]$ for $i = 1, 2$. It follows that $A_1 \cong A_2$ as C-modules. Let $\varphi : A_1 \to A_2$ be a C-isomorphism. Then we have $\varphi\big((b \otimes a^{\mathrm{op}})\big)(x) = (b \otimes^{\mathrm{op}})\varphi(x)$ for all $b \in B, a, x \in A$, i.e.,

$$(*) \qquad \varphi\big((f(b)xa\big) = g(b)\varphi(x)a \text{ for all } a, x \in A,\, b \in B.$$

Let $u = \varphi(1)$. Setting $b = 1 = x$ in (*) shows that $f(1) = 1 = g(1)$. Consequently, $\varphi(x) = \varphi(1)x = ux$ for all $x \in A$. As φ is surjective, there exists a $v \in A$ satisfying $uv = 1$. Since $A \cong \mathbb{M}_n(D)$, we have $u \in A^\times$. Setting $x = 1 = a$ in (*), we see that $uf(b) = \varphi\big(1 \cdot f(b)\big) = \varphi\big(f(b) \cdot 1\big) = g(b)u$ for all $b \in B$ as $f(b) \in A$ and φ is C-linear. The result follows. □

Corollary 106.19. *Let A be a finite dimensional central simple F-algebra. Then every F-algebra endomorphism of A is an inner A-automorphism.*

PROOF. Any A-algebra homomorphism takes 1_A to 1_A, so is not trivial. □

Corollary 106.20. *Let A be a finite dimensional central simple F-algebra. They any two isomorphic simple subalgebras of A are conjugate and, in particular, have conjugate centralizers.*

As an application of Skolem-Noether, we give an alternative proof of Wedderburn's Theorem about the commutativity of finite division rings.

Lemma 106.21. *Let D be a finite dimensional division F-algebra with K a maximal subfield of D. Then all maximal subfields of D are isomorphic to K as F-algebras if and only if K/F is separable and $D = \bigcup_{x \in D^\times} x^{-1}Kx$.*

Proof. ($\Rightarrow$): By Proposition 106.16, D contains a separable maximal subfield. As all maximal subfields are isomorphic by assumption, all maximal subfields are separable over F. Since every $x \in D$ lies in some maximal subfield E and $E \cong K$ as F-algebras, we have $E = x^{-1}Kx$ for some $x \in D^{\times}$ by the Skolem-Noether Theorem. The result follows.
($\Leftarrow$): Let $y \in D$. By hypothesis, there exists an element $x \in D^{\times}$ such that $y \in x^{-1}Kx$. Thus y is separable over F for all $y \in D^{\times}$, hence all maximal subfields in D are separable. If L is a maximal subfield of D, then $L = F(y)$ for some $y \in D^{\times}$ by the Primitive Element Theorem 57.9 and $y \in x^{-1}Kx$, for some $x \in D^{\times}$. Thus $L = F(y) = x^{-1}Kx$ by maximality. □

Theorem 106.22. (Wedderburn) *Every finite division ring D is a field.*

Proof. Let $F = Z(D)$, so D is a finite dimensional central division F-algebra, say of degree n. Then all maximal subfields of D have $|F|^n$ elements, and there is precisely one up to isomorphism, say K. By the lemma, $D^{\times} = \bigcup_{x \in D^{\times}} x^{-1}K^{\times}x$. If $D^{\times} \neq E^{\times}$, then the number of conjugates $x^{-1}K^{\times}x$ of $K^{\times}$, $x \in D^{\times}$, is at most $[D^{\times} : F^{\times}]$. Since $1 \in x^{-1}K^{\times}x$, for all $x \in D^{\times}$, we have $|\{x^{-1}hx \mid x \in D^{\times},\, h \in K^{\times}\}| < D^{\times}$, a contradiction. So $D = K$, hence $D = F$ □

We saw in Section 101 that generalized quaternion algebras have norms attached to them and are useful in deciding whether they were division rings or matrix rings. Using the Skolem-Noether Theorem, we can define a norm called the *reduced norm* on any finite dimensional central simple F-algebra. We now construct it.

Construction 106.23. Let A be a finite dimensional central simple F-algebra. We know by Corollary 106.17 that A has a separable splitting field K, finite over F. Let $\deg A = n$ and $\varphi : A^K \to \mathbb{M}_n(K)$ be an K-algebra isomorphism. If $\alpha \in \mathbb{M}_n(F)$, let $f_\alpha := \det(tI - \alpha)$ denote the characteristic polynomial of α in $K[t]$. If $a \in A$, define the *reduced characteristic polynomial* of a by $f_{\varphi(a)} \in K[t]$ over K. This is independent of the isomorphism φ by the Skolem-Noether Theorem 106.18. If $f_{\varphi(a)} = t^n + c_{n-1}t^n + \cdots + c_0 \in K[t]$, define the *reduced norm* of $a \in A$ over K by $\mathrm{Nrd}(a) := (-1)^n c_0$ and the *reduced trace* of a over K by $\mathrm{Trd}(a) := -c_0$ over K.

We want to show that the reduced norm of a finite dimensional central simple F-algebra is independent of the separable splitting field K and takes values in F. We begin by looking at the reduced characteristic polynomial.

Lemma 106.24. *Let A be a finite dimensional central simple F-algebra of degree n. Then the reduced characteristic polynomial of an element in A is independent of the finite separable splitting field of A over F. In particular, we can evaluate the reduced characteristic polynomial of any element using a finite Galois extension of F that splits A.*

Proof. The algebra A has a separable splitting field K, finite over F. Let $\widetilde{F}$ be an algebraic closure of F containing K. If L is another separable splitting field of A over F in $\widetilde{F}$, then we have $E = L(K) = K(L)$ is also a finite separable splitting field of A. It follows that the reduced characteristic polynomials of $\alpha \in A$ over K and over L are the same as they are the same over $\widetilde{F}$, since $\mathbb{M}_n(E) \subset \mathbb{M}_n(\widetilde{F})$ for all intermediate fields $\widetilde{F}/E/F$. Therefore, the reduced characteristic polynomial of a is independent of separable splitting field. In particular, we may assume that L/F is finite Galois with $L/K/F$. □

We now look at the action of a finite Galois group of a Galois splitting field on a finite dimensional central simple F-algebra.

Proposition 106.25. *Let A be a finite dimensional central simple F-algebra of degree n. Then $f_{\varphi(a)} \in F[t]$ for all $a \in A$.*

PROOF. Let K/F be a finite separable field extension with A split over K. Let $L/K/F$ be a finite field extension with L/F is Galois. For all $\sigma \in G(L/F)$, the map $\sigma \otimes 1_A : A^L \to A^L$ induced by $x \otimes a \mapsto \sigma(x) \otimes a$ for $x \in L$, and $a \in A$ is a K-algebra isomorphism. Let the map $\mathbb{M}_n(L) \to \mathbb{M}_n(L)$ be defined by taking the matrix $\big(x_{ij}\big)$ to $\big(\sigma(x_{ij})\big)$. Denote this map by σ also. We have a commutative diagram

$$\begin{array}{ccc} A^L & \xrightarrow{\sigma\otimes 1_A} & A^L \\ {\scriptstyle\varphi}\big\downarrow & & \big\downarrow{\scriptstyle\psi} \\ \mathbb{M}_n(L) & \xrightarrow{\sigma} & \mathbb{M}_n(L). \end{array}$$

with $\psi = \sigma\varphi(\sigma \otimes 1_A)^{-1}$, a ring isomorphism. Since

$$\psi(x \otimes 1) = \sigma\varphi\big(\sigma^{-1}(x) \otimes 1\big) = \begin{pmatrix} x & & \\ & \ddots & \\ & & x \end{pmatrix}$$

for all $x \in L$, it follows that ψ is an L-isomorphism. If $a \in A$, we have

$$\psi(1 \otimes a) = \sigma\varphi\big((\sigma \otimes 1_A)^{-1}\big)(1 \otimes a) = \sigma\big(\varphi(1 \otimes a)\big).$$

By the Skolem-Noether Theorem 106.18, $\varphi(1\otimes a)$ and $\psi(1\otimes a)$ have the same characteristic polynomials, so its coefficients lie in F. The result follows. □

Corollary 106.26. *Let A be a finite dimensional central simple F-algebra of degree n. Then for all $a, b \in A$ and all $x \in F$, we have*

1. *The reduced norm of A satisfies* $\mathrm{Nrd} : A \to F$. *Moreover,* $\mathrm{Nrd}(ab) = \mathrm{Nrd}(a)\,\mathrm{Nrd}(b)$ *and* $\mathrm{Nrd}(xa) = x^n\,\mathrm{Nrd}(a)$.
2. *An element $a \in A$ is a unit if and only if* $\mathrm{Nrd}(a)$ *is nonzero. In particular,* $\mathrm{Nrd} : A^\times \to F^\times$ *is a group homomorphism.*
3. *The reduced trace of A satisfies* $\mathrm{Trd} : A \to F$ *is F-linear. Moreover, satisfies* $\mathrm{Trd}(ab) = \mathrm{Trd}(ba)$.

PROOF. By the proposition, all the coefficients of f_a, $a \in A$ lie in F. In particular, $\mathrm{Nrd}(a)$ and $\mathrm{Trd}(a)$ lie in F for all $a \in A$. The other statements follow as the same properties hold for the corresponding properties for the determinant and trace of matrices. □

Remarks 106.27. Let A be a finite dimensional central simple F-algebra of degree n and K/F a finite separable field extension with A split over K. Let $\mathcal{B} = \{a_1, \dots a_{n^2}\}$ be a basis for A and $\varphi : A^K \to \mathbb{M}_n(K)$ an F-algebra isomorphism. Then $\mathrm{Nrd}(\sum_{i=1}^{n^2} x_i a_i) = \det\big(\sum_{i=1}^{n^2} x_i\varphi(a_i)\big)$ defines a polynomial $Nrd_A \in F[t_1, \dots, n^2]$. As we are taking a determinant, this is a homogeneous polynomial of degree n.

We arrive at the result that we sought:

Theorem 106.28. (Tsen) *Let F be a field of transcendence degree one over an algebraically closed field. Then every finite dimensional central simple F-algebra is isomorphic to a matrix ring over F.*

PROOF. If A is a finite dimensional central simple F-algebra of degree n, then the norm form Nrd_A is a homogeneous polynomial of degree n. As F is a C_1-field by Corollary 101.17, it has a nontrival zero if $n > 1$, hence cannot be a matrix ring over a noncommutative division F-algebra. $\square$

Exercises 106.29.
F is a field in all the exercises.

1. Let A be a finite dimensional F-algebra and K/F a finite field extension. Prove that A is a finite dimensional central simple F-algebra if and only if A^K is a finite dimensional central simple F-algebra.
2. Let A be a finite dimensional F-algebra. Show that A is a finite dimensional central simple F-algebra if and only if there exists a finite extension K/F such that A splits over K.
3. Let $\mathfrak{A}$ be a nonzero left ideal in a finite dimensional simple F-algebra A. Set $D = \mathrm{End}_A(\mathfrak{A})$. Show the map $\lambda_F : A \to \mathrm{End}_D(\mathfrak{A})$ defined by $a \mapsto \lambda_a : x \mapsto ax$ is an ring isomorphism.
4. The F-subalgebras in $\mathbb{M}_n(F)$ that are isomorphic to F^n are conjugate to the subalgebra of diagonal matrices.
5. Show a finite dimensional central simple F-algebra A of degree n is split if and only if it contains a F-subalgebra isomorphic to the direct product $F^n = F \times \cdots \times F$.
6. Let D be a central division F-algebra. If $a, b \in D$ are algebraic over F with the same minimal polynomial over F, show that $b = xax^{-1}$ for some $x \in D^\times$.
7. Let D be a central division F-algebra with $\mathrm{char}\, F = p > 0$. Let $a \in D \setminus F$ satisfy $a^{p^n} = a$ for some positive integer n. Show that there exists an element $x \in D^\times$ satisfying $xax^{-1} = a^i \neq a$ for some i.
8. Let A be an F-algebra. An *F-derivation* $\delta : A \to A$ is an F-linear map satisfying $\delta(ab) = a\delta(b) + \delta(a)b$. Suppose that A is a central simple F-algebra and $\delta : A \to A$ is an F-derivation. Show the following:
   (a) The subalgebras
   $$B = \{\begin{pmatrix} a & \delta(a) \\ 0 & a \end{pmatrix} \mid a \in A\} \text{ and } B = \{\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in A\}$$
   of $\mathbb{M}_2(A)$ are isomorphic F-algebras.
   (b) $\delta : A \to A$ is an *F-inner derivation.*, i.e., there exists an element $c \in A$ such that $\delta(x) = xc - cx$ for every x in A.
9. Let D be a finite dimensional central division F-algebra with F a C_2-field. Show $\mathrm{Nrd} : D \to F$ is surjective.

107. The Brauer Group

To further study the theory of finite dimensional central simple F-algebras, it is useful to give a structure arising from the set of isomorphism classes of such algebras. We shall show that the tensor product of finite dimensional central simple F-algebras is also a finite dimensional central simple F-algebra. Since any finite dimensional central simple F-algebra is a matrix ring over a central division F-algebra, unique up to isomorphism, the question arises about the underlying division ring in the tensor product of two such central simple algebras. We construct a group called the Brauer group of F group whose group structure is induced by the tensor of such algebras. This group is a torsion group and is an important invariant in the study of fields. Historically, it was crucial in the interpretation of parts of number theory (class field theory).

Construction 107.1. Let A_i be a finite dimensional central simple F-algebra with $A_i \cong \mathbb{M}_{n_i}(D_i)$, D_i a central division F-algebra for $i = 1, 2$. We say that A_1 is *(Brauer) equivalent to* A_2 if $D_1 \cong D_2$ as F-algebras. Let $[A_i]$ denote the (Brauer) equivalence class of A_1. Since $A_1 \otimes_F A_2$ is a finite dimensional central simple F-algebra and $A_1 \otimes_F A_1^{\mathrm{op}} \cong \mathbb{M}_{n_1^2}(F)$, we have

$$\mathrm{Br}(F) := \{[A] \mid A \text{ is a finite dimensional central simple } F\text{-algebra}\}$$

is an abelian group via $[F] = 1$, $[A_1][A_2] := [A_1 \otimes_F A_2]$, and $[A_1]^{-1} = [A_1^{\mathrm{op}}]$, called the *Brauer Group* of F. Let K/F be a field extension. Then

$$(A_1 \otimes_F A_2)^K = K \otimes_F A_1 \otimes_F A_2 \cong (K \otimes_F A_1) \otimes_K (K \otimes_F A_2) \cong A_1^K \otimes_F A_2^K$$

and

$$(A_1)^{K^{\mathrm{op}}} = (K \otimes_F A_1)^{\mathrm{op}} \cong K \otimes_F A_1^{\mathrm{op}} = (A_1^{\mathrm{op}})^K.$$

Therefore, we have a group homomorphism $\varphi : \mathrm{Br}(F) \to \mathrm{Br}(K)$ given by $[A] \mapsto [A^K]$. Let

$$\begin{aligned}\mathrm{Br}(K/F) := \ker\varphi = \{[A] \mid A\, &\text{is a finite dimensional central simple}\\ &F\text{-algebra with } A^K \text{ split}\}.\end{aligned}$$

Let $\widetilde{F}$ be a fixed algebraic closure of F and F_{sep} denote the *separable closure* of F in $\widetilde{F}$, i.e., the maximal separable extension of F in $\widetilde{F}$. The Galois group of $G(\widetilde{F}/F) = G(F_{\mathrm{sep}}/F)$ called the *absolute Galois group* of F. Since every finite dimensional central simple F-algebra has a finite separable splitting field, we saw that it has a finite Galois splitting field. It follows that

$$\mathrm{Br}(F) \;=\; \bigcup_{\substack{K/F \\ \text{finite separable}}} \mathrm{Br}(K/F) \;=\; \bigcup_{\substack{L/F \\ \text{finite Galois}}} \mathrm{Br}(L/F)$$

and $\mathrm{Br}(F) = \mathrm{Br}(F_{\mathrm{sep}}/F)$.

Proposition 107.2. *Let D be a finite dimensional central division F-algebra with* $\deg D = q$ *and K/F a finite field extension. Then*

(1) If D splits over K, then $q \mid [K : F]$.

(2) There exists a smallest positive integer r such that K can be embedded in $\mathbb{M}_r(F)$.

(3) *Let r be as in* (2). *Then D splits over K if and only if K is a maximal subfield of $\mathbb{M}_r(D)$.*

(4) *Let r be as in* (2) *and $E = Z_{M_r(D)}(K)$. Then E is a division F-algebra. Moreover, K is a maximal subfield of $\mathbb{M}_r(D)$ if and only if $E = K$.*

PROOF. Let $A = D^K$, a finite dimensional central simple F-algebra, and M an irreducible A-module. Then M is a D-vector space by $dm = (1 \otimes d)m$ for all $d \in D$, $m \in M$. Let $r = [M : D]$. We have

$$(k \otimes 1)(1 \otimes d) = k \otimes d = (1 \otimes d)(k \otimes 1) \text{ for all } k \in K,\, d \in D.$$

It follows that $\lambda_k : M \to M$ determined by $m \mapsto (k \otimes 1)m$ is D-linear for all $k \in K$, hence $K = K \otimes_F F \subset \mathrm{End}_D(M) = \mathbb{M}_r(D)$, where we view M as a right $\mathrm{End}_D(M)$-module. [We do this so that we do not get $\mathbb{M}_r(D^{\mathrm{op}})$.]

Claim 1. The above r satisfies (2):

Suppose that $K \subset \mathbb{M}_s(D)$. Choose N a D-vector space of dimension s such that $K \subset \mathrm{End}_D(N) \cong \mathbb{M}_s(D)$, where we write linear operators on the right, matrices as usual. Let N be an A-module via $(k \otimes d)v := dv\lambda_k$ for all $k \in K$, $d \in D$, and $v \in N$. Since A is a finite dimensional central simple F-algebra, M is irreducible, so $N \cong M^n$ for some n, hence

$$\mathrm{End}_D(N) \cong \mathrm{End}_D(M^n) \cong \mathbb{M}_{rn}(D), \quad i.e., \quad r \mid s.$$

This proves Claim 1.
Let $B = \mathrm{End}_B(M)$ and $E = Z_B(K)$, so $[B : D] = r^2$.

Claim 2. $E = \mathrm{End}_A(M)$:

Let $k \in D$, $m \in M$, and $f \in \mathrm{End}_A(M)$. Then

$$mf\lambda_k = (d \otimes 1)(mf) = \big((k \otimes 1)m\big)f = m\lambda_k f,$$

so $f \in E$. Conversely, if $f \in E$, then for all $k \in K$, $m \in M$, and $d \in M$,

$$\begin{aligned}\big((k \otimes d)m\big)f &= \big((k \otimes 1)(1 \otimes d\big)f = \big((1 \otimes d)m\lambda_k\big)f \\ &= (1 \otimes d)(mf)\lambda_k = (k \otimes 1)(1 \otimes d)(mf) = (k \otimes d)(mf),\end{aligned}$$

so $f \in \mathrm{End}_A(M)$ and Claim (2) is established.

We next show (4) holds. By Schur's Lemma, E is a division algebra. Suppose that $K = E$. If $K \subset L \subset B$, with L a field, then $L \subset Z_B(K) = E = K$. Therefore, K is a maximal subfield of B. Conversely, if $K < E$, then there exits an element $x \in E \setminus K$, so $K < K(x) \subset B$. But $K(x)$ is a field, so K is not a maximal subfield of B, a contradiction. This establishes (4).

We have $q = \deg D = \sqrt{[D : F]}$ and $[B : F] = [B : D][D : F]$. So by the Double Centralizer Theorem 106.8

$$(*) \qquad [K : F][E : F] = [B : F] = r^2 q^2.$$

We now show (1). suppose that K splits D, hence $A = D^K = \mathbb{M}_q(K)$ and $A \cong M^q$ as an A-module (as M corresponds to a column space of $\mathbb{M}_q(K)$. Therefore,

$$\begin{aligned}[A : F] &= [A : D][D : F] = q[M : D][D : F] = q^2 r \\ &= [A : K][K : F] = [D^K : K][K : F] = [D : F][K : F].\end{aligned}$$

Therefore,

$$qr = [A : D] = [K : F]. \tag{†}$$

which is (1).

Lastly, we prove (3).

($\Rightarrow$): We have $K \subset E$ and by (†), $qr = [K : F]$. Therefore, by (*), we have $[E : F] = qr$. It follows that $E = K$.

($\Leftarrow$): Suppose that K is a maximal subfield of B. We must show that D splits over K. We know that $K = E = Z_B(K)$ by (4). By Corollary 106.10 of the Double Centralizer Theorem 106.8, we know that

$$K \otimes_K B^{\text{op}} \cong \mathbb{M}_s\big(Z_B(K)\big) = \mathbb{M}_s(E) = \mathbb{M}_s(K)$$

with $s = [K : F]$. Hence $K \otimes B \cong \mathbb{M}_s(K)$ also. Consequently, B splits over K, hence so does D. □

Remark 107.3. Let D be a finite dimensional division F-algebra and K/F a finite field extension. Suppose that D splits over K. Then the proposition above shows that there exists a finite dimensional central simple R-algebra B such that $[B] = [D]$ in the Brauer group $\mathrm{Br}(F)$ satisfying

1. $K \subset B$ is a maximal subfield.
2. $[K : F]^2 = [B : F]$.
3. $K = Z_B(K)$.

We call such a K a *self centralizing maximal subfield* of B. [If B is not a division F-algebra, a maximal subfield L of B may not satisfy $Z_B(L) = L$.]

Corollary 107.4. *Let K/F be a field extension and A a finite dimensional central simple F-algebra satisfying $[K : F]^2 = [A : F]$. Suppose that A splits over K. Then there exists an embedding of F-algebras $K \hookrightarrow A$ of F-algebras with (the image of) K a self centralizing maximal subfield of A.*

Proof. By the Remark, there exists a finite dimension central simple F-algebra B satisfying $[B] = [A]$ in $\mathrm{Br}(F)$ and K is a self centralizing maximal subfield of B. Therefore, $[B : F] = [K : F]^2 = [A : F]$. It follows that $A \cong B$ as F-algebras. □

Corollary 107.5. *Let A be a finite dimensional central simple F-algebra. Then there exists a finite Galois extension L/F such that L splits A and L is a self centralizing maximal subfield of a finite dimensional central simple F-algebra B with B Brauer equivalent to A.*

Proof. We know that there exists a finite Galois extension L/F such that A splits over L. Consequently, L splits D if $A \cong \mathbb{M}_;(D)$, D a division F-algebra. The Remark shows that L is a self centralizing maximal subfield of a finite dimensional central simple F-algebra B satisfying B is Brauer equivalent to D in $\mathrm{Br}(F)$. The result follows. $\square$

We now investigate a group isomorphic to the Brauer group. We begin by generalizing our construction of the cyclic F-algebras in Section 105 that we mentioned in that section. We use the following

Notation 107.6. If a (multiplicative) group acts on a field L, we shall denote $\sigma(x)$ by ${}^{\sigma}x$ for $x \in L$ and $\sigma \in G$ when convenient.

Let L/F be a finite Galois extension. We now look (again) at the F-algebra called a *crossed product algebra* A mentioned in Section 105. We leave many details left to the reader.

Construction 107.7. Let L/F be a finite Galois extension and A be a vector space over L of dimension $[L : F]$ on basis $\mathcal{B} := \{u_\sigma \mid \sigma \in G(L/F)\}$. So $A = \coprod_{\sigma \in G(L/F)} u_\sigma$. We define multiplication on A as follows: Let σ, τ, η be arbitrary elements in $G(L/F)$ and set

1. $u_\sigma x = \sigma(x) u_\sigma = {}^{\sigma}x u_\sigma$ for all $x \in L$ defines an L-action on A.
2. $u_\sigma u_\tau = f(\sigma, \tau) u_{\sigma\tau}$ for some $f(\sigma, \tau) \in L^\times$.

In particular, $f : G(L/F) \times G(L/F) \to L^\times$ and f commutes with all elements in F. We define multiplication in A by

$$\Big(\sum_{\sigma \in G(L/F)} x_\sigma u_\sigma\Big)\Big(\sum_{\tau \in G(L/F)} y_\tau u_\tau\Big) = \sum_{\sigma, \tau \in G(L/F)} x_\sigma {}^{\sigma}y_\tau f(\sigma, \tau) u_{\sigma\tau}.$$

In order for A to be associative, we must have

$$u_\rho(u_\sigma u_\tau) = (u_\rho u_\sigma) u_\tau$$

for all $\rho, \sigma, \tau \in G(L/F)$.
As

$$\begin{aligned} u_\rho(u_\sigma u_\tau) &= u_\rho f(\sigma, \tau) u_{\sigma\tau} \\ &= {}^{\rho}f(\sigma, \tau) u_\rho u_{\sigma\tau} = {}^{\rho}f(\sigma, \tau) f(\rho, \sigma\tau) u_{\rho\sigma\tau}. \end{aligned}$$

and

$$(u_\rho u_\sigma) u_\tau = f(\rho, \sigma) u_{\rho\sigma} u_\tau = f(\rho, \sigma) f(\rho\sigma, \tau) u_{\rho\sigma\tau}$$

for all $\rho, \sigma, \tau \in G(L/F)$, the map f must satisfy

$$ {}^{\rho}f(\sigma, \tau) f(\rho, \sigma\tau) = f(\rho, \sigma) f(\rho\sigma, \tau) \tag{107.8}$$

for all $\rho, \sigma, \tau \in G(L/F)$. Check that under these conditions that A is an F-algebra.

Definition 107.9. Let L/F be a finite Galois extension. Call a map $f : G(L/F) \times G(L/F) \to L^\times$ satisfying equation (107.8) a *factor set*. The algebra A in the construction above is called a *crossed product algebra* on f, and is denoted $(L/F, f)$. We also call $\{u_\sigma \mid \sigma \in G(: L/F)\}$ a *canonical basis* for $(L/F, f)$

Example 107.10. Every cyclic F-algebra is a crossed product algebra. [Cf. the special case in Example 105.4(3).]

Construction 107.11. Let $f, g := G(L/F) \times G(L/F) \to L^\times$ with L/F a finite Galois extension. Define $fg : G(L/F) \times G(L/F) \to L^\times$ by

$$(fg)(\sigma, \tau) := f(\sigma, \tau)g(\sigma, \tau) \text{ for all } \sigma, \tau \in G(L/F).$$

The set

$$Z^2\big(G(L/F), L^\times\big) := \{f : G(L/F) \times G(L/F) \to L^\times \mid f \text{ a factor set}\}$$

is an abelian group with identity the *trivial factor set* 1 defined by $1(\sigma, \tau) = 1$ for all $\sigma, \tau \in G(L/K)$.

Let $c : G(K/F) \to L^\times$. Define

$$\delta c : G(K/F) \times G(K/F) \to L^\times \text{ by } (\delta c)(\sigma, \tau) := c(\sigma)\, {}^\sigma c(\tau)\, c(\sigma\tau)^{-1}.$$

Check. δc lies in $Z^2\big(G(L/F), L^\times\big)$.
It is called a *Principal Factor Set.* Set

$$B^2\big(G(L/F), L^\times\big) := \{\delta c \mid c : G(L/F) \to L^\times\},$$

a subgroup of $Z^2\big(G(L/F), L^\times\big)$.
The quotient group

$$H^2\big(G(L/F), L^\times\big) := Z^2\big(G(L/F), L^\times\big)/B^2\big(G(L/F), L^\times\big)$$

is called the $2nd$ *cohomology group* of $G(L/F)$ with coefficients in $L^\times$. Denote the class of $f \in Z^2\big(G(L/F), L^\times\big)$ in this quotient by $[f]$. If $[f] = [g]$ with $f, g \in Z^2\big(G(L/F), L^\times\big)$, write $f \sim g$.

Check. If $f \sim g$, then $(L/F, f) \to (L/F, \big((\delta c)f\big)$ given by $c(\sigma)u_\sigma \mapsto v_\sigma$, where $\{u_\sigma \mid \sigma \in G(L/F)\}$ and $\{v_\sigma \mid \sigma \in G(L/F)\}$ are invariant bases for $(L/F, f)$ and $\big(L/F, (\delta c)f\big)$, respectively, is an F-algebra isomorphism.

Let L/F be a finite Galois extension. We say $f \in Z^2\big(G(L/F), L^\times\big)$ is a *normalized factor set* if $f(\sigma, 1) = 1 = f(1, \sigma)$ for all $\sigma \in G(L/F)$.

Lemma 107.12. *Let L/F be a finite Galois extension and f an element in $Z^2\big(G(L/F), L^\times\big)$. Then there exists a normalized factor set $g \in Z^2\big(G(L/F), L^\times\big)$ satisfying $f \sim g$. In particular, $(L/F, f) \cong (L/F, g)$ as F-algebras.*

Proof. Define $c : G(L/F) \to L^\times$ by

$$c(\sigma) = \begin{cases} f(1,1)^{-1} & \text{if } \sigma = 1 \\ 1 & \text{if } \sigma \neq 1. \end{cases}$$

Then $(\delta(c)(\sigma, 1) = {}^\sigma c(1)$ for all σ in $G(L/F)$. Check if $g = (\delta c)f$, then g is a normalized factor set. □

Now suppose that $A = (L/K, f)$ with f a normalized factor set, $\{u_\sigma \mid \sigma \in G(L/F)\}$ a canonical basis. Then we have $u_\sigma \in A^\times$ for all $\sigma \in G(L/F)$ and identifying L and Lu_1 that $u_1 = 1_A$. [If f is not normalized, we still have $u_\sigma \in A^\times$ for all $\sigma \in G(L/F)$.]

Proposition 107.13. *Let L/F be a finite Galois extension and f an element in $Z^2\big(G(L/F), L^\times\big)$ a normalized factor set. Let $A = (L/F, f)$. Then A is a finite dimensional central simple F-algebra and L is a self centralizing maximal subfield of A.*

Proof. Let $\{u_\sigma \mid \sigma \in G(L/F)\}$ be a canonical basis for f, so $u_1 = 1$. We know that $F \subset Z(A)$, i.e., A is an F-algebra. Let $z = \sum_{G(L/F)} a_\sigma u_\sigma \in Z_A(L)$ with all $a_\sigma \in L$. Then for all $x \in L$, we have

$$\sum_{G(L/F)} x a_\sigma u_\sigma = xz = zx = \sum_{G(L/F)} a_\sigma u_\sigma x = \sum_{G(L/F)} a_\sigma \sigma(x) u_\sigma.$$

Therefore, $x a_\sigma = \sigma(x) a_\sigma$ for all $\sigma \in G(L/F)$ and all $x \in L$. Since $F = L^{G(L/F)}$, we must have $a_\sigma = 0$ for all $\sigma \neq 1$ in L. Hence $z = a_1 u_1 \in L$, so $L = Z_A(L)$ and we must have $L \subset A$ is a maximal subfield. Moreover, $[A : L] = |G(L/F)| = [L : F]$, so $[A : L] = [L : F]^2$. In particular, if we show A is a central simple F-algebra, then L is a self centralizing maximal subfield.

We first show that A is central. Let $z \in Z(A) \subset Z_A(L) = L$. Then $z u_\sigma = u_\sigma z = \sigma(z) u_\sigma$ for all $\sigma \in G(L/F)$. As $u_\sigma \in A^\times$ for all $\sigma \in G(L/F)$, we have $z = \sigma(z)$ for all $\sigma \in G(L/F)$. Therefore, $z \in F$, i.e., $F = Z(A)$, and A is F-central. Finally, we show that A is simple. If $\mathfrak{A}$ is a nonzero ideal in A, choose $0 \neq s = \sum_{i=1}^r a_{\sigma_i} u_{\sigma_i}$ in $\mathfrak{A}$ with r minimal. Suppose that $r > 1$. Choose $b \in L$ to satisfy $\sigma_1(b) \neq \sigma_2(b)$. Then we have

$$0 \neq s - \sigma_1(b)^{-1} s b = \sum_{i=2}^r \big(a_{\sigma_i} - \sigma_1(b)^{-1} a_{\sigma_i} \sigma_i(b)\big) u_{\sigma_i}$$

lies in $\mathfrak{A}$, contradicting the choice of r. Consequently, $r = 1$ and $0 \neq s = a_{\sigma_1} u_{\sigma_1}$ lies in $A^\times$. Therefore, $\mathfrak{A} = A$. In particular, A is simple. □

Corollary 107.14. *L/F be a finite Galois extension and f an element in $Z^2\big(G(L/F), L^\times\big)$. Let $A = (L/F, f)$. Then A is a finite dimensional central simple F-algebra and L is a centralizing maximal subfield of A.*

Proof. A is isomorphic to a crossed product algebra defined by a normalized factor set. □

Proposition 107.15. *Let L/F be a finite Galois extension and $f, g \in Z^2\big(G(L/F), L^\times\big)$. Then $[f] = [g]$ in $H^2\big(G(L/F), L^\times\big)$ if and only if $(L/F, f) \cong (L/F, g)$ as F-algebras.*

Proof. $(\Rightarrow)$ has been done.

$(\Leftarrow)$: We may assume that both f and g are normalized factor sets with $\{u_\sigma \mid \sigma \in G(L/F)\}$ a canonical basis for $A := (L/F, f)$ and $\{v_\sigma \mid \sigma \in G(L/F)\}$ a canonical basis for $B := (L/F, g)$. In particular, $w_1 = 1_A$ and $v_1 = 1_B$. Let $\varphi : A \to B$ be an F-algebra isomorphism. We must have $\varphi(u_1) = v_1$, so $\varphi(Lu_1) = L'v_i$ with $L' \cong L$ as F-algebras. Let $u'_\sigma = \varphi(u_\sigma)$ for all $\sigma \in G(L/K)$, so $\varphi(Lu_\sigma) = L'u'_\sigma$. Therefore, we have $B = \coprod_{G(L/F)} L'u'_\sigma$ and equations

$$\begin{aligned} u'_\sigma \varphi(x) &= \varphi({}^\sigma x) u'_\sigma \\ u'_\sigma u'_\tau &= \varphi\big(f(\sigma, \tau)\big) u_{\sigma\tau} \end{aligned} \tag{107.16}$$

for all $\sigma, \tau \in G(L/F)$. Since $Lv_1 \cong L'v_1$ are both simple subalgebras of B, by the Skolem-Noether Theorem 106.18, there exists an inner automorphism θ of B satisfying $\theta\big(\varphi(x)\big) = x$ for all $x \in L$. Define $w_\sigma = \theta(u'_\sigma)$ for all $\sigma \in G(L/F)$. Then applying the map θ to the equations (107.16) yields

$$\begin{aligned} B &= \coprod_{G(L/F)} Lw_\sigma \\ w_\sigma x &= \sigma(x) w_\sigma \\ w_\sigma w_\tau &= f(\sigma,\tau) w_{\sigma\tau} \end{aligned}$$

for all $\sigma, \tau \in G(L/F)$ and $x \in L$. As $v_\sigma v_{\sigma^{-1}} = g(\sigma, \sigma^{-1}) v_1 = g(\sigma, \sigma^{-1})$, we have $v_\sigma^{-1} = g(\sigma, \sigma^{-1})^{-1} v_{\sigma^{-1}}$ for all $\sigma \in G(L/F)$. Hence for all $x \in L$, we see that

$$\begin{aligned} w_\sigma v_\sigma^{-1} x = w_\sigma g(\sigma,\sigma^{-1})^{-1} x &= w_\sigma g(\sigma,\sigma^{-1})^{-1} \sigma^{-1}(x) v_{\sigma^{-1}} \\ &= x w_\sigma g(\sigma,\sigma^{-1})^{-1} v_{\sigma^{-1}} = x w_\sigma v_{\sigma^{-1}}. \end{aligned}$$

It follows that $w_\sigma v_\sigma^{-1}$ lies in $Z_B(L) = L$ and $w_\sigma {v_\sigma}^{-1} \neq 0$. Consequently, there exists a map $c : G(L/F) \to L^\times$ satisfying $w_\sigma = c(\sigma) v_\sigma$ for all $\sigma \in G(L/F)$ and

$$\begin{aligned} f(\sigma,\tau) w_{\sigma\tau} = w_\sigma w_\tau &= c(\sigma) v_\sigma c(\tau) v_\tau = c(\sigma)\, {}^\sigma c(\tau) v_\sigma v_\tau \\ &= c(\sigma) {}^\sigma c(\tau) g(\sigma,\tau) v_{\sigma\tau} = \big((\delta c) g)\big)(\sigma,\tau) w_{\sigma\tau} \end{aligned}$$

for all $\sigma, \tau \in G(L/F)$. Therefore, $f = (\delta c) g$ as needed. □

Corollary 107.17. *Let L/F be a finite Galois extension of degree n and $f \in Z^2\big(G(L/F)/L^\times\big)$. Then $(L/F, f) \cong \mathbb{M}_n(F)$ if and only if $[f] = 1$ in $H^2\big(G(L/F), L^\times\big)$.*

Proof. By Proposition 107.15, we need only show that $(L/F, 1) \cong \mathbb{M}_n(F)$. Let $(L/F, 1)$ have canonical basis $\{u_\sigma \mid \sigma \in G(L/F)\}$ and $\lambda : L \to \mathrm{End}_F(L)$ be defined by $x \mapsto \lambda_x : a \mapsto xa$. The map $(L/F, 1) \to \mathrm{End}_F(L)$ defined by $xu_\sigma \mapsto \lambda_x \sigma$ is an F-algebra homomorphism. Both are finite dimensional central simple F algebras of F-dimension n^2, so the map is an isomorphism. □

Remark 107.18. If A is a finite dimensional central simple F-algebra and M is a finitely generated A-module, the the A-isomorphism type of M is completely determined by $[M : F]$. Indeed if N is an irreducible A-module, then there exists a unique integer n such that $M \cong N^n$ as A-modules and $[M : F] = n[N : F]$.

Proposition 107.19. *Let L/F be a finite Galois extension and $f, g \in Z^2\big(G(L/F), L^\times\big)$. Then $(L/F, f) \otimes_F (L/f, g) \sim (L/F, fg)$.*

Proof. (Chase) Let $h = fg$. Set
$A = (L/K, f)$ with canonical basis $\{u_\sigma \mid \sigma \in G(L/F)\}$,
$B = (L/K, g)$ with canonical basis $\{v_\sigma \mid \sigma \in G(L/F)\}$, and
$C = (L/K, h)$ with canonical basis $\{w_\sigma \mid \sigma \in G(L/F)\}$.

Let $M = A \widetilde{\otimes}_L B$ be the tensor product of A and B as L-vector spaces, i.e., $\alpha a \widetilde{\otimes}_L b = a \widetilde{\otimes}_L \alpha b$ for all $\alpha \in L$. We give M the structure of a right $A \otimes_F B$- module by

$$(a \widetilde{\otimes}_L b)(a' \otimes b') := aa' \widetilde{\otimes}_L bb' \text{ for all } a, a' \in A,\ b, b' \in B.$$

Claim. M is a (left) C-module by

$$(cw_\sigma)(a\widetilde{\otimes}_L b) = (cu_\sigma a\widetilde{\otimes}_L v_\sigma b)$$

for all $a \in A$, $b \in B$, $c \in C$, and $\sigma \in G(L/F)$:

The only thing that is not straight-forward is associativity, i.e., $(\lambda\lambda')m = \lambda(\lambda' m)$ for all $\lambda, \lambda' \in C$, $m \in M$ which we show, leaving the rest as an exercise. To show this, it suffice to assume that $\lambda = cw_\sigma$, $\lambda' = c'w_\tau$, and $m = a\widetilde{\otimes}_L b$, for $c, c' \in L$, $\sigma, \tau \in G(L/F)$, $a \in A$, and $b \in B$. But

$$\begin{aligned}
(\lambda\lambda')m &= c\,{}^\sigma c' h(\sigma,\tau) w_{\sigma\tau}(a\widetilde{\otimes}_L b) \\
&= c\,{}^\sigma c' f(\sigma,\tau) u_{\sigma\tau} a\widetilde{\otimes}_L g(\sigma,\tau) v_{\sigma\tau} b = c\,{}^\sigma c' u_\sigma u_\tau a\widetilde{\otimes}_L v_\sigma v_\tau b \\
&= \lambda(c' u_\tau a\widetilde{\otimes}_L v_{\sigma\tau} b) = \lambda(\lambda' m)
\end{aligned}$$

as needed, and the Claim is established.

It follows that M is a $C-(A \otimes_F B)$-bimodule. This means that the map $\rho : A \otimes_F B \to \mathrm{End}_C(M)$ by $a \otimes b \mapsto \rho_{a\otimes b}$, i.e,, right multiplication by $a \otimes b$, for all $a \in A$, $b \in B$ is an F-algebra homomorphism.

Let $n = [L : F] = [A : L] = [B : L][C : L]$. Then

$$[M : F] = [M : L][L : F] = n^3 = n[C : F].$$

Since C is a finite dimensional central simple F-algebra, $M \cong C^n$ as C-modules by Remark 107.18, hence $\mathrm{End}_C(M) \cong \mathbb{M}_n(C)$. It follows that $[\mathbb{M}_n(C) : F] = n^2[C : F] = n^4 = [A \otimes_F B : F]$. Therefore, the F-algebra homomorphism ρ must be an isomorphism. $\square$

Theorem 107.20. *Let L/F be a finite Galois extension. Then*

$$\Phi_{L/F} : H^2\big(G(L/F), L^\times\big) \to Br(L/F) \ \textit{ defined by } \ [f] \mapsto (L/F, f)$$

is a group isomorphism.

Proof. By Proposition 107.19, Φ is a group homomorphism, hence by Corollary 107.17, it injective. So we need only show that it is onto. Let A be a finite dimensional central simple F-algebra split by L. Replacing A by a (Brauer) equivalent algebra, we may assume that $L \subset A$ is a simple subalgebra. By the Skolem-Noether Theorem 106.18, for each $\sigma \in G(L/F)$, there exists an element $u_\sigma \in A^\times$ satisfying $\sigma(x) = u_\sigma x u_\sigma^{-1}$ for all $x \in L$., i.e., we have

$$u_\sigma = {}^\sigma x u_\sigma \ \text{ for all } \ \sigma \in G, \ \text{ all } \ x \in L.$$

Claim. $A = \coprod_{G(L/F)} Lu_\sigma$ as an L-vector space:

If we show $\coprod_{G(L/F)} Lu_\sigma = \sum_{G(L/F)} Lu_\sigma$, then this is true, since we know that $[A : F] = [L : F]^2$ and $\coprod_{G(L/F)} Lu_\sigma \subset A$. Suppose that $\mathcal{B} = \{u_\sigma \mid \sigma \in G(L/F)\}$ is L-linearly dependent. Let $w = \sum_{i=1}^r a_{\sigma_i} u_{\sigma_i} = 0$ with $u_{\sigma_i} \in \mathcal{B}$, $a_{\sigma_i} \in L$ be chosen with r minimal. Each $u_\sigma \in A^\times$ for $\sigma \in G(L/F)$, $r > 1$. Choose $b \in L$ satisfying $\sigma_1(b) \neq \sigma_2(b)$. Then

$$0 = \sigma_1(b)w - wb = \sum_{i=2}^r \big(\sigma_1(b) - \sigma_i(b)\big) a_{\sigma_i} u_{\sigma_i},$$

contracts the minimality of r. This establishes the Claim.

Equation (*) implies that $u_\sigma u_\tau u_{\sigma\tau}^{-1}$ lies in $Z_A(L) = L$. Hence there exist $f(\sigma,\tau)$ in $L^\times$ satisfying $u_\sigma u_\tau = f(\sigma,\tau)u_{\sigma\tau}$ for all $\sigma,\tau \in G(L/K)$. The associativity for A now implies that $f \in Z^2\big(G(L/F), L^\times\big)$. □

Remark 107.21. Amitsur has shown that there exist finite dimensional central division F-algebras that are not isomorphic to cross product algebras.

Proposition 107.22. *Let A be a finite dimensional central simple F-algebra of index n. Then $[A^n] = 1$ in $\mathrm{Br}(F)$.*

Proof. By Theorem 107.20 and Corollary 107.14, we may assume that $A = (L/F, f)$ with L/F a finite Galois extension say of degree m and L a self centralizing maximal subfield of A with canonical basis $\{u_\sigma \mid \sigma \in G(L/F)\}$ for A. In particular, we have $u_\sigma x = \sigma(x)u_\sigma$ for all $x \in L$. We have $A \cong \mathbb{M}_m(D)$ for some central division F-algebra and $\operatorname{ind} A \mid \deg A$. We also have $\deg A = [L : F] = mn$. Let M be a simple A-module, so M is a D-vector of dimension m (as isomorphic to a column of $\mathbb{M}_n(D)$). Then

$$mn \dim_L M = [L : F] \dim_L M = \dim_F M = \dim_D M \dim_F D = mn^2,$$

so $\dim_L M = n$. Let $\{v_1,\dots,v_n\}$ be an L-basis for M. For each $x \in A$, define $C(x) \in \mathbb{M}_m(D)$ by

$$xv_j = \sum_{i=1}^m C(x)_{ij} v_i.$$

We have

$$u_\sigma u_\tau v_j = f(\sigma,\tau)u_{\sigma\tau}v_j = f(\sigma,\tau)\sum_{i=1}^m C(u_{\sigma\tau})_{ij}v_i \tag{1}$$

and

$$\begin{aligned} u_\sigma u_\tau v_j &= u_\sigma\big(\sum_{k=1}^m C(u_\tau)_{kj}v_k\big) = \big(\sum_{k=1}^m \sigma\big(C(u_\tau)_{kj}\big)u_\sigma v_k \\ &= \sum_{i,k=1}^m \sigma\big(C(u_\tau)_{kj}\big)\big(C(u_\sigma)_{ik}v_i\big). \end{aligned}$$

Let $\sigma(C) = \big(\sigma(C_{ij})\big)$ in $\mathbb{M}_m(F)$, i.e., the matrix in which σ acts on every entry of C. So we have

$$u_\sigma u_\tau v_j = \sum_{i=1}^m \sigma\big(C(u_\tau)_{ij}\big)v_i. \tag{2}$$

Comparing (1) and (2) yields

$$f(\sigma,\tau)C(u_{\sigma\tau}) = C(u_\sigma)\sigma\big(C(u_\tau)\big)$$

in $\mathbb{M}_m(L)$. Setting $c_\sigma = \det\big(C(u_\sigma)\big)$ for all $\sigma \in G(L/F)$, we deduce that

$$f(\sigma,\tau)^m c_{\sigma\tau} = c_\sigma\sigma(c_\tau)$$

for all $\sigma,\tau \in G(L/F)$. It follows that $f(\sigma,\tau)^m = c_\sigma\sigma(c_\tau)c_{\sigma\tau}^{-1}$ is the trivial factor set. □

Corollary 107.23. *The Brauer group of a field is a torsion group.*

Proof. $\mathrm{Br}(F) = \bigcup_{\substack{L/F \\ \text{finite Galois}}} \mathrm{Br}(L/F)$ □

Definition 107.24. Let A be a a finite dimensional central simple F-algebra. The order of $[A]$ in the Brauer group of F is called the *exponent* or *period* of A. and is denoted $\exp A$.

Corollary 107.25. *Let A be a a finite dimensional central simple F-algebra. Then* $\exp A$*, the exponent of A, divides* $\operatorname{ind} A$*, the index of A.*

Remark 107.26. If A is a a finite dimensional central simple F-algebra, then the $\exp A$ and $\operatorname{ind} A$ have the same prime divisors. We leave this as an exercise.

Remark 107.27. We indicate (with few details) how one can put together all the $H^2(G(L/F), L^\times)$ with L/F a finite Galois extension to get a 2nd cohomology group isomorphic to the Brauer group $\mathrm{Br}(F)$.
Let $\widetilde{F}$ be an algebraic closure of F and $\Gamma = G(F_{sep}/F)$, the (absolute) Galois group of F, where F_{sep} is the separable closure of F in $\widetilde{F}$. We saw that Γ_F has a topology, the profinite topology. Let $F_{sep}^\times$ have the discrete topology, and define

$$\begin{aligned} Z^2_{cont}(\Gamma_F, F_{sep}^\times) &:= \{f \in Z^2(\Gamma_F, F_{sep}^\times) \mid f \text{ continuous}\}, \\ B^2_{cont}(\Gamma_F, F_{sep}^\times) &:= \{f \in B^2(\Gamma_F, F_{sep}^\times) \mid f \text{ continuous}\}, \\ H^2_{cont}(\Gamma_F, F_{sep}^\times) &:= Z^2_{cont}(\Gamma_F, F_{sep}^\times)/B^2_{cont}(\Gamma_F, F_{sep}^\times). \end{aligned}$$

Let $E/L/F$ be finite extensions in Γ_F with L/F and E/F Galois extensions. One shows that the group inclusion $i_{E/F} : \mathrm{Br}(L/F) \to \mathrm{Br}(E/F)$ and the group epimorphism $G(E/F) \to G(L/F)$ given by $\sigma \mapsto \sigma|_L$ (so $G(E/F)/G(E/L) \cong G(L/F)$) of Galois theory induce a group monomorphism

$$\inf_{E/L} : H^2\big(G(L/F), L^\times\big) \to H^2\big(G(E/F), E^\times\big)$$

such that

$$\begin{tikzcd} \mathrm{Br}(L/F) \arrow[r, "i_{E/L}"] \arrow[d, "\Phi_{L/F}"'] & \mathrm{Br}(E/F) \arrow[d, "\Phi_{E/F}"] \\ H^2\big(G(L/F), L^\times\big) \arrow[r, "\inf_{E/L}"] & H^2\big(G(E/F), E^\times\big). \end{tikzcd}$$

commutes. One then can show that $H^2_{cont}(\Gamma_F, F_{sep}^\times)$ is built from the $H^2\big(G(L/F), L^\times\big)$ as L runs over the finite Galois extensions L/F. More specifically, $H^2_{cont}(\Gamma_F, F_{sep}^\times) = \varinjlim H^2\big(G(L/F), L^\times\big)$, the direct limit over $\inf_{L/F}$ where the L/F run over finite Galois extensions in $\widetilde{F}$. (Cf. Exercise 107.30(7) for the definition of direct limit.) We then obtain a group isomorphism $\Phi_F : Br(F) \to H^2_{cont}(\Gamma_F, F_{sep}^\times)$ compatible with all $\Phi_{L/F}$.

We end this section with remarks amplifying comments made about cyclic algebras in Section 105. We omit the proofs, since we shall not develop cohomology theory.

Remark 107.28. Let L/F be a cyclic field extension of degree n with $G(L/F) = \langle\sigma\rangle$ and $A = (L/F, \sigma, a)$ a cyclic F-algebra. Then the following can be shown:

1. Suppose that $L/K/F$ is an intermediate field with $[K : F] = m$. Then $(L/F, \sigma|_K, a) \cong (L/K, \sigma, a^{\frac{m}{n}})$. In particular, A is a division F-algebra if $\overline{a}$ has order n in $F^\times/N_{L/F}(L^\times)$. The last statement follows as we have $n = \exp A = \deg A$ in this case (cf. the exercises for this section).
2. There is a group isomorphism $F^\times/N_{L/F}(L^\times) \to H^2\big(G(L/F), L^\times\big)$ mapping $xN_{L/F}(L^\times) \mapsto (L/F, \sigma, x)$. [This arises from the periodicity of Tate cohomology for finite cyclic groups.]
3. A is a central division F-algebra if $[A]$ has order n in $\operatorname{Br}(L/F)$.

Remark 107.29. Historically, much of our study of finite dimensional centeral simple algebras reached a high point in algebraic number theory, where the classification of finite dimensional central division F-algebras with $F/\mathbb{Q}$ is a finite extension was determined in a joint paper of Brauer, Hasse, and Noether (some of the results were independently proven by Albert). Brauer had defined cross product algebras (and the Brauer group) which was simplified by Noether, generalizing cyclic algebras over arbitrary fields. Recall if $F/\mathbb{Q}$ is a finite extension (an algebraic number field), then

$$\mathbb{Z}_F := \{x \in K \mid f(x) = 0, \text{ for some monic polynomial } f \in \mathbb{Z}[t]\}$$

is the ring of algebraic integers in F. If $\mathfrak{p}$ is a prime ideal in $\mathbb{Z}_F$, let $(\mathbb{Z}_F)_\mathfrak{p}$ denote its completion (cf. Exercise 98.32(4)) at $\mathfrak{p}$ and $F_\mathfrak{p}$ the quotient field of $(\mathbb{Z}_F)_\mathfrak{p}$. All central simple $F_\mathfrak{p}$-algebras were known to be cyclic, in fact, one central division ring for each integer n (up to isomorphism). Let K/F be a finite extension. Then $\mathfrak{p}\mathbb{Z}_K$ factors into a product of primes in $\mathbb{Z}_K$. If $\mathfrak{P}$ is one such, we write $\mathfrak{p} \mid \mathfrak{P}$ and we get a field extension $K_\mathfrak{P}/F_\mathfrak{p}$. Let $X(F) = \operatorname{Max}\mathbb{Z}_K \cup X_\infty(\mathbb{Z}_K)$ where $X_\infty(\mathbb{Z}_K) := \{f : \mathbb{Z}_F \to \mathbb{C} \mid f \text{ a field homomorphism}\}$. So each $f \in X_\infty(Z_K)$ has image in $\mathbb{R}$, with completion $\mathbb{R}$, called a real infinite prime or not, called a complex infinite prime, with completion $\mathbb{C}$. We then extend our notation to all elements in $X(F)$. Let L/F be a cyclic extension. Hasse proved that $x \in F$ satisfies $x \in N_{L/F}(L)$ if and only if $x \in N_{L_\mathfrak{P}/F_\mathfrak{p}}(L_\mathfrak{P})$ for all $\mathfrak{p} \in X(F)$ and all $\mathfrak{P} \in X(L)$ with $\mathfrak{P} \mid \mathfrak{p}$. One says that x is a norm if and only if x is a norm locally for all primes (finite and infinite). As cyclic F-algebras split under a norm condition, this *local-global principle* applies to cyclic F-algebras. They then showed every crossed product F-algebra was a cyclic F-algebra and every central simple F-algebra was a cyclic algebra.

Exercises 107.30.

1. Let A and B be finite dimensional central simple F-algebras. Show the following:
   (a) Let K/F is a finite field extension with $\operatorname{ind} A$ relatively prime to $[K : F]$. Then $\operatorname{ind} A^K \mid \operatorname{ind} A$. In particular, if in addition, A is a division algebra, so is A^K.
   (b) We have $\operatorname{ind}(A \otimes_F B)$ divides $(\operatorname{ind} A)(\operatorname{ind} B)$.
   (c) We have $\operatorname{ind} \underbrace{(A \otimes_F \cdots \otimes_F A)}_{m}$ divides $\operatorname{ind} A$ for all $m \geq 1$.
2. Let D be a central division F-algebra and K/F a finite field extension with $[K : F]$ a prime dividing $\deg D$. Prove that the following are equivalent:

   (a) K is isomorphic to a subfield of D.
   (b) D^K is not a division algebra.
   (c) $\deg D = [K : F] \operatorname{ind} D^K$.

3. Let A be a finite dimensional central simple F-algebra with $\operatorname{ind} A = p^e m$, p a prime not dividing m and $e \geq 1$. Prove that there exists a finite separable extension K/F satisfying $p \nmid [K : F]$ and $\operatorname{ind} A^K = p^e$.
4. Show Example 107.10 is a cross product algebra.
5. Let A and B be finite dimensional central simple F-algebras and K/F a finite field extension. Show the following:
   (a) Every prime divisor of $\operatorname{ind} A$ divides $\exp A$.
   (b) We have $\exp A^K \mid \exp A$.
   (c) If $\operatorname{ind} A$ and $[K : F]$ are relatively prime, then $\exp A^K = \exp A$.
   (d) We have $\exp(A \otimes_F B)$ divides the least common multiple of $\exp A$ and $\exp B$.
   (e) We have $\exp(\underbrace{A \otimes_F \cdots \otimes_F A}_{m}) = (\exp A)/d$ where d is the gcd of $\exp A$ and $[K : F]$.
   (f) If $\operatorname{ind} A$ and $\operatorname{ind} B$ are relatively prime, then $(\operatorname{ind}(A \otimes_F B) = (\operatorname{ind} A)(\operatorname{ind} B)$ and $(\exp(A \otimes_F B) = (\exp A)(\exp B)$. In particular, if in addition, A and B are division algebras, so is $A \otimes_F B$.
6. Let D be a central division F-algebra with $\deg D = p_1^{e_1} \cdots p_r^{e_r}$ its standard factorization. Then there exists an F-algebra isomorphism $D = D_1 \otimes_F \cdots \otimes_F D_r$ with D_i a central F-division algebra with $\deg D_i = p^{e_i}$ for all $i = 1, \ldots, r$, unique up to isomorphism.
7. Let I be a partially ordered set under $\leq$ that also satisfies for all $i, j \in I$, there exists a $k \in I$ such that $i \leq k$ and $j \leq k$; and $\{M_i\}_I$ be a collection of R-modules. Suppose for all $i \leq j$ in I, there exist R-homomorphisms $\theta_{i,j} : M_i \to M_j$ with $\theta_{i,i} = 1_{M_i}$ and

$$\begin{tikzcd}
M_i \arrow[dd, "\theta_{ij}"'] \arrow[dr, "\theta_{ik}"] & \\
& M_k \\
M_j \arrow[ur, "\theta_{jk}"'] &
\end{tikzcd}$$

commutes whenever $i \leq j \leq k$. Show that there exists an R-module M and for each $i \in I$ an R-homomorphism $\psi_i : M_i \to M$ satisfying the following universal property: For all non-negative $i \leq j$ in I,

$$\begin{tikzcd}
M_i \arrow[dd, "\theta_{ij}"'] \arrow[dr, "\psi_i"] & \\
& M \\
M_j \arrow[ur, "\psi_j"'] &
\end{tikzcd}$$

commutes, and if there exist R-homomorphisms $\theta_i : M_i \to M$ satisfying for all $i \le j$ in I,

$$\begin{tikzcd} M_i \arrow[d, "\theta_{ij}"'] \arrow[dr, "\varphi_i"] & \\ & M' \\ M_j \arrow[ur, "\varphi_j"'] & \end{tikzcd}$$

commutes, then there exists a unique R-homomorphism $\mu : M_i \to M'$ satisfying

$$\begin{tikzcd} M_i \arrow[dd, "\theta_{ij}"'] \arrow[dr, "\psi_i"] \arrow[drr, "\varphi_i"] & & \\ & M \arrow[r, "\mu"] & M' \\ M_j \arrow[ur, "\psi_j"] \arrow[urr, "\varphi_j"'] & & \end{tikzcd}$$

commutes for all $i \le j$ in I. Such a M is unique up to a unique isomorphism and is called the *direct limit* of the M_i and denoted by $\lim\limits_{\rightarrow} M_i$.

8. Show reversing all the arrows in Exercise 7 still gives a valid result and the unique R-module M up to isomorphism is called the *inverse* or *projective limit* of the M_i and denoted by $\lim\limits_{\leftarrow} M_i$.

[Remark: For example, If Γ_F is absolute Galois group of a field F (in some fixed algebraic closure), then $\Gamma_F = \lim\limits_{\leftarrow} G(L/F)$, where the inverse limit is over all finite Galois extensions L/F. We know that $G(L/F) = \Gamma_F/\Gamma_L$ for all Galois extension L/F, and we have

$$H^2_{cont}(\Gamma_F, F^\times_{sep}) = \lim_{\rightarrow} H^2(G(L/F), L^\times) = H^2(\lim_{\leftarrow} G(L/F), L^\times)$$

where L/F runs over all finite Galois extensions of F.]

108. Polynomial Rings over a Division Algebra

When we investigated polynomial rings over a commutative ring, we noted that the analogous theory broke down if the ring was not commutative. As an example, we showed that the polynomial $t^2 - 1$ over the division ring of Hamiltonion quaternions had infinitely many roots. (Cf. Remark 34.10(2).) A major reason for this is that if R is a ring that is not commutative and x an element in R, then the evaluation map $e_x : R[t] \to R$ given by $\sum_{i=0}^n a_i t^i \mapsto \sum_{i=0}^n a_i x^i$ is not necessarily a ring homomorphism, although it is an additive homomorphism. For example, if with $ab \ne ba$ and $f = (t-a)(t-b)$ in $R[t]$, then

$$f = t^2 - (a+b)t + ab, \quad \text{hence} \quad f(a) = -ba + ab \ne 0,$$

but $e_a(t-a)e_a(t-b) = 0$. The problem arises as t is central in $R[t]$, i.e., $at = ta$ for all $a \in R$. One useful result, still holding in $R[t]$, is the division algorithm as the usual proof works, i.e., if $f \in R[t]$ is monic (or leading term a unit), for any $g \in R[t]$, there exists $h, r \in R[t]$ satisfying

$$g = hf + r \ \text{ with } \ r = 0 \ \text{ or } \deg r < \deg f.$$

As mentioned in the special case in Appendix D, Lemma D.5, there are in fact two such algorithms, i.e., left and right, so there exist h', r' in $R[t]$ with $g = fh' + r'$ with $r' = 0$ or $\deg r' < \deg f$. We shall concentrate on the first, which we call *right division* by f. As in Appendix D, the special case when f is linear, say $f = t - x$, is of interest, i.e.,

$$g = h \cdot (t - x) + r \ \text{ with } \ r \in R.$$

We call x a *right root* of $g \in R[t]$ if $g(x) = 0$ in the above, i.e., if $g = \sum_{i=0}^n a_i t^i$, then $\sum_{i=0}^n a_i x^i = 0$. (So a left root of g would be $\sum_{i=0}^n x^i a_i = 0$.) Since we are interested in the map $e_a : R[t] \to R$ by plugging a into t on the right in this section, we shall let $g(a) := e_a(g)$ to be this evaluation on the right. In particular, if x is a right root of g, then $g \in R[t](t - x)$. Our analogy to the commutative case is

Lemma 108.1. *Let R be a ring, $0 \neq g \in R[t]$. Then x in R is a right root of f if and only if $t - x$ is a right divisor of g. In particular,*

$$\{h \in R[t] \mid x \ \ \textit{is a right root of } h\} = R[t](t - x).$$

PROOF. Let

$$f = (\sum_{i=0}^n a_i t^i)(t - x) = \sum_{i=0}^n a_i t^{i+1} - \sum_{i=0}^n a_i x t^i.$$

Then x is a right root of f, as $\sum_{i=0}^n a_i x^{i+1} - \sum_{i=0}^n a_i x x^i = 0$. □

A very useful observation, is the following observation:

Proposition 108.2. *Let D be a division ring. Suppose we have an equation $f = gh$ of polynomials in $D[t]$ with x in D satisfying $r = h(x)$ is nonzero (so a unit in D). Then*

$$f(x) = g(rxr^{-1})h(x).$$

In particular, if x is a right root of f but not of h, then rxr^{-1} is a right root of g.

PROOF. Let $g = \sum_{i=0}^n a_i t^i$, hence $f = \sum_{i=0}^n a_i h(t) t^i$. Therefore,

$$\begin{aligned} f(x) &= \sum_{i=0}^n a_i h(x) x^i = \sum_{i=0}^n a_i r x^i \\ &= \sum_{i=0}^n a_i (rxr^{-1})^i r = g(rxr^{-1})h(x). \end{aligned}$$ □

This results in an analogue of the number of roots for a polynomial over a field, an analogue that often comes up in the study. If x is a nonzero element in a division ring D, we shall let $\mathcal{C}(x) := \{rxr^{-1} \mid r \in D^\times\}$ be the conjugacy class of x in $D^\times$ and call $\{0\}$ the conjugacy class of 0 in D. We shall call any of these conjugacy classes in D.

Theorem 108.3. *Let D be a division ring and f a polynomial in $D[t]$ of degree d. Then the right roots of f lie in at most n conjugacy classes in D (not necessarily distinct). In particular, if $f = (t - a_1) \cdots (t - a_n)$ with $a_1, \ldots, a_n \in D$ and d is a right root of f, then d is conjugate to a_i for some i, $i = 1, \ldots, n$.*

Proof. The result is immediate if $n = 1$, so assume that $n \geq 2$. Let $x \in D$ be a right root of f. Then $f = g \cdot (t - x)$ with $g \in D[t]$ of degree less than n. If y is another root of f different from x, then by Propositon 108.2, a conjugate of y is a root of g, and the first statement follows by induction. If $g = (t - b_1) \cdots (t - b_{n-1})$, the second statement also follows by Proposition 108.2 and induction. $\square$

Just as in field theory, we are interested in the analogue of algebraic elements. If D is a division ring, its center F is a field, and we can look at elements x in D algebraic over F. Of course, $F(x)$ is also a subfield of D, and if x is algebraic it has an irreducible polynomial. But if x is algebraic over F, then every conjugate of x is also algebraic over F, since x a right root of $g \in D[t]$ implies that rxr^{-1} is a a right root of $rg(t)r^{-1}$ for all nonzero r in D.

Definition 108.4. Let D be a division ring with center F. Let x in D and $\mathcal{C}$ be the conjugacy class of x in D. We say that $\mathcal{C}$ is algebraic over F if x is algebraic over F (if and only if every element of $\mathcal{C}$ is algebraic over F). Suppose that x in D is algebraic over F with $m_F(x) \in F[t]$ its minimal polynomial. Then $m_F(x) = m_F(rxr^{-1})$ for every nonzero element r in D, We call this polynomial the *minimal polynomial* of $\mathcal{C}$ and denote it by $m_F(\mathcal{C})$.

The analogue of the field case now becomes:

Lemma 108.5. *Let D be a division ring with center F and $\mathcal{C}$ a conjugacy class in D algebraic over F. Suppose that h is a nonzero polynomial in $D[t]$ satisfying $h(c) = 0$ for all c in $\mathcal{C}$. Then $\deg h \geq \deg m_F(\mathcal{C})$.*

Proof. Suppose the result is false. Among all counterexamples choose one h in $D[t]$ of minimal degree m with $m < \deg m_F(\mathcal{C})$. Let $h = \sum_{i=0}^{m} a_i t^i$. We may assume that $a_m = 1$, i.e., that h is monic. By the commutative case, we may also assume that not all the a_i lie in F, say $a_j \notin F$. Choose b in D satisfying $a_j b \neq b a_j$. Let $c \in \mathcal{C}$, so by assumption $\sum_{i=0}^{m} a_i c^i = 0$. Hence

$$0 = b\Big(\sum_{i=0}^{m} a_i c^i\Big)b^{-1} = \sum_{i=0}^{m} (b a_i b^{-1})(b c^i b^{-1}) \text{ with } a_j \neq b a_j b^{-1}.$$

Then the polynomial $h_1 = \sum_{i=0}^{m-1} (a_i - b a_i b^{-1}) t^i$ vanishes on $b\,\mathcal{C}\,b^{-1} = \mathcal{C}$, contradicting the minimality of m. $\square$

This lemma implies:

Proposition 108.6. *Let D be a division ring with center F and $\mathcal{C}$ a conjugacy class in D algebraic over F. If $h \in D[t]$, then h vanishes on $\mathcal{C}$ if and only if $h \in D[t] m_F(\mathcal{C})$.*

Proof. Set $f = m_F(\mathcal{C})$. By Lemma 108.1, for all $x \in \mathcal{D}$, we have $f \in D[t] \cdot (t - x)$. It follows that if $h \in D[t]f$, then $h(x) = 0$ for all x in $\mathcal{C}$. Conversely, suppose that

$h(x) = 0$ for all $x \in \mathcal{C}$. By the (right) division algorithm, $h = qf + g$ in $D[t]$ with $g = 0$ or $\deg g < \deg f$. Since $h(x) = 0 = f(x)$ for all $x \in \mathcal{C}$, we also have $g(x) = 0$ for all $x \in \mathcal{C}$. By Lemma 108.5, we must have $g = 0$, so $h \in D[t]f$. □

Corollary 108.7. *Let D be an infinite division ring. Then no nonzero polynomial in $D[t]$ vanishes on D.*

Proof. We may assume that D is not a field. Suppose the corollary is false and that $h = t^m + a_{m-1}t^{m-1} + \cdots + a_1 t + a_0$ in $D[t]$ vanishes on all of D. In particular, $h(0) = 0$, so $a_0 = 0$. We may also assume that we have chosen h with m minimal. Let F be the center of D. By the proof of Lemma 108.5, we see that we must have $a_i \in F$ for all i, i.e., $h \in F[t]$. It follows that F must be a finite field. Let $x \in D$. Then x is a root of h, so is algebraic over F. In particular, the subfield $F[x]$ of D is a finite extension of F, so also a finite field. Hence there exists an integer $n = n(x)$ satisfying $x^n = x$. By Proposition 108.5, D is a field, a contradiction. □

Lemma 108.8. *Let D be a division ring with center F and $f \in F[t]$. Suppose that there exist polynomials $g_1, g_2 \in D[t]$ satisfying $f = g_1 g_2$. Then $f = g_2 g_1$.*

Proof. As $f \in F[t] \subset Z(D[t])$ (check), $g_1 g = g g_1 = g_1 g_2 g_1$. Since $D[t]$ is a non-commutative domain (obvious definition), $g = g_2 g_1$. □

We now show that the minimal polynomial of an algebraic conjugacy class in a division ring D splits in $D[t]$.

Theorem 108.9. (Wedderburn) Let D be a division ring with center F and $\mathcal{C}$ a conjugacy class in D algebraic over F. Suppose that $\deg m_F(\mathcal{C}) = n$. Then there exist $a_1, \dots, a_n$ in D satisfying

$$(*) \qquad m_F(\mathcal{C}) = (t - a_1) \cdots (t - a_n).$$

Moreover, a_1 in $\mathcal{C}$ can be arbitrarily chosen and then any cyclic permutation of $(*)$ is still $m_F(\mathcal{C})$.

Proof. Let $f = m_F(\mathcal{C})$. Fix any element $a_1 \in \mathcal{C}$. Then $f \in D[t](t - a_1)$ by Lemma 108.1. Write $f = g \cdot (t - a_r) \cdots (t - a_1)$ in $D[t]$ with $a_1, \dots, a_r \in \mathcal{C}$ and r maximal. Let $h = (t - a_r) \cdots (t - a_1)$. We show h vanishes on $\mathcal{C}$. If not then there exist an $x \in \mathcal{C}$ with $h(x) \neq 0$. By Proposition 108.2, there exists an $a_{r+1} \in \mathcal{C}$, a conjugate of x satisfying $g(a_{r+1}) = 0$. It follows that $f = h \cdot (t - a_{r+1}) \cdots (t - a_1)$ for some $h \in D[t]$ by Lemma 108.1. This contradicts the maximality of r. Thus h vanishes on $\mathcal{C}$, hence lies in $D[t]f$ by Proposition 108.6. It follows that $r = n$ and $f = (t - a_n) \cdots (t - a_1)$ in $D[t]$ by Lemma 108.5. The last statement in the theorem follows by Lemma 108.8. □

Definition 108.10. A division ring D is called *right algebraically closed* (respectively, *left algebraically closed*) if every non-constant polynomial in $D[t]$ has a right (respectively, left) root.

We note that Frobenius's Theorem for division rings over the reals 104.1 holds over any real closed field using the same proof. In particular, the Hamiltonian quaternions $\mathcal{H}$ are defined over any real closed field F. We use the same notation as in Construction 33.1. In particular, we have the map $^- : \mathcal{H} \to \mathcal{H}$ given by $x_0 + x_1 i + x_2 j + x_3 k \mapsto x_0 - x_1 i - x_2 j - x_3 k$

is an antiautomorphism fixing F. This antiautomorphism induces an antiisomorphism $^{-}: \mathcal{H}[t] \to \mathcal{H}[t]$ given by $\sum_{i=0}^{n} a_i t^i \mapsto \sum_{i=0}^{n} \overline{a_i} t^i$. In particular, if $f \in \mathcal{H}[t]$, then $\overline{f} f$ lies in $F[t]$.

Proposition 108.11. *Let F be a real closed field. Then the Hamiltonian quaternions $\mathcal{H}$ over F is right and left algebraically closed.*

Proof. Let $f = \sum_{i=0}^{n} a_i t^i$ be a non-constant polynomial in $\mathcal{H}[t]$. We show that f has a right root. This is certainly true if $n = 1$, so we may assume that $n > 1$. As the element $i \in \mathcal{H}$ satisfies, $F(i)$ is algebraically closed and $\overline{f} f$ lies in $F[t]$, it has a (right) root α. If α is a right root of f, we are done, so we may assume not. Then a conjugate β of α is a right root of $\overline{f}$ by Proposition 108.2. As $\sum_{i=0}^{n} \overline{\beta}^i a_i = \sum_{i=0}^{n} \overline{a_i} \beta^i = 0$, we have $\overline{\beta}$ is a left root of f. By the left analogue of Proposition 108.2, $f = (t - \overline{\beta}) g$ for some $g \in \mathcal{D}[t]$. Since $\deg g = n - 1$, the polynomial g has a right root in $\mathcal{H}$, hence so does f. Consequently, $\mathcal{H}$ is right algebraically closed. The proof that $\mathcal{H}$ is left algebraically closed is analogous. □

Using the Artin-Schreier Theorem 77.2, we can now prove:

Theorem 108.12. (Baer) *Let D be a non-commutative division ring with center F. Suppose that D is finite over F, i.e., $\dim_F D < \infty$, and every nonzero polynomial in $F[t]$ has a right root in D, e.g., D is right algebraically closed. Then F is real closed and $D = \mathcal{H}$.*

Proof. Let $n = \dim_F D$. Let f be a non-constant polynomial in $F[t]$. Then f has a right root α in D. In particular, $\deg m_F(\alpha) \leq n$ by Lemma 108.5.

Claim. F is perfect.

We may assume that F has positive characteristic p. Let L be an algebraic closure of F and α be an element of L. Let α^{1/p^m} be the unique p^mth root of α in L. We must show $\alpha^{1/p}$ lies in F. Since $F \subset F(\alpha^{1/p}) \subset \cdots \subset F(\alpha^{1/p^m}) \subset \cdots$ and $[F(\alpha^{1/p^m}) : F] \leq n$, there exists an m such that $\alpha^{1/p^{m+1}}$ lies in $F(\alpha^{1/p^m})$, hence $\alpha^{1/p}$ lies in $F^{p^m}(\alpha) \subset F$, as needed.

Now let E be a intermediate field satisfying $L/E/F$ chosen with E a simple extension of F (i.e., is $E = F(x)$ some x) of maximal degree. Since E/F is separable, E is determined by a separable polynomial f in $F[t]$ and has a root in D. Therefore, $[E : F] \leq n$. In particular, L/F must be a finite extension. Since D is not commutative, $F < L$. By the Artin-Schreier Theorem 77.2, F is real closed and hence $D = \mathcal{H}$ by Frobenius's Theorem 104.1. □

Lemma 108.13. *Let D be a division ring, V a left vector space over F, and M a subset of V closed under addition and containing at least two linear independent elements in V. Suppose that $\varphi : M \to V$ is an additive map such that for all $x \in M$, $\varphi(x) = \lambda_x x$ for some $\lambda_x \in F$. Then there exists an element $\lambda \in F$ satisfying $\varphi(x) = \lambda x$ for all $x \in M$,*

Proof. Let x be a nonzero element in M. By assumption $M \not\subset Fx$. Let y be any element in $M \setminus (M \cap Fx)$. Then

$$(*) \qquad \lambda_x x + \lambda_y y = \varphi(x) + \varphi(y) = \varphi(x + y) = \lambda_{x+y}(x + y) = \lambda_{x+y} x + \lambda_{x+y} y.$$

As x, y are linearly independent, it follows that $\lambda_x = \lambda_{x+y} = \lambda_y$. In particular, if z is a nonzero element of M not in Fx, we have $\lambda_x = \lambda_z$. If $z \in Fx \subset M$, then replacing x by z in $(*)$ shows $\lambda_z = \lambda_y = \lambda_x$. The result follows. $\square$

Theorem 108.14. (Cartan-Brauer-Hua) *Let $K \subset D$ be division rings and $M \subset D$ be closed under addition and containing 1. Suppose that $xK \subset Kx$ for all x in M. Then either $M \subset K$ or $M \subset Z_D(K)$, the centralizer of K in D. In particular, if $xKx^{-1} \subset K$ for all x in D, then either $K = D$ or K lies in the center of D.*

PROOF. Suppose that $M \not\subset K$. Then M contains at least two linearly independent vectors in the right vector space D over K. Let $k \in K$. If $x \in M$, then $xk \in xK \subset Kx$, hence $xk = \lambda_x x$ for some λ_x in K. Applying the lemma to the map $\rho_k : M \to M$ given by $x \mapsto xk$ yields an element $\lambda \in K$ satisfying $xk = \lambda x$ for all x in M. Setting $x = 1$ shows $\lambda = k$. It follows that $kx = xk$ for all x in M and $k \in K$. This gives the first statement. Applying the first statement with $M = D$ yields the second statement. $\square$

Theorem 108.15. *Let D be a division ring and x an element of D having only finitely many conjugates in D. Then x has only one conjugate, i.e., lies in the center of D.*

PROOF. Let $Z_D(x) := Z_{D^\times}(x) \cup \{0\} = \{d \in D \mid xd = dx\}$, a division ring lying in D. Then the hypothesis means that $[D^\times : Z_{D^\times}(x)]$ is finite. Each conjugate axa^{-1} of x gives rise to a division algebra $Z_D(axa^{-1})$ with $Z_{D^\times}(axa^{-1})$ of finite index in D and there are finitely many such. Let K be the intersection of these finitely many division rings. Then K is a division ring satisfying $[D^\times : K^\times]$ is finite by Poincaré's Lemma (Exercise 10.16(7)). Moreover, $K^\times \triangleleft D^\times$. It follows by the Cartan-Brauer-Hua Theorem above that either $K = D$ or $K \subset Z(D)$. If $x \in Z(D)$, we are done. So suppose that $K \subset Z(D)$). Then we must have $[D^\times : Z(D^\times)] \le [D^\times : Z_{D^\times}(K^\times)] < \infty$. In particular, if $Z(D)$ is a finite field, then D is a finite division ring, hence commutative by Proposition 108.5. So we may assume that $Z(D)$ is infinite. Let $y_1, y_2, \dots$ be an infinite number of elements in $Z(D)$ and $x_0 = x$, $x_1 = x + y_1$, $x_2 = x + y_2$, $\dots$. Since $[D^\times : Z(D^\times)]$ is finite, there exist $y_i \ne y_j$ with x_i and x_j lying in the same coset of $Z(D^\times)$. Therefore, $x + y_i = z(x + y_j)$ for some $z \in Z(D^\times)$. As $y_i \ne y_j$, we have $z \ne 1$. Consequently, $(1 - z)x = zy_j - y_i$ lies in $Z(D)$ with $1 - z$ nonzero, hence invertible. Therefore, x is an element in $Z(D)$ and the result is proven. $\square$

Lemma 108.16. *Let D be a division ring with center F and $\mathcal{C}$ a conjugacy class in D such that $m_F(\mathcal{C})$ is quadratic in $F[t]$. Suppose $f \in D[t]$ has two right roots in $\mathcal{C}$. Then $f \in D[t]m_F(\mathcal{C})$ and $f(\mathcal{C}) = 0$.*

We leave the proof of this as an exercise.

Proposition 108.17. *Let F be a real closed field and $\mathcal{H}$ the Hamilton quaternions over F. Let $f \in \mathcal{H}[t]$ be nonzero. Then the following are equivalent:*

(1) The polynomial f has infinitely many roots in D.
(2) There exist elements a and b in F with b nonzero satisfying $f(a + bi) = 0 = f(a - bi)$.
(3) The polynomial f has a right factor $q(t)$ in $D[t]$ with $q \in F[t]$ a quadratic irreducible.

If these conditions hold, then f vanishes on the conjugacy class of $a + bi$

Proof. (1) $\Rightarrow$ (3): By Proposition 108.11, f has two roots in some conjugacy class $\mathcal{C}$ in $\mathcal{H}$. Then $m_F(\mathcal{C})$ lies in $F[t]$ and is an irreducible quadratic. By Lemma 108.16, we have f lies in $\mathcal{H}[t]m_F(\mathcal{C})$ (and $m_F(\mathcal{C})$ vanishes on $\mathcal{C}$).

(3) $\Rightarrow$ (1): Let c be a root of $m_F(\mathcal{C})$ in $F(i)$. Then by Theorem 108.15, x has infinitely many conjugates in $\mathcal{H}$. Each of these is a root of $m_F(\mathcal{C})$ hence of f.

(3) $\Rightarrow$ (2) is clear and (2) $\Rightarrow$ (3) follows from Lemma 108.1 and Proposition 108.2. $\square$

Exercises 108.18.

1. Let D be a division ring with center F and a, b in D both algebraic over F. Then a and b are conjugate in D if and only if $m_F(a) = m_F(b)$.
2. Prove Lemma 108.16.
3. Let R be a real closed field and $\mathcal{H}$ the Hamilton quaternions over F. Suppose that $f = \sum_{i=0}^{n}$ in $\mathcal{H}[t]$ with $a_0 \in \mathcal{H} \setminus F$ and $a_1, \dots a_n \in F$. Show that f has at most n roots in $\mathcal{H}$. In particular, show $t^n - a_0$ has exactly n solutions in D, and they all lie in $F(a_0)$.

CHAPTER XIX

Introduction to Representation Theory

In this chapter, we give applications of the Artin-Wedderburn Theorem to finite group theory. In particular, we study group homomorphisms $\varphi : G \to \mathrm{GL}_n(F)$ with G a finite group and F a field, especially the case that the characteristic of F is zero. We give two important applications, the first a famous theorem of Hurwitz on the products of sums of squares, and the second the theorem of Burnside showing that groups of order $p^a q^b$ are solvable where p and q are primes and a, b non-negative integers. To prove the second we introduce and study characters of representations, i.e., the trace of a representation $\varphi : G \to \mathrm{GL}_n(F)$.

Throughout this chapter F will denote a field and G a group.

109. Representations

Throughout this section R will denote a nonzero commutative ring.

Definition 109.1. Let R be a commutative ring and G a group (respectively monoid) (usually written multiplicatively) . The *group ring* (respectively, *monoid ring*) $R[G]$ of G over R is the free R-module on basis $G = \{g \mid g \in G\}$ made into a ring by the following multiplication:

$$(\sum_G a_g g)(\sum_G b_h h) = \sum_G c_k k \text{ with } a_h, b_h \in G \text{ almost all zero where}$$

$$c_k = \sum_{gh=k} a_g b_h.$$

We have $1_{R[G]} = 1_R e_G$ where e_G is the unity of G.

The map $\varphi_{R[G]} : R \to R[G]$ given by $r \mapsto r1_{R[G]}$ is a ring monomorphism, so we view $R \subset R[G]$. As $R \subset Z(R[G])$, the center of $R[G]$, the ring $R[G]$ is an R-algebra.

Examples 109.2. Let R be a commutative ring and G a group.

1. If $H = \{t^i \mid i \geq 0\}$, then $R[H] = R[t]$.
2. If $H = \mathbb{N}^n$ with $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$, (an additive monoid), then $R[H] \cong R[t_1, \dots, t_n]$.
3. $R[\mathbb{Z}] \cong R[t, t^{-1}]$.
4. $R[\mathbb{Z}^n] \cong R[t_1, t_1^{-1}, \dots, t_n, t_n^{-1}]$.
5. $R[G]$ is commutative if and only if G is abelian.
6. $G \subset R[G]^\times$.
7. If G is cyclic of order n, then $R[G] \cong R[t]/(t^n - 1)$.

Remarks 109.3. Let R be a commutative ring, G a (multiplicative) group, M an $R[G]$-module, and N an R-module.

1. The map $\lambda : G \to \mathrm{Aut}_R(M)$ given by $g \mapsto \lambda_g = \lambda(g) : m \mapsto gm$, is a group homomorphism. (The inverse of λ_g is $\lambda_{g^{-1}}$.)
2. A group homomorphism $\varphi : G \to \mathrm{Aut}_R N$ is called a *representation* of G. Such a φ makes N into an $R[G]$-module via
$$g \cdot x := \varphi(g)(x) \text{ for all } g \in G,\, x \in N.$$
Conclusion. A representation $\varphi : G \to \mathrm{Aut}_R N$ is equivalent to an $R[G]$-structure on N. We say that φ is *irreducible* if N is an irreducible $R[G]$-module (via φ).
3. If $\tau : G' \to G$ is a group homomorphism, then any representation of G induces a representation of G' by composition.
4. As G is an R-basis for the R-free module $R[G]$, any representation $\varphi : G \to \mathrm{Aut}_R N$ can be extended to an R-algebra homomorphism which we also write as $\varphi : R[G] \to \mathrm{End}_R N$ and also call a *representation*. Conversely, any R-algebra homomorphism $\varphi : R[G] \to \mathrm{End}_R N$ restricts to a representation $\varphi : G \to \mathrm{Aut}_R N$.

   Conclusion. A representation $\varphi : G \to \mathrm{Aut}_R N$ is equivalent to an R-algebra homomorphism $\varphi : R[G] \to \mathrm{End}_R N$.
5. Let $\lambda : G \to \sum(G)$ be the left regular representation, i.e., $x \mapsto \lambda_x : g \mapsto xg$. Then λ induces a representation $\lambda : R[G] \to \mathrm{End}_R(R[G])$ called the *(left) regular representation* of G (relative to R). It is *faithful* , i.e., $\ker \lambda = 0$.
6. Suppose that N is a finitely generated free R-module of rank n on (an ordered) basis $\mathcal{B}$. Then $N \cong R^n$ and we have a group isomorphism $\psi^{-1} : \mathrm{GL}_n(R) \to \mathrm{Aut}_R(N)$ where $\psi(T) = [T]_{\mathcal{B}}$, the matrix representation of T relative to the basis $\mathcal{B}$. If $\varphi : G \to \mathrm{Aut}_R N$ is a representation, we get a group representation $\widetilde{\varphi} = \psi\varphi : G \to \mathrm{GL}_n(R)$ called an *R-representation* of *degree n*. We say that $\widetilde{\varphi}$ *affords* N or φ *affords* N if $\mathcal{B}$ is clear. [It is also common to write $\mathrm{GL}(N)$ for $\mathrm{Aut}_R(N)$.] Hence fixing a basis $\mathcal{B}$ for a finitely generated free R-module N gives rise to a representation $\widetilde{\varphi} : G \to \mathrm{GL}_n(R)$ where n is the rank of N, hence an $R[G]$-module structure to R^n. One often writes $\widetilde{\varphi}(n)$, $n \in N$, for $\varphi(g)(n)$ if we know N.
7. Suppose that N and N' are two free R-modules of the same rank n with $\varphi : G \to \mathrm{Aut}_R(N)$ and $\psi : R \to \mathrm{Aut}_R(N')$ affording N and N', respectively. We say φ and ψ are *equivalent* and write $\varphi \sim \psi$, if there exists an R-isomorphism $T : N \to N'$ satisfying
$$\begin{tikzcd} N \arrow[r, "T"] \arrow[d, "\varphi(g)"'] & N' \arrow[d, "\psi(g)"] \\ N \arrow[r, "T"'] & N'. \end{tikzcd}$$
commutes for all g in G, i.e., if N is an $R[G]$-module via φ and N' is an $R[G]$-module via ψ in the above, then $T : N \to N'$ is an $R[G]$-isomorphism. If $N = R^n = N'$, of course, we write $GL_n(R)$ for $\mathrm{Aut}_R N$.

   Conclusion. We have $\varphi \sim \psi$ in the above if and only if $N \cong N'$ as $R[G]$-modules.

8. We say the $R[G]$-module M is *G-trivial* if $gm = m$ for all m in M and all g in G.
9. Define
$$M^G := \{m \in M \mid gm = m \text{ for all } g \in G\},$$
a submodule of M called the set of *G-fixed points* of M.
10. Let $\varphi : G \to \mathrm{GL}_n(R)$ afford the finitely generated free R-module N. Then N is G-trivial via φ if and only if φ is the trivial map, i.e., $g \mapsto I$ for all g in G. We call this the *trivial representation* of G. If $n = 1$, it is irreducible.
11. Suppose both M, N, and P are $R[G]$-modules. Then $\mathrm{Hom}_R(M, N)$ is an $R[G]$-module defined by
$$(\sigma f)(m) := \sigma\Big(f\big(\sigma^{-1}(m)\big)\Big)$$
for all $\sigma \in G$, $m \in M$, and $f \in \mathrm{Hom}_R(M, N)$.
Check. If $f \in \mathrm{Hom}_R(M, N)$, $g \in \mathrm{Hom}_R(N, P)$, then $\sigma(g \circ f) = (\sigma g) \circ (\sigma f)$ and
$$\mathrm{Hom}_{R[G]}(M, N) = \big(\mathrm{Hom}_R(M, N)\big)^G.$$
12. We view R as a trivial $R[G]$-module. If M is an $R[G]$-mmodule, then the isomorphism of $R[G]$-modules $\mathrm{Hom}_R(R, M) \to M$ given by $f \mapsto f(1)$ induces an isomorphism $\mathrm{Hom}_{R[G]}(R, M) \cong M^G$.
13. Suppose that G is a finite group. Define the *norm* of G by
$$N_G := \sum_G g \text{ in } R[G].$$
As $\sigma N_G = N_G = N_G\sigma$ for all σ in G, we have N_G lies in $R[G]^G$ and $N_G x$ lies in M^G for all x in M, i.e., $N_G : M \to M^G$ given by $x \mapsto N_G x$ is an $R[G]$-homomorphism, so $N_G M \subset M^G$. The quotient $M^G/N_G M$ is an object of study in algebraic number theory. Of course, this also says that N_G lies in the center of $R[G]$. More generally, if H is a normal subgroup of G, then N_H lies in the center of $R[G]$.
14. Suppose that G is a finite group and both M and N are $R[G]$-modules. Let $f : M \to N$ be an R-homomorphism. Define the *trace* of f to be the $R[G]$-homomorphism
$$\mathrm{Tr}_G\, f := N_G f = \sum_G \sigma f : M \to M.$$
In particular, if f is an $R[G]$-homomorphism, then $\mathrm{Tr}_G\, f = |G| f$.

Lemma 109.4. *Let R be a commutative ring and G a finite group. Suppose that M, M', N, N' are $R[G]$-modules and $f : M \to N$ an R-homomorphism. If $\varphi : M' \to M$ and $\psi : N \to N'$ are $R[G]$-homomorphisms, then $\mathrm{Tr}_G(\psi f \varphi) = \psi \circ \mathrm{Tr}_G\, f \circ \varphi$.*

Proof. By definition, $(\sigma f)(m) = \sigma\big(f(\sigma^{-1}m)\big)$, so
$$\begin{aligned}\mathrm{Tr}_G(\psi f \varphi) &= \sum_G \sigma \circ (\psi f \varphi) = \sum_G (\sigma \circ \psi) \circ (\sigma \circ f) \circ (\sigma \circ \varphi)\\ &= \psi \circ (\sum_G \sigma f) \circ \varphi = \psi \circ \mathrm{Tr}_G\, f \circ \varphi.\end{aligned}$$
□

Theorem 109.5. (Maschke's Theorem) *Let F be a field and G a finite group. Suppose that either* $\operatorname{char} F = 0$ *or* $\operatorname{char} F \not| \, |G|$. *Then $F[G]$ is semisimple.*

PROOF. Let V be an $F[G]$-module and W an $F[G]$-submodule. We must show that W is a direct summand of V as an $F[G]$-module. Let $i_W : W \to V$ denote the inclusion map. We must show that this is a split $F[G]$-monomorphism by Exercise 38.18(12). Write $V = W \oplus W'$ as F-vector spaces. [Of course, W' need not be an $F[G]$-module.] Let $\pi_W : V \to W$ be the vector space projection of V onto W. By assumption, $1/|G| \in F^\times$, so we can define the 'average'

$$\varphi := \frac{1}{|G|} \operatorname{Tr}_G \pi_W : V \to W.$$

The map φ is an $F[G]$-homomorphism and, by the lemma, we have

$$\operatorname{Tr}_G(i_W \pi_W i_W) = i_W \circ \operatorname{Tr}_G \pi_W \circ i_W = |G| \, i_W \circ \varphi \circ i_W.$$

Since

$$\operatorname{Tr}_G(i_W \pi_W i_W) = \operatorname{Tr}_G \big(i_W(\pi_W i_W)\big) = \operatorname{Tr}_G(i_W 1_W) = \operatorname{Tr}_G i_W = |G| \, i_W,$$

we conclude that $i_W \circ (\varphi \circ i_W) = i_W = i_W 1_W$ as $|G|$ is a unit in $F[G]$. As i_W is a monomorphism, we have $\varphi \circ i_W = 1_W$, i.e., i_W is a split $F[G]$-monomorphism as desired. $\square$

Maschke's Theorem allows us to apply the Artin-Wedderburn Theorem to $F[G]$ when G is a finite group and $\operatorname{char} F = 0$ or $\operatorname{char} F$ does not divide $|G|$. We also get one further piece of useful information, as we can compute the dimension of the center of $F[G]$, which we now do.

Definition 109.6. Let R be a commutative ring and G a group. If g is an element of G and the conjugacy class $\mathcal{C}(g)$ of g in G finite, we let $C_g := \sum_{\mathcal{C}(g)} h$ called the *class sum* of g in $R[G]$.

Lemma 109.7. *Let R be a commutative ring and G a finite group. If $C_{g_1}, \dots, C_{g_r}$ are the distinct class sums of G in $R[G]$, then the center $Z(R[G])$ is a free R-module on basis $\mathcal{B} := \{C_{g_1}, \dots, C_{g_r}\}$. In particular, the rank of $Z(R[G])$ is equal to the number of conjugacy classes of G.*

PROOF. Let σ be an element of G. Then for all g in G, we have $\sigma C_g \sigma^{-1} = C_g$, so C_g lies in $Z(R[G])$. Since $G = \bigvee \mathcal{C}(g_i)$ is an R-basis for $R[G]$, the set $\mathcal{B}$ is R-linearly independent. Thus we need only show that $\mathcal{B}$ spans $Z(R[G])$. Let $y = \sum_G a_g g$ lie in $Z(R[G])$. Then for each σ in G, we have $y = \sigma y \sigma^{-1} = \sum_G a_g \sigma g \sigma^{-1}$. As G is an R-basis for $R[G]$, we have $a_g = a_{\sigma g \sigma^{-1}}$ for all σ in G, and the result follows. $\square$

We now apply the Artin-Wedderburn Theorem to the case that F is an algebraically closed field and G is a finite group with $\operatorname{char} F = 0$ or $\operatorname{char} F \not| \, |G|$, i.e., Maschke's Theorem holds, e.g., $F = \mathbb{C}$, to conclude the following:

Theorem 109.8. *Let G be a finite group and F an algebraically closed field of characteristic zero or positive characteristic not dividing the order of G. Let $\mathfrak{A}_1, \dots, \mathfrak{A}_r$ be a basic set for $F[G]$ and $n_i = \dim_F \mathfrak{A}_i$, $i = 1, \dots, r$. Then*

1. $F[G] \cong \coprod_{i=1}^{r} \mathfrak{A}_i^{n_i}$ *as* $F[G]$ *modules.*
2. $F[G] \cong \bigtimes_{i=1}^{r} \mathbb{M}_{n_i}(F)$ *as rings.*
3. r *is the number of conjugacy classes of* G*.*
4. $|G| = \sum_{i=1}^{r} n_i^2$ *and at least one of the* n_i *is one.*

Proof. There are no finite dimensional F-division rings over F except for F, as any such would contain an element x not in F, but then $F(x)$ would be a commutative division ring distinct from F. It also follows that if B_i is the Wedderburn component corresponding to $\mathbb{M}_{n_i}(F)$, we have $\dim_F B_i = n_i$ is the dimension of a column space of $\mathbb{M}_{n_i}(F)$. Therefore, we have (1), (2), and (3). Finally, $Z\big(M_{n_i}(F)\big) = F$, so

$$G = \dim_{\mathbb{C}} \mathbb{C}[G] = \dim_{\mathbb{C}} \left(\bigtimes_{i=1}^{r} \mathbb{M}_{n_i}(\mathbb{C})\right) = \sum_{i=1}^{r} n_i^2,$$

and the trivial representation of G into $\mathrm{GL}_1(F) = F^{\times}$ is irreducible of degree one giving (4). $\square$

We can weaken the hypothesis of the theorem that F be algebraically close to having all the Wedderburn components matrix rings over F. We shall investigate this in the next section.

Remarks 109.9. Let G be a finite group and F a field of characteristic zero or of positive characteristic not dividing $|G|$.

1. A representation $\varphi : G \to \mathrm{GL}_1(F)$, i.e., of degree one is called a *linear representation*). By Maschke's Theorem, it must be irreducible.
2. As a special case of (1), suppose that G is an *elementary* 2-*group*, i.e., $G \cong (\mathbb{Z}/2\mathbb{Z})^n$ for some n and F is a field of characteristic different from two. Then Maschke's Theorem holds. There exist at least 2^n linear inequivalent representations of G given by the trivial representation and the $2^n - 1$ representations each defined by taking one nonzero element of G to -1 and all other elements if G to 1. Since $\dim_F F[G] = 2^n$ and $F[G]$ is semisimple, these must be all the irreducible representations of G.
3. Let $G^{ab} = G/[G, G]$. Then any representation of G^{ab} gives rise to a representation of G of the same degree via composition with the canonical epimorphism $^{-} : G \to G^{ab}$.
4. If F is algebraically closed, then any irreducible representation $\varphi : G^{ab} \to \mathbb{M}_n(F)$ must be of degree one.

Example 109.10. Let G be a finite group of order n. We give some examples of complex representations, i.e., representations $\varphi : G \to \mathrm{GL}_n(\mathbb{C})$.

1. We interpret Theorem 109.8 in the case of complex representations. Let $\mathfrak{A}_1, \dots, \mathfrak{A}_r$ be a basic set for $\mathbb{C}[G]$ and $n_i = \dim_{\mathbb{C}} \mathfrak{A}_i$, $i = 1, \dots, r$. Let φ_i be the irreducible representation afforded by $\mathfrak{A}_i$. Then the regular representation $\lambda : G \to \mathrm{GL}_n(\mathbb{C})$ is similar to

$$\begin{pmatrix} \varphi_{\mathfrak{A}_1} & n_1 & & \cdots & & & \\ & \ddots & & & & & \\ & & \varphi_{\mathfrak{A}_1} & & & & \\ \vdots & & & \ddots & & & \\ & & & & \varphi_{\mathfrak{A}_r} & n_r & \\ & & & & & \ddots & \\ & & & & & & \varphi_{\mathfrak{A}_r} \end{pmatrix}$$

2. Let $G = \langle x \rangle$ be a cyclic group of order n and ω a primitive root of unity in $\mathbb{C}$. Then $\rho_i : G \to \mathbb{C}^\times = \mathrm{GL}_1(\mathbb{C})$ determined by $x \to \omega^i$ for $i = 1, \dots n$ are all the inequivalent irreducible representations of G.
3. Let G be the dihedral group D_3 with $G = \{a, b \mid a^3 = 1 = b^2,\ bab^{-1} = a^{-1}\}$. We know the trivial representation is of degree one, and there are three conjugacy classes in D_3. Therefore, there is another a linear character and one of degree 2. The nontrivial linear character $G^{ab} \cong \mathbb{Z}/2\mathbb{Z}$ induces the irreducible character $G \to \mathbb{C}^\times$ determined by $a \mapsto 1$ and $b \mapsto -1$. Let ω be a primitive cube root of unity. The representation $\varphi : G \to \mathrm{GL}_n(\mathbb{C})$ determined by
$$a \mapsto \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \quad \text{and} \quad b \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
is the third irreducible representation.

Exercises 109.11.

1. Let F be an algebraically closed field of characteristic zero, G be a finite group and $\varphi : G \to \mathrm{GL}_n(F)$ a representation. Show that φ is irreducible if and only if for each x in G, there exists an element λ in F such that $\lambda(x) = \lambda I$.
2. Let F be an algebraically closed field of characteristic zero, G be a finite group, and z an element of the center of $F[G]$. Show if V is an irreducible $F[G]$-module, then there exists an element λ in F satisfying $zv = \lambda v$ for all $v \in V$.
3. Let F be an algebraically closed field of characteristic zero and G be a finite group. Show if there exists a faithful irreducible $F[G]$-module M (i.e., irreducible and $gm = m$ for all $m \in M$, then $g = e_G$), then the center of G is cyclic.
4. Determine all complex representations of a finite abelian group.
5. Let $G = \langle x \rangle$ be a cyclic group of order n and ω a primitive root of unity in $\mathbb{C}$. In the notation of Example 109.10(2) determine the $\mathbb{C}[G]$-submodule of $\mathbb{C}[G]$ determined by ρ_1, by ρ_i.
6. Let G be the dihedral group D_3 and ω a primitive root of unity in $\mathbb{C}$. In the notation of Example 109.10(2) determine the irreducible $\mathbb{C}[G]$-submodules of $\mathbb{C}[G]$.
7. Find all complex irreducible representations of the dihedral group D_4.
8. Find all complex irreducible representations of the quaternion group Q_3.
9. Find all real irreducible representations of the quaternion group Q_3.

110. Split Group Rings

The purpose of this section is to extend theorems about of group rings over fields with the group finite and Maschke's Theorem holds when the underlying field is algebraically closed to the case that the underlying division rings in the Wedderburn decomposition of a group ring are always the base field. This is useful in the general theory.

Definition 110.1. Let D be a division ring and M a (left) D-vector space. If A is a subring of $\mathrm{End}_D(M)$, we say that A *act densely* on M if given any D-linear independent vectors $v_1, \dots, v_m$ in M, $m > 0$, and vectors $v'_1, \dots, v'_n$ in M, then there exists a $T \in A$ satisfying $Tv_i = v'_i$ for $i = 1, \dots, n$.

Theorem 110.2. (Jacobson Density Theorem) *Let R be a nonzero commutative ring, M an irreducible R-module, and $D = \mathrm{End}_R(M)$ (a division ring by Schur's Lemma). Then*

$$\lambda : R \to \mathrm{End}_D(M) \ \textit{ by } \ a \mapsto \lambda_a : m \mapsto am$$

is a ring homomorphism. Set

$$L_R(M) := \mathrm{im}\,\lambda \subset \mathrm{End}_D(M).$$

Then $L_R(M)$ acts densely on M as a D-vector space. In particular, if $\dim_D M$ is finite, then λ is a ring epimorphism.

Proof. Let a be an element on R and f an element of $\mathrm{End}_D(M)$. Then for all m in M, we have

$$\lambda_a f(m) = af(m) = f(am) = f\lambda_a(m),$$

so λ_a lies in $\mathrm{End}_D(M)$ for all a in R, and clearly, λ is a ring homomorphism. Let $v_1, \dots, v_m$ in M be D-linearly independent with $m \geq 1$ and $v'_1, \dots, v'_n$ in M. As M is R-irreducible, $M = Rv_i$ for every $i = 1, \dots, n$. Thus the result is trivial if $m = 1$. So assume that $m > 1$. We are finished if we can establish the following:

Claim: It suffices to produce $s_m \in R$ satisfying

$$s_m v_i = \begin{cases} 0, & \text{if } i < m. \\ \neq 0, & \text{if } i = m. \end{cases}$$

Indeed, suppose that such an s_m exists. As $s_m v_m \neq 0$, we have $M = Rs_m v_m$ by irreducibility. Write $v'_m = rs_m v_m$ with $r \in R$ and let $b_m = rs_m$. In an analogous way, there exist $b_j \in R$ satisfying $b_j v_j = \delta_{ij} v'_j$, with δ_{ij} the Kronecker delta. Then $b = b_1 + \cdots + b_m$ works. This establishes the claim.

Now assume that the result is false. This means, we have

$$(*) \qquad \text{If } a \in R \text{ satisfies } av_i = 0,\ 1 \leq i \leq m-1, \text{ then } av_m = 0.$$

By induction, $M^{m-1} = \{(av_1, \dots, av_{m-1}) \mid a \in R\}$, so (*) implies if $a, b \in R$, then

$$(av_1, \dots, av_{m-1}) = (bv_1, \dots, bv_{m-1}) \Rightarrow av_m = bv_m.$$

This means that

$$\mu : M^{m-1} \to M \text{ given by } (av_1, \dots, av_{m-1}) \mapsto av_m$$

is a well-defined R-homomorphism. Let $\iota_j : M \to M^{m-1}$ be the R-homomorphism given by $m \mapsto (0, \ldots, \underbrace{m}_{j}, 0, \ldots, 0)$ as usual, and set $\varphi_j = \mu\,\iota_j$ in $D = \mathrm{End}_R(M)$. Then

$$v_m = \mu(av_1, \ldots, av_{m-1}) = \sum_{j=1}^{m-1} \mu(0, \ldots, \underbrace{m}_{j}, 0, \ldots, 0) = \sum_{j=1}^{m-1} \varphi_j(v_j),$$

contradicting $v_1, \ldots, v_m$ are D-linearly independent. $\square$

Remark 110.3. Let F be a field and A a finite dimensional F-algebra (i.e., $\dim_F A < \infty$). Let M be a finitely generated A-module, hence a finite dimensional F-vector space. Then $\mathrm{End}_A(M) \subset \mathrm{End}_F(M)$ is a subring. By definition $F \subset Z(A)$, the center of A, so

$$\lambda : A \to \mathrm{End}_F(M) \text{ by } a \mapsto \lambda_a : m \mapsto am$$

is a ring homomorphism. Set

$$L_A(M) := \mathrm{im}\,\lambda \cong A/\ker\lambda = A/\mathrm{ann}_R\, M.$$

We view

$$F \subset \mathrm{End}_A(M) \subset \mathrm{End}_F(M)$$

as subrings via $a \mapsto a1_M$.

Theorem 110.4. (Burnside) *Let F be a field, A an F-algebra (not necessarily finitely generated), and M an irreducible A-module that is also a finite-dimensional vector space over F (so cyclic as finitely generated).*

(1) If F is algebraically closed, then $F = \mathrm{End}_A(M)$.
(2) If $F = \mathrm{End}_A(M)$, then $L_A(M) = \mathrm{End}_F(M)$.

Proof. $\mathrm{End}_A(M)$ is a finite dimensional vector space over F as M is. By Schur's Lemma 102.16, $D = \mathrm{End}_A(M)$ is a division ring.

(1): Let $x \in D$. Then $F \subset Z(D)$, so $F(x) \subset D$ is a commutative division ring, i.e., $F(x)$ is a field. As $[F(x) : F] \le \dim_F D < \infty$, the field extension $F(x)/F$ is algebraic, hence $F(x) = F$, as F is algebraically closed.

(2): As $\dim_R(M)$ is finite, this follows from the Jacobson Density Theorem 110.2. $\square$

Definition 110.5. Let A be a semi-simple ring and a finite dimensional F-algebra with F a field. Suppose that $A = B_1 \oplus \cdots \oplus B_m$ is a Wedderburn decomposition. We say that A is *F-split* if B_i is a matrix ring over F for every $i = 1, \ldots, m$.

We leave the following as an exercise:

Proposition 110.6. *Let F be a field, A a semi-simple finite dimensional F-algebra. Then A is F-split if and only if $F \cong \mathrm{End}_A(M)$ for every irreducible A-module M. In particularly, if F is algebraically closed, then A is F-split.*

Remark 110.7. Suppose that F is a field, G a finite group not divisible by the characteristic of F. Then $F[G]$ is semi-simple. Suppose, in addition, that $F[G]$ is F-split and $M_1, \ldots, M_r$ a complete set of representatives for the isomorphism classes of irreducible $F[G]$-modules. Then $Z(F[G]) = \prod_{i=1}^{r} F$ and r is the number of conjugacy classes of G by Lemma 109.7.

Using the Artin-Wedderburn Theorem, Maschke's Theorem, Burnside's Lemma, and the above, we now have the following generalization of Theorem 109.8:

Summary 110.8. Let F be a field, G a finite group such that $\operatorname{char} F \not| \, |G|$ (so $F[G]$ is semi-simple by Maschke's Theorem). Suppose that $F[G]$ is F-split, e.g., if F is algebraically closed. Let

$$M_1, \ldots, M_r$$

be a complete set of representatives for the isomorphism classes of irreducible $F[G]$-modules and $n_i = \dim_F M_i$ for $1 \leq i \leq r$. Then

1. $F[G] \cong \prod_{1=1}^{r} M_i^{n_i}$ as $F[G]$-modules.
2. $F[G] \cong \bigtimes_{i=1}^{r} \mathbb{M}_{n_i}(F)$ as rings.
3. r is the number of conjugacy classes of G.
4. $|G| = \sum_{i=1}^{r} n_i^2$.
5. $L_{F[G]}(M_i) = \operatorname{End}_F(M_i)$ for $i = 1, \ldots, r$.

Exercises 110.9.

1. Let F be a field, A an F-algebra (not necessarily finitely generated), and M a completely reducible A-module. Show if $\operatorname{End}_A(M) \cong F$, then M is an irreducible A-module.
2. Prove Proposition 110.6.

111. Addendum: Hurwitz's Theorem

We shall give a nice application of the theory developed so far. If $n \geq 3$, the *general quaternion group* Q_n is the group on n generators $a_1, \ldots, a_{n-1}, \varepsilon$ subject to the relations

$$\begin{aligned} \varepsilon^2 &= 1 \\ a_i^2 &= \varepsilon && \text{for each } i = 1, \ldots, n-1 \\ a_i a_j &= \varepsilon a_j a_i && \text{for all } i, j = 1, \ldots, n-1 \text{ satisfying } i \neq j. \end{aligned}$$

For example, Q_3 is the usual quaternion group. We shall see below that this group has order 2^n.

Properties 111.1. Let Q_n, $n \geq 3$, be the general quaternion group.

1. The element ε lies in the center of Q_n.
2. The element ε satisfies $\varepsilon = a_i a_j a_i^{-1} a_j^{-1}$ for all $i \neq j$. In particular, ε lies in $[Q_n, Q_n]$.
3. $[Q_n, Q_n] = \langle \varepsilon \rangle = \{1, \varepsilon\}$ (as $Q_n/\langle \varepsilon \rangle$ is abelian and $\varepsilon \in [Q_n, Q_n]$). In particular, $Q_n/\langle \varepsilon \rangle$ is an elementary 2-group on the image of the generators $a_1, \ldots, a_{n-1}$. Hence $|Q_n/\langle \varepsilon \rangle| = 2^{n-1}$ and $|Q_n| = 2^n$.
4. There exist (at least) 2^{n-1}-linear representations of Q_n over any field of characteristic different from two by Remark 109.9(2).
5. If n is odd, then the center $Z(Q_n)$ of Q_n, is $\langle \varepsilon \rangle$:

   It suffices to show that $z = a_1 \cdots a_l$ for $1 \leq l \leq n$ is not central. But if such a z lies in $Z(Q_n)$, then

$$a_l a_1 \cdots a_l = a_1 \cdots a_l a_l = \varepsilon^{l-1} a_l a_1 \cdots a_l,$$

so $l-1$ is even, hence l is odd and $l < n-1$. But

$$a_{n-1}a_1\cdots a_l = a_1\cdots a_l a_{n-1} = \varepsilon^l a_{n-1}a_1\cdots a_l,$$

so l is even, a contradiction.

6. If n is even, then $Z(Q_n) = \langle \varepsilon, a_1\cdots a_{n-1}\rangle$:

   This is similar to the previous argument, using $a_1\cdots a_n$ is central as $n-1$ is odd, so

$$a_1\cdots a_{n-1}a_j = \varepsilon^{n-2}a_j a_1\cdot a_{n-1} = a_j a_1\cdots a_{n-1},$$

   since a_j occurs once in $a_1\cdots a_{n-1}$.

7. If g in Q_n is not central, then its conjugacy class $\mathcal{C}(g) = \{g, \varepsilon g\}$:

   As $g \notin Z(Q_n)$, we have $|\mathcal{C}(g)| > 1$. Let $^- : Q_n \to Q_n/\langle\varepsilon\rangle$ be the canonical surjection. As $\overline{Q_n}$ is abelian, $g_o \in \{g, \varepsilon g\}$ if $g_o = xgx^{-1}$ with $x \in Q_n$.

We next look at the degrees of the irreducible representations of the general quaternion group over an algebraically closed field of characteristic different from two.

Calculation 111.2. Let Q_n, $n \geq 3$, be the general quaternion group and r the number of conjugacy classes of Q_n. Let F be a algebraically closed field of characteristic different from two. As Q_n is a 2-group, we know that r is the number of irreducible F-representations of Q_n and there exist 2^{n-1} linear F-representations by the properties of established above. Therefore, using the properties above, we have

$$2^n = |G| = \sum_{i=1}^{r} d_i^2 \text{ with } d_i = 1 \text{ for } 1 \leq i \leq 2^{n-1}.$$

Case 1. n is odd:

We have

$$r = |Z(Q_n)| + \frac{|Q_n| - |Z(Q_n)|}{2} = 2 + \frac{2^n - 2}{2} = 2^{n-1} + 1.$$

So there exists one further irreducible F-representation and it is of degree d_r. Since $2^n = 1\cdot 2^{n-1} + d_r^2$, we have $d_r = 2^{\frac{n-1}{2}}$.

Case 2. n is even:

We have

$$r = |Z(Q_n)| + \frac{|Q_n| - |Z(Q_n)|}{2} = 4 + \frac{2^n - 4}{2} = 2^{n-1} + 2,$$

so there exist two further irreducible F-representations of degrees d_{r-1} and d_r and these satisfy $2^n = 1\cdot 2^{n-1} + d_{r-1}^2 + d_r^2$, hence $2^{n-1} = d_{r-1}^2 + d_r^2$. Write $d_{r-1} = 2^k f_1$ and $d_r = 2^l f_2$ with f_1, f_2 odd. It follows that $k = l$, hence

$$2^{n-2k-1} = f_1^2 + f_2^2 \equiv 2 \mod 4.$$

If either f_1 or f_2 is greater than one, we would have $4 \mid f_1^2 + f_2^2$, which is impossible. Consequently, $d_{r-1} = d_r = 2^{\frac{n-2}{2}}$.

We shall use the calculation above to solve a very nice problem in field theory. Let $x_1, \dots, x_n$ and $y_1, \dots, y_n$ be variables over a field F. We wish to know when there exists a formula

$$z_1^2 + \cdots + z_n^2 = (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) \tag{*}$$

over $F[x_1, \dots, x_n, y_1, \dots y_n]$ with the z_i bilinear in the x_i's and y_j's, i.e., bilinear when plugging in any values in F for the variables. We know that there are such formulae in certain cases, e.g.,

If $n = 1$. (This is trivial.)
If $n = 2$. (Cf. with the norm from $\mathbb{C}$ to $\mathbb{R}$.)
If $n = 4$. (Cf. with the norm form of a quaternion algebra.)
If $n = 8$. (Cf. with the norm form Cayley's octonion algebra — a non-associate algebra, i.e., satisfies all properties of an algebra except associativity, and in addition all nonzero elements have inverses.)

Theorem 111.3. (Hurwitz) *Let F be a field of characteristic not two. Then* (*) *exists if and only if $n = 1, 2, 4,$ or 8.*

Proof. (Eckmann) If $n = 1, 2, 4, 8$, such a formula exists. We have written it down except for the case of $n = 8$, which we leave to the reader to look up. Let $X = (x_1, \dots, x_n)$ and suppose that (*) holds. Let $z_i = \sum_{j=1}^n a_{ij}(X) y_j$ for $i = 1, \dots, n$. Each of the $a_{ij}(X)$ must be linear in the x_i's. We have

$$\begin{aligned}\sum_{i=1}^n z_i^2 &= \sum_{i=1}^n \big(\sum_{j=1}^n a_{ij}(X) y_j\big)^2 \\ &= \sum_{i,j=1}^n a_{ij}(X)^2 y_j^2 + 2\sum_{i=1}^n \sum_{1 \le j < k \le n} a_{ij}(X) a_{ik}(X) y_j y_k.\end{aligned}$$

Using (*) and comparing coefficients, we see that

$$\begin{aligned}\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n a_{ij}(X)^2 \qquad \text{for } j = 1, \dots, n. \\ \sum_{i=1}^n a_{ij}(X) a_{ik}(X) &= 0 \qquad \text{for all } j \ne k.\end{aligned}$$

Let $A = \big(a_{ij}(X)\big)$ in $\mathbb{M}_n\big(F[x_1, \dots, x_n, y_1, \dots, y_n]\big)$. We have

$$A^t A = \big(\sum_{i=1}^n x_i^2\big) I. \tag{†}$$

Write

$$A = A_1 x_1 + \cdots + A_n x_n \text{ with } A_i \in \mathbb{M}_n(F).$$

Substitute this into (†) and multiply out. Comparing coefficients yields equations:

$$\begin{aligned}&(1) \quad A_i^t A_j + A_j^t A_i = 0 \qquad \text{for all } i, j = 1, \dots, n \text{ and } i \ne j. \\ &(2) \quad A_i^t A_i = I \qquad \text{for all } i = 1, \dots, n.\end{aligned}$$

In particular, the A_i are orthogonal matrices. We next normalize these equations by setting

$$B_i = A_i A_n^t \text{ for } 1 \le i \le n, \text{ hence } B_n = A_n A_n^t = I.$$

The B_i's then satisfy:

$$(1') \quad B_i^t B_j + B_j^t B_i = 0 \quad \text{for all } i, j = 1, \dots, n-1 \text{ and } i \ne j.$$
$$(2') \quad B_i^t B_i = I \quad \text{for all } i = 1, \dots, n.$$

As $B_n = I$, we have $B_i^t B_n + B_n^t B_i = 0$ for $1 \le i < n$, hence

$$B_i^t = -B_i \text{ for } i = 1, \dots, n-1,$$

so $(1')$ and $(2')$ become

$$(1'') \quad B_i B_j + B_j B_i = 0 \quad \text{for all } i, j = 1, \dots, n-1 \text{ and } i \ne j.$$
$$(2'') \quad B_i^2 = -I \quad \text{for all } i = 1, \dots, n.$$

Let $\widetilde{F}$ be an algebraic closure of F and $\rho : Q_n \to \mathrm{GL}_n(\widetilde{F})$ be the representation induced by $a_i \mapsto B_i$, for $i = 1, \dots, n-1$ and $\varepsilon \mapsto -I$. Each linear representation of Q_n must take $\varepsilon \mapsto 1$ as $a_i a_j = \varepsilon a_j a_i$ for $i \ne j$ and F is commutative. Since $\operatorname{char} F \ne 2$, ρ is a direct sum of irreducible representations of degree greater than one using Exercise 103.7(1). In particular,

If n is odd, then $2^{\frac{n-1}{2}} \mid n$, hence $n = 1$.

If n is even, then $2^{\frac{n-2}{2}} \mid n$.

Since $2^{(n-2)/2} > n$ for $n \ge 10$, we have $n \le 8$ and we check:

If $n = 8$, then $2^{\frac{8-2}{2}} \mid 8$.

If $n = 6$, then $2^{\frac{6-2}{2}} \nmid 6$.

If $n = 2, 4$, then $2^{\frac{n-2}{2}} \mid n$. $\square$

A more general theorem (that we do not prove) is

Theorem 111.4. (Hurwitz-Radon Theorem) *Let $x_1, \dots, x_l$ and $y_1, \dots, y_n$ be variables over a field F of characteristic not two. Then there exists a formula*

$$z_1^2 + \cdots + z_n^2 = (x_1^2 + \cdots + x_l^2)(y_1^2 + \cdots + y_n^2)$$

over $F[x_1, \dots, x_l, y_1, \dots y_n]$ with the z_i bilinear in the x_i's and y_j's if and only if $n = 2^{4\alpha+\beta}$ with $l \le 8\alpha + 2^\beta$.

If one does not insist on the z_i being bilinear in the x_i's and y_i's and works in $F(x_1, \dots, x_n, y_1, \dots y_n)$, then Pfister showed that the product of two sums of 2^n squares is a sum of 2^n squares.

Exercise 111.5. Find all complex irreducible representations of the general quaternion group Q_n, $n \ge 3$.

112. Characters

Throughout this section F will be a field and G a group.

Definition 112.1. Let $\varphi : G \to \mathrm{GL}_n(F)$ be a representation. The map $\chi_\varphi : G \to F = \mathbb{M}_1(F)$ defined by

$$\chi_\varphi(x) := \operatorname{trace} \varphi(x) \text{ for all } x \in G$$

is called the *character* of the representation φ. The degree of the representation φ is also called the *degree* of χ_φ , so it is the integer $\chi_\varphi(e_G)$ which we shall denoted by $\chi_\varphi(1)$, i.e., write e_G as 1. In particular, χ_φ is not a group homomorphism in general. We say that χ_φ is an *irreducible character* if φ is an irreducible representation. For example, the character associated to the trivial representation called the *trivial character* is an irreducible character of degree one.

Remarks 112.2. Let V be an $F[G]$-module of dimension n as a vector space over F and $\varphi = \varphi_V : G \to \mathrm{GL}_n(F)$ a representation affording $V \cong F^n$ relative to some fixed basis. If $\psi : G \to \mathrm{GL}_n(F)$ affords V relative to another basis, then $\varphi(x)$ and $\psi(x)$ are conjugate for all x in V, hence $\chi_\varphi(x) = \chi_\psi(x)$ for all v in V. Therefore, χ_φ depends only on the equivalence class of φ. We often write χ_V for χ_φ. As $F[G]$ is F-free on basis G, the character χ_φ induces an F-linear functional $F[G] \to F$ that we also denote by χ_φ.

Proposition 112.3. *Let V, V', and V'' be $F[G]$-modules, finite dimensional over F. Then*

(1) $\chi_V(\sigma x \sigma^{-1}) = \chi_V(x)$ for all σ in G.
(2) If

$$0 \to V' \to V \to V'' \to 0$$

is an exact sequence of finite dimensional $F[G]$-modules, then

$$\chi_V = \chi_{V'} + \chi_{V''}.$$

Proof. We already showed (1). As for (2), we may assume that $V' \subset V$ with $\mathcal{B}'$ a basis for V' extended to a basis to a basis $\mathcal{B}$ for V. If $\varphi_V : G \to \mathrm{GL}_n(F)$, then with the obvious notation,

$$\varphi_V(x) = \begin{pmatrix} \varphi_{V'}(x) & * \\ 0 & \varphi_{V''}(x) \end{pmatrix},$$

and the result follows. □

Example 112.4. Let G be a finite group with $\operatorname{char} F$ not dividing $|G|$, so $F[G]$ is semisimple by Maschke's Theorem 109.5. Let

$$\lambda : G \to \mathrm{Aut}_F\big(F[G]\big) \text{ be given by } x \mapsto \lambda_x : y \mapsto xy.$$

Extend this representation linearly to an F-algebra homomorphism $\lambda : F[G] \to \mathrm{End}_F\big(F[G]\big)$. This is just the *(left) regular representation.* Let $\{\mathfrak{A}_1, \dots, \mathfrak{A}_r\}$ be the basic set for $F[G]$. Suppose that $F[G]$ is F-split, e.g., if F is algebraically closed, and $\varphi_{\mathfrak{A}_1}, \dots, \varphi_{\mathfrak{A}_r}$ are all the irreducible representations of G afforded by the $\mathfrak{A}_i$ with the $\chi_{\mathfrak{A}_i}$ the corresponding characters. Set $n_i = \deg \chi_{\mathfrak{A}_i} = \chi_{\mathfrak{A}_i}(1)$. Then we have

$$\chi_{F[G]} = \chi_\lambda = \sum_{i=1}^{r} n_i \chi_{\mathfrak{A}_i} = \sum_{i=1}^{r} \chi_{\mathfrak{A}_i}(1)\chi_{\mathfrak{A}_i}.$$

Note: λ_x with x not e_G in G permutes G without fixed points and G is a basis for $F[G]$, so
$$\operatorname{trace}\varphi_\lambda(x) = \chi_\lambda(x) = 0 \text{ for all } x \neq e_G.$$
More generally, if $\varphi : G \to \mathrm{GL}_n(F)$ is a representation affording a finite dimensional F-vector space V, then there exist unique integers $m_i \geq 0$ satisfying

1. $V \cong \coprod \mathfrak{A}_i^{m_i}$ as $F[G]$-modules.
2. $\chi_V = \sum m_i \chi_{\mathfrak{A}_i}$.
3. $\chi_V(1) = \dim_F V$.

We now show that if F is a field of characteristic zero, then the equivalence class of a representation of a finite group $G \to \mathrm{GL}_n(F)$ is completely determined by its character. This is false if F has positive characteristic dividing the order of G.

Theorem 112.5. *Let F be a field of characteristic zero, G a finite group, and V, V' two finitely generated $F[G]$-modules. Then*
$$V \cong V' \text{ if and only if } \chi_V = \chi_{V'}.$$

Proof. ($\Rightarrow$) is trivial.

($\Leftarrow$): We know that the F-algebra homomorphism $\lambda : F[G] \to \mathrm{End}_F(V)$ induced by $x \mapsto \lambda_x : v \to xv$ affords V and that $F[G]$ is semi-simple by Maschke's Theorem. Let $\{\mathfrak{A}_1, \dots, \mathfrak{A}_r\}$ be a basic set and $B_i = B_{\mathfrak{A}_i}$, $1 \leq i \leq r$ the simple components, so $F[G] = B_1 \oplus \cdots \oplus B_r$ is the Wedderburn decomposition. Let $e_i = 1_{B_i}$, $1 \leq i \leq r$, so $e_1, \dots, e_r$ are central orthogonal idempotents. V is a finitely generated $F[G]$-module, so there exist unique $m_i \geq 0$ satisfying $V \cong \coprod \mathfrak{A}_i^{m_i}$ as $F[G]$-modules, which we view as an equality.

To prove the theorem, we need to show that the F-linear functional $\chi_V : F[G] \to F$ determines all the m_i. Fix a j, $1 \leq j \leq r$, and let $d_j = \dim_F \mathfrak{A}_j \neq 0$. [Note: We are not assuming $F[G]$ is F-split, so the value d_j is not so clear.] As e_j is central, for each i, the map $\lambda_{e_j} : \mathfrak{A}_i \to \mathfrak{A}_j$ given by $a \mapsto e_j a$ is an $F[G]$-homomorphism. Viewing $\mathfrak{A}_i$ as a B_i-module, we have $\lambda_{e_j} = \delta_{ij} 1_{B_j}$. Thus we can view $\lambda_{e_j} : V \to V$ by $v \mapsto e_j v$ as an $F[G]$-homomorphism and conclude that $\operatorname{im}\lambda_{e_j} = \lambda_{e_j}(V) = \mathfrak{A}_j^{m_j}$. Let $\mathcal{B}_{kj}$ be an (ordered) F-basis for the kth component of $\mathfrak{A}_j^{m_j}$. Then $\mathcal{B}_j = \mathcal{B}_{1j} \cup \cdots \cup \mathcal{B}_{m_j j}$ is an F-basis for $\mathfrak{A}_j^{m_j}$ and $\mathcal{B} = \bigcup \mathcal{B}_j$ is an F-basis for V. We then have $[\lambda_{e_j}]_{\mathfrak{B}}$ is the matrix
$$\begin{pmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & I & m_j & & & \\ & & & \ddots & & & \\ & & & & I & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$
with I the $d_j \times d_j$ identity matrix. Consequently,
$$\chi_V(e_j) = \operatorname{trace}[\lambda_{e_j}]_{\mathfrak{B}} = d_j m_j,$$

so

$$m_j = \frac{\chi_V(e_j)}{d_j} = \frac{\chi_V(e_j)}{\dim_F \mathfrak{A}_j} \text{ in } F$$

is determined by χ_V. □

Exercise 112.6. Let G be a finite p-group and F be a field of positive characteristic p dividing $|G|$. Suppose that $\varphi : G \to \mathrm{GL}_n(F)$ is an irreducible representation, show that φ is the trivial representation, hence $\chi_\varphi = n1_F$. In particular, if V and V' are finitely generated $F[G]$-modules (hence finite dimensional vector spaces over F), then $\chi_V = \chi_{V'}$ if $\dim V \equiv \dim V' \mod p$.

113. Orthogonality Relations

Throughout this section, we shall let $\widetilde{F}$ denote an algebraic closure of the field F.

Definition 113.1. A group G is called a *torsion group* if every element of G has finite order. (G may be infinite.)

Proposition 113.2. *Let G be a torsion group, V an n-dimension vector space over F, V an $F[G]$-module. Let x be an element of G and N the order of the cyclic subgroup $\langle x\rangle$ in G. Then $\chi_V(x)$ is a sum of n terms in which each term is an Nth root of unity in an algebraic closure $\widetilde{F}$ of F.*

Proof. Let $N = \dim_F V$ and $\varphi : G \to \mathrm{GL}_N(F)$ afford V. Set $\alpha = \varphi(x)$ and $n = |\langle x\rangle|$. Taking a Jordan canonical form of α in $\mathrm{GL}_N(\widetilde{F})$ we see that trace α is a sum of N elements each of which is an eigenvalue of α (counted with multiplicity). So it suffices to show that every eigenvalue of α is an nth root of unity in $\widetilde{F}$. If $v \in V$ is a nonzero eigenvector of α with eigenvalue ε, then $x^n = e_G$ implies that $\alpha^n = 1$, so $v = \alpha^n v = \varepsilon^n v$. Consequently, $1 = \varepsilon^n$ as needed. □

Recall that the integral closure of $\mathbb{Z}$ in K, $K/\mathbb{Q}$ a field extension, is denoted by $\mathbb{Z}_K$. Then the proposition implies:

Corollary 113.3. *Let G be a torsion group and n a positive integer satisfying $x^n = e$ for every element x in G. Suppose that F is an algebraically closed field of characteristic zero and ω a primitive nth root of unity in F. If $\varphi : G \to \mathrm{GL}_N(F)$ is a representation, then $\chi_\varphi : G \to \mathbb{Z}_{\mathbb{Q}(\omega)}$.*

Definition 113.4. Let $\chi : G \to F$ be a character. The *kernel* of χ is defined to be the set

$$\ker\chi := \chi^{-1}\big(\chi(1)\big) = \{x \in G \mid \chi(x) = \chi(1)\}.$$

Proposition 113.5. *Let G be a finite group, F a field of characteristic zero, and $\varphi : G \to \mathrm{GL}_n(F)$ a representation. Then $\ker\chi_\varphi = \ker\varphi$. In particular, $\ker\chi_\varphi$ is a normal subgroup of G.*

Proof. Certainly, $\ker\varphi \subset \ker\chi_\varphi$, so we need only show the reverse inclusion. So suppose that $\chi_\varphi(x) = \chi_\varphi(1)$. By the proposition, we have

$$n = \chi_\varphi(1) = \varepsilon_1 + \cdots + \varepsilon_n$$

with each ε an $|G|$th root of unity over F hence over $\mathbb{Q}$ (and viewed in $\mathbb{C}$). So $n \leq |\varepsilon_1| + \cdots + |\varepsilon_n| \leq n$. Check if $|\varepsilon_1 + \cdots + \varepsilon_n| = |\varepsilon_1| + \cdots + |\varepsilon_n|$, then $\varepsilon_i = \varepsilon_j$ for all i, j. It follows that $n = \varepsilon_1 + \cdots + \varepsilon_n = n\varepsilon_1$, hence $\varepsilon_i = \varepsilon_1$ for all i. Since the ε_i are the eigenvalues of $\varphi(x)$ in an algebraic closure of F by the previous proof, we must have the characteristic polynomial of $\varphi(x)$ is $f_{\varphi(x)} = (t-1)^n$. Since $\varphi(y)$ is a root of $t^{|G|} - 1$ for all $y \in G$, i.e, $\varphi(y)^{|G|} - 1 = 0$, the matrix $\varphi(x)$ is a root of the gcd of $(t-1)^n$ and of $t^{|G|} - 1$. The latter polynomial cannot have multiple roots, as we are working in characteristic zero, so this gcd is $t - 1$. It follows that $\varphi(x) = I$ and x lies in $\ker \varphi$. $\square$

Theorem 113.6. *Let G be a finite group and $\varphi : G \to \mathrm{GL}_n(F)$ and $\varphi' : G \to \mathrm{GL}_m(F)$ two irreducible representations. Let*

$$a_{ij}, a'_{kl} : G \to F \ \ \text{satisfy} \ \ \varphi(x) = \big(a_{ij}(x)\big), \ \varphi'(x) = \big(a'_{kl}(x)\big)$$

for all x in G, i.e., a_{ij}, a'_{kl} are the coordinate functions of φ, φ', respectively. Then:

(1) If φ and φ' are not equivalent, then

$$\sum_G a_{ij}(x) a'_{kl}(x^{-1}) = 0 \ \ \text{for all} \ \ i, j, k, l.$$

(2) Suppose that $\mathrm{char}\, F$ *is zero or does not divide the order of G and $F[G]$ is F-split. Then n is nonzero in F and*

$$\sum_G a_{ij}(x) a_{kl}(x^{-1}) = \delta ij \, \delta jk \, \frac{|G|}{n}.$$

Proof. Let V with (ordered) basis $\mathcal{B} = \{v_1, \dots, v_n\}$ be afforded by φ and V' with (ordered) basis $\mathcal{B}' = \{v'_1, \dots, v'_m\}$ be afforded by φ'. Check if $f : V \to V'$ is an F-linear transformation, then $Tf : V \to V'$ given by

$$Tf = \sum_G \varphi'(x^{-1}) \circ f \circ \varphi(x),$$

i.e.,

$$Tf\big(\varphi(\sigma)v\big) = Tf\big(\sigma(v)\big) = \sigma Tf(v) = \varphi'(\sigma) Tf(v)$$

for all v in V and for all σ in G, is an $F[G]$-homomorphism. (Cf. Tr_G.) For fixed i and l, define an F-linear transformation f_o by

$$f_o : V \to V' \ \text{ given by } \begin{cases} v_i \mapsto v_l \\ v_j \mapsto 0 & \text{if } j \neq i. \end{cases}$$

(1): By assumption $V \not\cong V'$ as $F[G]$-modules. Since V and V' are irreducible $F[G]$-modules, we must have $f_o = 0$. As $[f_o]_{\mathcal{B},\mathcal{B}'} = \big((\delta_{rl}\delta_{si})_{rs}\big) = \big((f_o)_{rs}\big)$ and

$$\begin{aligned} (Tf_o)_{kj} &= \sum_G \sum_{r,s} \big(\varphi'(x^{-1})\big)_{kr} (f_o)_{rs} \big(\varphi(x)\big)_{sj} \\ (*) \qquad &= \sum_G \sum_{r,s} a'_{kr}(x^{-1}) \delta_{rl} \delta_{si} a_{sj}(x) \\ &= \sum_G a'_{kl}(x^{-1}) a_{ij}(x), \end{aligned}$$

we have proven (1) as $Tf_o = 0$.
(2): As V is $F[G]$-irreducible and F-split, we have $F = \operatorname{End}_{F[G]}(V)$. Let $\mathcal{B}' = \mathcal{B}$ and $v_i' = v_i$ for all v_i, then $Tf_o = \lambda(f_o)I$ for some $\lambda(f_o)$ in F. Consequently, $\operatorname{trace}(Tf_o) = n\lambda(f_o)$. However, we also have

$$\begin{aligned}\operatorname{trace}(Tf_o) &= \operatorname{trace}\big(\sum_G \varphi(x^{-1})[f_o]_{\mathcal{B}}\varphi(x)\big)\\ &= \sum_G \operatorname{trace}\big(\varphi(x^{-1})[f_o]_{\mathcal{B}}\varphi(x)\big)\\ &= |G|\operatorname{trace}([f_o]_{\mathcal{B}}) = |G|\,\delta_{il}.\end{aligned}$$

Hence n is nonzero in F and

$$Tf_o = \lambda(f_o)I = \frac{|G|}{n}\,\delta_{il}\,I.$$

Plugging this into (*) yields

$$\frac{|G|}{n}\,\delta_{il}\delta_{kj} = (Tf_o)_{kj} = \sum_G a_{kl}(x^{-1})a_{ij}(x). \qquad \square$$

Corollary 113.7. (Frobenius-Schur Theorem) *Let G be a finite group and F a field of characteristic zero or with $\operatorname{char}(F) \nmid |G|$. Suppose that $F[G]$ is F-split and $\varphi^{(k)} : G \to \operatorname{GL}_{n_k}(F)$, $k = 1,\dots,s$, is a full set of inequivalent irreducible representations of G. If $\varphi^{(k)}(g) = \big(\alpha^k_{ij}(g)\big)$ for all $g \in G$ and $k = 1,\dots,s$, then we have:*

(1) $\mathcal{B} := \{a^{(k)}_{ij} \mid i,j = 1,\dots,n_k,\ k = 1,\dots,s\}$ is a linearly independent set.
(2) $s = r$, the number of conjugacy classes of G.
(3) $|\mathcal{B}| = |G|^2$.
(4) $\mathcal{B}$ is a basis for for the dual space of $F[G]$, $\operatorname{Hom}_F(F[G], F)$.

Proof. Statement (1) follows from Theorem 113.6, (2) and (3) by Summary 110.8, and (4) from (1), (2), and (3). $\square$

Of course, we call two characters $\chi, \chi' : G \to F$ *inequivalent* if they are not equivalent.

Theorem 113.8. (Orthogonal Relations) *Let G be a finite group and χ, χ' two inequivalent characters.*

(1) If both χ and χ' are irreducible, then

$$\sum_G \chi(x)\chi'(x^{-1}) = 0.$$

(2) Suppose that $\operatorname{char} F$ is zero or does not divide the order of G and $F[G]$ is F-split. If χ is irreducible of degree n, then $\chi(1) = n \neq 0$ in F and

$$\frac{1}{|G|}\sum_G \chi(x)\chi(x^{-1}) = 1.$$

(3) If $\operatorname{char} F = 0$, then
(a) $\sum_G \chi(x)\chi(x^{-1})$ is an integer.

(b) $|G|$ *divides* $\sum_G \chi(x)\chi(x^{-1})$ *in* $\mathbb{Z}$.

(c) If $|G| = \sum_G \chi(x)\chi(x^{-1})$*, then* χ *is irreducible.*

Proof. Let $\chi = \chi_\varphi$ and $\chi' = \chi_{\varphi'}$ with $\varphi = (a_{ij})$ and $\varphi' = (a'_{ij})$ the representations giving χ and χ' respectively.

(1): By the theorem,

$$\begin{aligned}\sum_G \chi(x)\chi'(x^{-1}) &= \sum_G \Big(\big(\sum_i a_{ii}(x)\big)\big(\sum_j a_{jj}(x^{-1})\big)\Big)\\ &= \sum_{i,j}\sum_G a_{ii}(x)a'_{jj}(x^{-1}) = 0.\end{aligned}$$

(2): If $n = \chi(1)$, we know n is nonzero in F by the theorem and

$$\begin{aligned}\sum_G \chi(x)\chi(x^{-1}) &= \sum_G \Big(\big(\sum_i a_{ii}(x)\big)\big(\sum_j a_{jj}(x^{-1})\big)\Big)\\ &= \sum_{i,j}\sum_G a_{ii}(x)a'_{jj}(x^{-1}) = \sum_{ij}\frac{|G|}{n} = n\frac{|G|}{n} = |G|.\end{aligned}$$

(3): If $\widetilde{F}$ is an algebraic closure of F, then we can view χ as a character $\chi : G \to \widetilde{F}$. Let $\chi_1, \dots, \chi_r$ be all the irreducible characters $G \to \widetilde{F}$. Then $\chi = \sum m_i\chi_i$ for some non-negative integers m_i. We see that

$$(*)\qquad \begin{aligned}\sum_G \chi(x)\chi(x^{-1}) &= \sum_G \big(\sum_i m_i\chi_i(x)\big)\big(\sum_j m_j\chi_j(x^{-1})\big)\\ &= |G|\sum_i m_i^2\end{aligned}$$

by (1) and (2). Since $\sum m_i^2$ is an integer and characteristic of F is zero, we have $\sum_G \chi(x)\chi(x^{-1})$ and $|G|$ dividing $\sum_G \chi(x)\chi(x^{-1})$.

Finally suppose that $|G| = \sum_G \chi(x)\chi(x^{-1})$. Then by (*), we have $1 = \sum m_i^2$, so there exists a i such that $m_i = 1$ and $m_j = 0$ for all $j \neq i$, i.e., χ is irreducible as a character when viewed as $\chi : G \to \widetilde{F}$.

Claim. $\chi : G \to F$ is an irreducible:

Let V be the F-vector space with basis $\mathcal{B}$ afforded by φ (recall $\chi = \chi_\varphi$) and W a nonzero $F[G]$-submodule of V with F-basis $\mathcal{C}$. Let $\widetilde{W} \subset \widetilde{V}$ be the $\widetilde{F}$-vector spaces with bases $\mathcal{C}$ and $\mathcal{B}$, respectively. Both are $\widetilde{F}[G]$-modules with $\widetilde{W}$ a submodule of $\widetilde{V}$. As $\chi : G \to \widetilde{F}$ is afforded by $\widetilde{V}$ and is irreducible, we see that $\widetilde{W} = \widetilde{V}$. So

$$\dim_F W = \dim_{\widetilde{F}} \widetilde{W} = \dim_{\widetilde{F}} \widetilde{V} = \dim_F V < \infty.$$

It follows that $W = V$ and $\chi : G \to F$ is irreducible. □

Theorem 113.9. *Let* G *be a finite group. Suppose that* char F *is zero or does not divide the order of* G *and* $F[G]$ *is* F*-split. If* $\chi_1, \dots, \chi_r$ *are all the irreducible characters of* G

and x, y lie in G, then

$$\sum_{i=1}^{r} \chi_i(x)\chi_i(y^{-1}) = \begin{cases} 0, & \text{if } y \notin \mathcal{C}(x). \\ |Z_G(x)|, & \text{if } y \in \mathcal{C}(x). \end{cases}$$

PROOF. Let $x_1, \dots x_{r'}$ be a system of representatives for the conjugacy classes of G. As $F[G]$ is F-split, $r = r'$. Let $C = (c_{ij})$ with $c_{ij} = \chi_i(x_j)$ $\big($independent of the choice of $x \in \mathcal{C}(x_j)\big)$. This matrix is called the *character table* of G. Let $D = (d_{ij})$ with

$$d_{ij} = \frac{|\mathcal{C}(x_i)|}{|G|}\chi_j(x_i^{-1}) = \frac{1}{|Z_G(x_i)|}\chi_j(x_i^{-1}).$$

Check that

$$\begin{aligned}(CD)_{ij} &= \sum_k \frac{|\mathcal{C}(x_k)|}{|G|}\chi_i(x_k)\chi_j(x_k^{-1}) \\ &= \frac{1}{|G|}\sum_G \chi_i(x)\chi_j(x^{-1}) \\ &= \frac{1}{|G|}\,|G|\delta_{ij} = \delta_{ij}.\end{aligned}$$

Thus $CD = I$, hence $DC = I$ and

$$(DC)_{ij} = \sum_{k=1}^{r} \frac{1}{|Z_G(x_i)|}\chi_k(x_i^{-1})\chi_k(x_j).$$

Let $x \in \mathcal{C}(x_i)$ and $y \in \mathcal{C}(x_j)$. If $i \neq j$, we see that $(DC)_{ij} = 0$ leads to $\sum_{k=1}^{r} \chi_k(x)\chi_k(y^{-1}) = 0$ and if $i = j$, then $(DC)_{ii} = 1$ leads to

$$|Z_G(x_i)| = \frac{|G|}{|\mathcal{C}(x)|} = \sum_{k=1}^{r} \chi_k(x)\chi_k(x^{-1}) = \sum_{k=1}^{r} \chi_k(x)\chi_k(y^{-1}). \qquad \square$$

Conclusion 113.10. We have shown if $\operatorname{char} F$ is zero or does not divide the order of G with $F[G]$ being F-split, $\chi_1, \dots, \chi_r$ being all the irreducible characters of G, and $x_1, \dots, x_r$ being a system of representatives of the conjugacy classes of G, then for all $i, j = 1, \dots, r$, and $x \in G$, we have

$$\sum_{g \in G} \chi_i(x)\chi_j(x^{-1}) = |G|\delta_{ij}$$

$$\sum_{k=1}^{r} \chi_k(x_i)\chi_k(x_j^{-1}) = |Z_G(x_i)|\delta_{ij}.$$

Definition 113.11. Let F be a field and G a finite group. A function $f : G \to F$ is called a *class function on* G if $f(g) = f(xgx^{-1})$ for all $x \in G$. For example, every character of G is a class function on G. Let $\operatorname{class}_F(G)$ denote the set of class functions $G \to F$. This is clearly an F-vector space. Moreover, $\operatorname{class}_F(G)$ is an r-dimensional F-vector space. Indeed if $\{C_1, \dots, C_r\}$ is the set conjugacy classes of G, then the characteristic functions $f_i \in \operatorname{class}_F(G)$, $i = 1, \dots, r$, i.e., $f_i(x) = \delta_{ij}$ for $x \in C_j$, form a basis for $\operatorname{class}_F(G)$.

Corollary 113.12. *Let G be a finite group. Suppose that* $\operatorname{char} F$ *is zero or does not divide the order of G and $F[G]$ is F-split. Let $\chi_1, \dots, \chi_s$ be all the irreducible characters of G. Then $\{\chi_1, \dots, \chi_s\}$ is a basis for* $\operatorname{class}_F(G)$.

PROOF. The irreducible characters $\chi_1, \dots, \chi_s$ are linearly independent by the Orthogonality Relations (Theorem 113.8). As s is the number of conjugacy classes by Summary 110.8, the result follows. $\square$

Remark 113.13. Let G be a finite group. We look at complex characters for G. Define an inner product on $(\mathbb{C}[G])^* = \operatorname{Hom}_{\mathbb{C}}(\mathbb{C}[G], \mathbb{C})$, the dual space of $\mathbb{C}[G]$, by

$$\langle\ ,\ \rangle = \langle\ ,\ \rangle_{\mathbb{C}} : (\mathbb{C}[G])^* \times (\mathbb{C}[G])^* \to \mathbb{C}$$

given by

$$\langle f, h\rangle = \langle f, h\rangle_{\mathbb{C}} = \frac{1}{|G|} \sum_G f(x)\overline{h}(x)$$

where $\overline{h}$ is defined by $\overline{h}(x) := \overline{h(x)}$ for x in $\mathbb{C}[G]$. Since $\chi(x)$ is a sum of roots of unity for a character χ of G and $\chi(x^{-1}) = \overline{\chi(x)}$, we have

$$\langle \chi, \chi'\rangle = \frac{1}{|G|} \sum_G \chi(x)\overline{\chi'(x)} = \frac{1}{|G|} \sum_G \chi(x)\chi'(x^{-1})$$

for characters χ, χ' of G, where we restrict $\langle\ ,\ \rangle$ to the subspace spanned by the characters. We also have

$$\overline{\chi} : G \to \mathbb{C} \text{ given by } \overline{\chi(x)} := \chi(x^{-1})$$

is a character. Moreover, χ is irreducible if and only if $\overline{\chi}$ is. So complex characters come in two flavors, *real*, i.e., those have values in $\mathbb{R}$ and *complex pairs* χ, $\overline{\chi}$. i.e., those complex characters arising with non-real values.
Let $\mathcal{B} = \{\chi_1, \dots, \chi_r\}$ be the set of all irreducible complex characters of G . By Conclusion 113.12 and the Orthogonal Relations (Theorem 113.8), $\mathcal{B}$ is an orthonormal basis for the complex inner product space $\mathbb{C}[G]^*$ via $\langle\ ,\ \rangle$ and if χ is a complex character of G, then $\chi = \sum_{i=1}^r a_i\chi_i$ for some non-negative integers $a_i = \langle \chi, \chi_i\rangle$, $i = 1, \dots, r$. Therefore, $\langle \chi, \chi\rangle = \sum_{i=1}^r a_i^2 = |G|^2$. It follows by Summary 110.8, if W and W' are finite dimensional $\mathbb{C}[G]$-modules, then

$$\dim_{\mathbb{C}} \big(\operatorname{Hom}_{\mathbb{C}[G]}(W, W')\big) = \langle \chi_W, \chi'_W\rangle_{\mathbb{C}}.$$

Examples 113.14. 1. Let G be the cyclic group $\langle a\rangle$ of order 3 and ω a cube root of unity. Then the character table over the complex numbers is:

	1	a	a^2
χ_1	1	1	1
χ_2	1	ω	ω^2
χ_3	1	ω^2	ω

2. Let G be the dihedral group $D_3 = \langle a, b \mid a^3 = 1 = b^2,\ bab^{-1} = a^{-1}\rangle$ with 1, a, b a system of representatives of the conjugacy classes. Then the character table over the complex numbers is:

	1	a	b
χ_1	1	1	1
χ_2	-1	1	-1
χ_3	2	-1	0

Exercises 113.15.

1. (Idempotent Theorem) Let G be a finite group. Suppose that $\operatorname{char} F$ is zero or does not divide the order of G and $F[G]$ is F-split. Let $\mathfrak{A}$ be an irreducible $F[G]$-module and $B_{\mathfrak{A}}$ the simple component corresponding to $\mathfrak{A}$. Then show the unit $f = 1_{B_{\mathfrak{A}}}$ of $B_{\mathfrak{A}}$ satisfies
$$f = \frac{\chi_{\mathfrak{A}}(1)}{|G|} \sum_G \chi_{\mathfrak{A}}(x^{-1})x.$$
2. Let G be a finite group and V a finite dimensional $\mathbb{C}[G]$-module affording $\varphi : G \to \mathrm{GL}(V)$. Show if $g \in G$, then
   (*i*) $\varphi(g)$ is diagonalizable.
   (*ii*) $\varphi(g)$ is equal to the sum (counted with multiplicity) of the eigenvalues of $\varphi(g)$ all of which are $\chi_V(1)$ nth roots of unity.
   (*iii*) $|\chi_V(g)| \leq \chi_V(1)$.
3. Show the complex character $\chi_{\mathbb{C}[G]}$ of the regular representation of a finite group G (called the *regular character of* G) satisfies
$$\chi_{\mathbb{C}[G]}(x) = \begin{cases} |G| & \text{if } x = 1 \\ 0 & \text{if } x \neq 1. \end{cases}$$
Show if χ_V is an irreducible complex character of G, then $\langle \chi_V, \chi_{\mathbb{C}[G]}\rangle = \deg \chi_V$.

4. Let G be a finite group and N a normal subgroup of G. Let $\varphi : G \to G/N$ be the canonical epimorphism. Let χ be a complex character on G/N. Show that $\tilde{\chi} = \chi \circ \varphi$ is a complex character on G and irreducible if and only if χ is irreducible. In addition, show that $\tilde{\chi}_{\mathbb{C}[G/N]} = \sum \chi(1)\chi$ where the sum is taken over all irreducible complex characters on G satisfying $N \subset \ker \chi$ and $\chi_{\mathbb{C}[G/N]}$ is the regular character of G/N. (Cf. the previous exercise) .
5. Let G be the dihedral group $D_4 = \langle a, b \mid a^4 = 1 = b^2,\ bab^{-1} = a^{-1}\rangle$ with 1, a^2, a, b, ab a system of representatives of the conjugacy classes. Show the character table over the complex numbers is

	1	a^2	a	b	ab
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	1	-1	1	-1
χ_4	1	1	-1	-1	1
χ_5	2	-2	0	0	0

6. Fill in the details to Remark 113.13

114. Burnside's $p^a q^b$ Theorem

Throughout this section, we let $\widetilde{F}$ denote an algebraic closure of the field F. Recall that if K is an algebraically closed field of characteristic zero, we shall view an algebraic closure of $\mathbb{Q}$ in K to lie in $\mathbb{C}$ by taking an appropriate isomorphic image.

Theorem 114.1. (Arithmetic Lemma) (Frobenius) *Let F be a field of characteristic zero with algebraic closure $\widetilde{F}$ and G a finite group such that $F[G]$ is F-split. If $\chi : G \to F$ is an irreducible character and x an element in G, then*

$$\frac{|\mathcal{C}(x)|}{\chi(1)}\chi(x) \text{ is an algebraic integer, i.e., lies in } \mathbb{Z}_{\widetilde{\mathbb{Q}}},$$

where $\widetilde{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ in $\widetilde{F}$.

Proof. Let $\chi = \chi_V$ with V the irreducible $F[G]$-module afforded by $\varphi : F[G] \to \operatorname{End}_F(V)$ and $C = C_x = \sum_{\mathcal{C}(x)} y$, the class sum of x in $F[G]$. We know that C lies in the center $Z\bigl(F[G]\bigr)$ by Lemma 109.7. By definition, if $z \in Z\bigl(F[G]\bigr)$, then $\varphi(y)\varphi(z) = \varphi(z)\varphi(y)$ for all $y \in F[G]$ means that $\varphi(z)$ lies in $\operatorname{End}_{F[G]}(V)$. As $F[G]$ is F-split, $\operatorname{End}_{F[G]}(V) = F1_V$ by Burnside's Theorem 110.4. Therefore, $\varphi(z)$ lies in $F1_V$ for all $z \in Z\bigl(F[G]\bigr)$. In particular, $\varphi(C) = \lambda 1_v$ for some λ in F. Let $n = \chi(1) = \dim V$ and $\mathcal{B}$ a (ordered) basis for V. We know that $\chi(x)$ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ with $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$ by Lemma 113.3, so

$$\chi(1)\lambda = n\lambda = \operatorname{trace}\bigl([\varphi(C)]_{\mathcal{B}}\bigr) = \operatorname{trace}\bigl([\varphi(\sum_{\mathcal{C}(x)}(y))]_{\mathcal{B}}\bigr) = |\mathcal{C}(x)|\chi(x)$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, hence

$$\lambda = \frac{|\mathcal{C}(x)|}{\chi(1)}\chi(x) \text{ lies in } \widetilde{\mathbb{Q}}.$$

We must show that λ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$.

We restrict the map φ to a ring homomorphism (hence a $\mathbb{Z}$-algebra homomorphism)

$$\varphi = \varphi|_{\mathbb{Z}[G]} : \mathbb{Z}[G] \to \operatorname{End}_F(V).$$

Since $C = \sum_{\mathcal{C}(x)} 1 \cdot z$ lies in $\mathbb{Z}(G)$, it must lie in $Z\big(\mathbb{Z}[G]\big)$. Further restrict φ to a ring homomorphism

$$\varphi = \varphi|_{Z(\mathbb{Z}[G])} : Z(\mathbb{Z}[G]) \to \operatorname{End}_F(V).$$

As above, we have $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big) \subset \operatorname{End}_{F[G]}(V) = F1_V$, i.e., this restriction means $\varphi : Z(\mathbb{Z}[G]) \to F1_V = \operatorname{End}_{F[G]}(V)$ and C lies in $Z\big(\mathbb{Z}[G]\big)$. Since G is a finite group, $\mathbb{Z}[G]$ is a finitely generated abelian group, hence so is the subring $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big)$ – using either $Z\big(\mathbb{Z}[G]\big)$ is generated by class sums or $\mathbb{Z}$ is noetherian. In particular, every element of $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big)$ is integral over $\mathbb{Z}$ by Claim 79.4. So $\varphi(C) = \lambda 1_V$ implies that λ lies in $\widetilde{\mathbb{Q}} \cap \mathbb{Z}_F \subset \widetilde{\mathbb{Q}} \cap \mathbb{Z}_{\widetilde{F}} = \mathbb{Z}_{\widetilde{\mathbb{Q}}}$, where $\widetilde{F}$ is an algebraic closure of F. □

Theorem 114.2. *Let F be a field of characteristic zero and G a finite group such that $F[G]$ is F-split. If $\chi : G \to F$ is an irreducible character, then $\chi(1) \mid |G|$ in $\mathbb{Z}$.*

Proof. Let $x_1, \dots, x_r$ be a system of representatives for the conjugacy classes of G and $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$. Then $\dfrac{|\mathcal{C}(x_i)|}{\chi(1)}\chi(x_i)$ and $\chi(x^{-1})$ lie in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, hence

$$\frac{|G|}{\chi(1)} = \sum_G \frac{1}{\chi(1)}\chi(x)\chi(x^{-1}) = \sum_{i=1}^r \frac{\mathcal{C}(x_i)}{\chi(1)}\chi(x)\chi(x^{-1})$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap \mathbb{Q} = \mathbb{Z}$. □

Schur proved a stronger result, viz., under the hypothesis of the theorem, $\chi(1) \mid [G : Z(G)]$ in $\mathbb{Z}$. We shall prove this in the Addendum 115.

Corollary 114.3. *Let F be a field of characteristic zero and G a finite group such that $F[G]$ is F-split, say $F[G] \cong \bigtimes_{i=1}^r \mathbb{M}_{n_i}(F)$. Then*

(1) r is the number of conjugacy classes of G.
(2) $|G| = \sum_{i=1}^r n_i^2$.
(3) There exists an i, $1 \le i \le r$ satisfying $n_1 = 1$.
(4) $n_i \mid |G|$ for $i = 1, \dots, r$.

Proof. We have previously shown (1) — (3) and (4) follows from the theorem. □

Lemma 114.4. (Burnside's Lemma) *Let F be an algebraically closed field of characteristic zero, G a finite group, and $\varphi : G \to \mathrm{GL}_n(F)$ an irreducible representation. If x in G satisfies $|\mathcal{C}(x)|$ is relatively prime to $n = \chi_\varphi(1)$, then either $\chi_\varphi(x) = 0$ or $\varphi(x) = \lambda I$ for some λ in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, with $\widetilde{\mathbb{Q}}$ the algebraic closure of $\mathbb{Q}$.*

PROOF. Let $\chi = \chi_\varphi$ and $m = |\mathcal{C}(x)|$, so m and n are relatively prime. We can write $1 = am + bn$ for some integers a, b; hence

$$\lambda := \frac{\chi(x)}{n} = a\frac{m}{n}\chi(b) + b\chi(x)$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, with $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$ by the Arithmetic Lemma. By Proposition 113.2 and the proof of Proposition 113.5, we know that $\chi(x) = \varepsilon_1 + \cdots + \varepsilon_n$ for some $|G|$th roots of unity ε_i in $\widetilde{\mathbb{Q}}$ that are also the eigenvalues of $\varphi(x)$ and satisfy

$$(*) \qquad |\lambda| = \frac{1}{n}|\chi(x)| = \frac{1}{n}|\varepsilon_1 + \cdots + \varepsilon_n| \leq \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|$$

with equality if and only if $\varepsilon_i = \varepsilon_j$ for all i, j.

Case 1. Not all the ε_i are equal:

By (*), we have $|\lambda| < 1$. Let ω be a primitive $|G|$th root of unity over $\mathbb{Q}$ and $L = \mathbb{Q}(\omega)$. Then λ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap L = \mathbb{Z}_L$ and ε_k lies in $\mathbb{Z}_L$ for all k. Suppose that $\varepsilon_i \neq \varepsilon_j$. Since $L/\mathbb{Q}$ is a finite Galois extension (even abelian), $\sigma(\varepsilon_i) \neq \sigma(\varepsilon_j)$ for all σ in $G(L/\mathbb{Q})$. Hence by (*), we have $|\sigma(\lambda)| < 1$ for all σ in $G(L/\mathbb{Q})$. It follows that $\mathrm{N}_{L/\mathbb{Q}}(\lambda)| = \prod_{G(L/\mathbb{Q})} |\sigma(\lambda)| < 1$ and $\mathrm{N}_{L/\mathbb{Q}}(\lambda)$ lies in $\mathbb{Z}_L \cap \mathbb{Q} = \mathbb{Z}$, so $\mathrm{N}_{L/\mathbb{Q}}(\lambda) = 0$. Therefore, $\lambda = 0$ and $\chi_\varphi(x) = \chi(x) = \lambda n = 0$.

Case 2. All the ε_i are equal, say $\varepsilon = \varepsilon_i$:

We have $\lambda = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i = \varepsilon$ and ε is the only eigenvalue of $\varphi(x)$, so the characteristic polynomial of $\varphi(x)$ is $f_{\varphi(x)} = (t-\varepsilon)^n$ in $L[t]$. As $x^{|G|} = 1$, the matrix $\varphi(x)$ is a root of $t^{|G|} - 1$. Hence $\varphi(x)$ also a root of the gcd of $(t-\varepsilon)^n$ and $t^{|G|} - 1$ which is $t - \varepsilon$ in $L[t]$, since $t^{|G|} - 1$ has no multiple roots. Therefore, $\varphi(x) = \varepsilon I = \lambda I$. □

Theorem 114.5. (Burnside) *Let G be a finite group and p a prime. If there exists an element x in G satisfying $\mathcal{C}(x) = p^e > 1$, then G is not a simple group.*

PROOF. Let $\chi_1, \ldots, \chi_r : G \to \mathbb{C}$ be all the distinct irreducible complex characters. One of these must be the trivial character, say it is χ_1. Let $n_i = \chi_i(1)$, so $n_1 = 1$ (and $\chi_1(x) = 1$). Since $\mathcal{C}(x) \neq \{x\}$, we know that $x \neq 1$, hence $1 \notin \mathcal{C}(x)$. Therefore, we have

$$0 = \sum_{i=1}^{r}\chi_i(x)\chi_i(1^{-1}) = \sum_{i=1}^{r}\chi_i(1)\chi_i(x) = 1 + \sum_{i=2}^{r} n_i\chi_i(x).$$

Case 1. There exists an integer $1 < j \leq r$ such that $p \not| \, n_j$ and $\chi_j(x) \neq 0$:

By assumption, $\chi_j(1)$ and $|\mathcal{C}(x)|$ are relatively prime. Since $\chi_j(x) \neq 0$, by Burnside's Lemma, there exists a complex number λ satisfying $\varphi_j(x) = \lambda I$, where $\chi_j = \chi_{\varphi_j}$. In particular, $\varphi_j(x)$ lies in $Z\big(\mathbb{M}_{n_j}(\mathbb{C})\big)$, so $\varphi_j(x)\varphi_j(y) = \varphi_j(y)\varphi_j(x)$ for all $y \in G$. As $|\mathcal{C}(x)| > 1$, $x \notin Z(G)$, so there exists a $y \in G$ satisfying $1 \neq xyx^{-1}y^{-1}$ lying in $\ker\varphi_j$. As $j > 1$, χ_j is not the trivial character, so $1 < \ker\varphi_j = \ker\chi_j < G$ by Proposition 113.5. Consequently, G is not simple.

Case 2. For all $1 < j \leq r$, we have $\chi_j(x) = 0$ whenever $p \not| \, n_j$:

In this case, we have

$$0 = 1 + \sum_{i=2}^{r} n_j \chi_j(x) = 1 + p \sum_{\substack{p>1 \\ p|n_j}} \frac{n_j}{p} \chi_j(x). \tag{*}$$

As n_j/p is an integer when $p|n_j$, we have $1/p$ lies in $\mathbb{Z}_{\mathbb{C}} \cap \mathbb{Q} = \mathbb{Z}$ by (*), a contradiction. Thus Case 2 cannot occur. $\square$

Theorem 114.6. (Burnside's $p^a q^b$-Theorem) *Let G be a finite group of order $p^a q^b$ with p, q distinct primes and a, b non-negative integers. Then G is a solvable group.*

Proof. By our previous work and induction, it suffices to show that G is not simple if a and b are both positive. Let Q be a Sylow q-subgroup of G and $e_G \neq x$ an element in $Z(Q)$. Then $Q \subset Z_G(x)$, hence $|\mathcal{C}(x)| = [G : Z_G(x)] = p^n$ for some integer $n \geq 0$. If $n = 0$, then $Z_G(x) = G$, so $1 < Z(G) \triangleleft G$ as $x \neq e_G$, and the result follows. If $n > 0$, then G is not simple by Burnside's Theorem. $\square$

When Burnside wrote the first edition of his historic book on finite group theory, he decided not discuss representation theory as it was not intrinsic to the theory of groups. After he proved the $p^a q^b$-Theorem in 1904, he realized that representation theory was now an essential tool in studying group theory, so he included it in a revised edition of his book. A proof of his theorem was proven avoiding representation in the 1970's. It is much more difficult than the one using representation theory.

115. Addendum: Schur's Theorem

In this section, we establish the improvement to Theorem 114.2 mentioned in §114. To do so, we assume the reader is conversant with tensor products. (Cf. §119.)

Definition 115.1. Let F be a field and G a group. Suppose that V is an irreducible $F[G]$-module that is finite dimensional as an F-vector space. We say that V is *absolutely irreducible* if $\widetilde{F} \otimes_F V$ is an irreducible $\widetilde{F}[G]$-module with $\widetilde{F}$ an algebraic closure of F.

Check 115.2. An $F[G]$-module V, finite dimensional as an F-vector space, is absolutely irreducible if and only if $\mathrm{End}_{F[G]}(V) = F$.

Lemma 115.3. *Let G, G' be (arbitrary) groups. Suppose that V is an absolutely irreducible $F[G]$-module and V' is an absolutely irreducible $F[G']$-module. Let $V \otimes_F V'$ be the $F[G \times G']$-module induced by the G-action $(g, g')(v \otimes v') = gv \otimes g'v'$ for all $g \in G$, $g' \in G'$, $v \in V$, and $v' \in V'$. Then $V \otimes_F V'$ is an absolutely irreducible $F[G \times G']$-module.*

Proof. Let $\varphi : F[G] \to \mathrm{End}_F(V)$ and $\varphi' : G' \to \mathrm{End}_F(V')$ afford V and V', respectively. By (the modified form Proposition 110.6 of) Burnside's Theorem 110.4, we know both φ and φ' are surjective. Consider the following commutative diagram:

$$\begin{tikzcd} F[G \times G'] \arrow[r, "\rho"] \arrow[d, "f"'] & \mathrm{End}_F(V \otimes_F V') \\ F[G] \otimes_F F[G'] \arrow[r, "h"] & \mathrm{End}_F(V \otimes_F V') \arrow[u, "g"'] \end{tikzcd},$$

where

$$\begin{aligned}
&\rho : (g, g') \mapsto (v \otimes v' \mapsto gv \otimes g'v') \\
&f : (g, g') \mapsto g \otimes g' \\
&g : h \otimes h' \mapsto \big(v \otimes v' \mapsto h(v) \otimes h'(v')\big),
\end{aligned}$$

for all $v \in V$, $v' \in V'$, $g \in G$, $g' \in G'$, $h \in \mathrm{End}_F(V)$, and $h' \in \mathrm{End}_F(V')$, induce the F-algebra homomorphisms in the diagram. By dimension count, the F-algebra homomorphism g is an isomorphism (as it is F-linear). The map f is clearly a surjection. As the map $F[G] \otimes_F F[G'] \to F[G \times G']$ induced by $g \otimes g' \mapsto (g, g')$ determines the inverse to f, we conclude that f is also an F-algebra isomorphism. Since $\varphi \otimes \varphi'$ is surjective, it follows that ρ is also surjective. Taking the fixed points of the action of $G \times G'$ on $V \otimes_F V'$, we see that

$$\mathrm{End}_{F[G\times G']}(V \otimes_F V') = \big(\mathrm{End}_F(V \otimes V')\big)^{G\times G'} = F.$$

Therefore, by Check 115.2, $V \otimes_F V'$ is an absolutely irreducible $F[G \times G']$-module. □

Theorem 115.4. (Schur) *Let F be a field of characteristic zero and G a finite group. Suppose that $F[G]$ is F-split and $\chi : G \to F$ is an irreducible character. Then $\chi(1) \mid [G : Z(G)]$.*

Proof. (Tate). Let $\chi = \chi_V$. By Check 115.2, V is absolutely irreducible as $\mathrm{End}_F(V) = F$. Let $n \in \mathbb{Z}^+$. Then $V^{\otimes n} := V \underbrace{\otimes_F \cdots \otimes_F}_{n} V$ is an absolutely irreducible $F[G \times \cdots \times V]$-module by Lemma 115.3. Let the map $\varphi : G \to \mathrm{Aut}_F(V)$ afford V. Since $\mathrm{End}_{F[G]}(V) = F$ and $\varphi\big(Z(G)\big) \subset \mathrm{End}_{F[G]}(V)$, as $\varphi\big(Z(G)\big)$ commutes with $\varphi(G)$, we conclude that $\varphi|_{Z(G)} : Z(G) \to F^\times$. Set

$$H = \{(g_1, \ldots, g_n) \mid g_i \in Z(G),\ g_1 \cdots g_n = 1\} \subset G \times \cdots \times G,$$

a subgroup. For all $(g_1, \ldots, g_n) \in H$ and $v_1, \ldots, v_n \in V$, we have

$$\begin{aligned}
(g_1, \ldots, g_n)(v_1 \otimes \cdots \otimes v_n) &= g_1 v_1 \otimes \cdots \otimes g_n v_n \\
&= \varphi(g_1)v_1 \otimes \cdots \otimes \varphi(g_n)v_n = \varphi(g_1)\cdots\varphi(g_n)v_1 \otimes \cdots \otimes v_n \\
&= \varphi(g_1 \cdots g_n)(v_1 \otimes \cdots \otimes v_n) = v_1 \otimes \cdots \otimes v_n,
\end{aligned}$$

as $\varphi(g_i)$ lies in $\mathrm{End}_F(V) = F$ for all $g_i \in Z(G)$, $i = 1, \ldots, n$. Since the $v_1 \otimes \cdots \otimes v_n$ generate $V^{\otimes n}$, it follows that H acts trivially on $V^{\otimes n}$.

Since $H \subset Z[G \times \cdots \times G]$, we have $H \triangleleft G \times \cdots \times G$. By the above, the H-action on $V^{\otimes n}$ is trivial. Consequently, we may view $V^{\otimes n}$ as an irreducible $F[(G \times \cdots \times G)/H]$-module. By Theorem 114.2, we see that

$$\chi(1)^n \,\Big|\, |(G \times \cdots \times G)/H| = \frac{|G|^n}{|H|}. \tag{115.5}$$

Let $g_1, \ldots, g_{n-1} \in Z(G)$. If $g_n \in G$, then $g_1 \ldots g_{n-1} g_n \in H$ if and only if $g_n = (g_1 \cdots g_{n-1})^{-1}$. Therefore,

$$\chi(1)^n \,\Big|\, \frac{|G|^n}{|Z(G)|^{n-1}}.$$

Hence, there exists an element $e = e(m) \in \mathbb{Z}$ satisfying $|G|^n/|Z(G)|^{n-1} = e\chi(1)^n$. It follows that

$$\Big(\frac{[G:Z(G)]}{|Z(G)|}\Big)^n = \Big(\frac{|G|}{|Z(G)|}\Big)^n \text{ lies in } \frac{1}{|Z(G)|}\mathbb{Z}$$

for all $n \in \mathbb{Z}^+$. Since $[G:Z(G)]/\chi(1)$ lies in $\mathbb{Q}$, it follows that we have $[G:Z(G)]/\chi(1)$ lies in $\mathbb{Z}$, i.e., $\chi(1) \mid [G:Z(G)]$. □

Exercises 115.6.

1. Prove Check 115.2.
2.
3. Show an $F[G]$-module V, finite dimensional as an F-vector space, is absolutely irreducible if and only if $K \otimes_F V$ is irreducible for all field extensions K/F.
4. Verify equation (115.5).

116. Induced Representations

In this section, we shall study how subgroups of a finite group induce representations of the full group as well as the induced character theory. In this section we shall assume familiarity of tensor products of modules over an arbitrary ring. (Cf. Section 119.)

Definition 116.1. Let F be a field and G a finite group. If V is a finite dimensional $F[G]$-module and $H \subset G$ a subgroup, then restriction of scalars to $F[H]$ induces an $F[H]$-module structure of V called the *restriction* of V to H and denoted by $\operatorname{res}^G_H(V)$. If V is a finite dimensional F-vector space and affords the representation $\sigma_V : G \to \mathrm{GL}(V)$, then we write $\operatorname{res}^G_H(\sigma_V)$ for the representation afforded by $\operatorname{res}^G_H(V)$ and $\operatorname{res}^G_H \chi$ for the character of this representation. If W is an $F[H]$-module, then define the *induced* $F[G]$*-module* by $\operatorname{ind}^G_H(W) := \mathrm{F}[G] \otimes_{F[H]} W$. For example by properties of $\otimes$, we have $\operatorname{ind}^G_H(F[H]) = F[G] \otimes_{F[H]} F[H]$ which is identified with $F[G]$. Since $F[G]$ is an $(F[G], F[H])$-bimodule, we have $\operatorname{ind}^G_H(W) = \{\sum_{g_i \in G} g_i \otimes w_i \mid g_i \in G,\ w \in W\}$ becomes an $F[G]$-module via the G-action $g(\sum_{g_i \in G} \otimes w) = \sum_{g_i \in G} gg_i \otimes w$. If W is a finite dimensional F-vector space and affords $\sigma_W : H \to \mathrm{GL}(W)$, let $\operatorname{ind}^G_H(\chi_W)$ denote the character induced by $\operatorname{ind}^G_H(W)$.

Construction 116.2. Let F be a field, G a finite group and $H \subset G$ is a subgroup. Suppose that $g_1, \dots, g_n$ is a left transversal of H in G with $g_1 = e_G = 1_{F[G]} = 1$. So we have $F[G] = \bigoplus_{i=1}^m g_i F[H]$. Therefore, if W is an $F[H]$-module,

$$\operatorname{ind}^G_H(W) = F[G] \otimes_{F[H]} W = \bigoplus_{i=1}^n (g_i F[H] \otimes_{F[H]} W) = \bigoplus_{i=1}^n g_i \otimes_{F[H]} W$$

as $\bigoplus$ and $\bigotimes$ commute and $\bigotimes_{F[H]}$ is $F[H]$-balanced. We have $g_1 \otimes_{F[H]} W = 1 \otimes_{F[H]} W$ an $F[H]$-submodule of $\operatorname{res}^G_H\big(\operatorname{ind}^G_H(W)\big)$. So $w \mapsto 1 \otimes w$ for $w \in W$ induces an $F[H]$-monomorphism $W \to \operatorname{ind}^G_H(W)$. We also have that the direct summand $g_i(1 \otimes_{F[H]} W)$ of $\operatorname{ind}^G_H(W)$ can be expressed as $g_i(1 \otimes_{F[H]} W)$. Therefore, we have $g_i(1 \otimes_{F[H]} W) \cong g_i \otimes_{F[H]} W$ as vector spaces over F. For each i, $i = 1, \dots, n$, and $g \in G$, there exists a unique j, $1 \le j \le n$, satisfying $gg_i \in g_j H$. So for each i, there also exists a unique

h_i in H satisfying $gg_i = g_j h_i$. It follows that g permutes the $\{g_i \otimes_{F[H]} W\}$ with action $g(g_i \otimes w) = g_j h_i \otimes w = g_j \otimes h_i w$ for all $w \in W$.

Now suppose that W is a finite dimensional F-vector space. Let $\sigma_W : H \to \mathrm{GL}(W)$ afford W and $\{w_1, \dots, w_m\}$ be an F-basis for W. Then for all $h \in H$, we have $hw_j = \sum_{i=1}^m \alpha_{ij}(h) w_i$ with all $\alpha_{ij}(h) \in F$. It follows that $\mathcal{B} := \{g_i \otimes w_j \mid 1 \le i \le n,\ 1 \le j \le m\}$ is an F-basis for $\mathrm{ind}_H^G(W)$ as $\bigoplus$ and $\bigotimes$ commute. With this notation, we have

$$g(g_i \otimes w_s) = g_j h_i \otimes w_s = g_j \otimes h_i w_s = \sum_{r=1}^m \alpha_{rs}(h_i)(g_j \otimes w_r) \text{ for all } 1 \le s \le m.$$

Since $gg_i = g_j h_i$ above means that $h_i = g_j^{-1} g g_i$, we can extend the maps $\alpha_{ij} : H \to F$ to maps $\dot{\alpha}_{ij} : G \to F$ defined by

$$\dot{\alpha}_{ij}(g) = \begin{cases} \alpha_{ij}(g) & \text{if } g \in H \\ 0 & \text{if } g \notin H. \end{cases}$$

Then the action of g in G on the F-basis $\mathcal{B}$ for $\mathrm{ind}_H^G(W)$ is given by

$$g(g_i \otimes w_s) = \sum_{r=1}^m \alpha_{rs}(h_i)(g_j \otimes w_r) = \sum_{j=1}^n \sum_{r=1}^m \dot{\alpha}_{rs}(g_j^{-1} g g_i)(g_j \otimes w_r).$$

Now extend σ_W to an $F[G]$-module by $\dot{\sigma}_W(g) = \big(\dot{\alpha}_{ij}(g)\big)$. Relative to the F-basis $\mathcal{C} = \{g_1 \otimes w_1, \dots, g_1 \otimes w_m, g_2 \otimes w_1, \dots, g_n \otimes w_m\}$, the matrix representation of $\mathrm{ind}_H^G(W)$ is

$$g \mapsto \sigma_{\mathrm{ind}_H^G(W)}(g) = \begin{pmatrix} \dot{\sigma}(g_1^{-1} g g_1) & \cdots & \dot{\sigma}(g_1^{-1} g g_m) \\ \vdots & & \vdots \\ \dot{\sigma}(g_n^{-1} g g_1) & \cdots & \dot{\sigma}(g_n^{-1} g g_m) \end{pmatrix}.$$

It follows that $\mathrm{ind}_H^G(\chi_W)$ is defined by

$$\mathrm{ind}_H^G(\chi_W)(g) = \sum_{i=1}^n \dot{\chi}_W(g_i^{-1} g g_i) \tag{116.3}$$

where

$$\dot{\chi}_W(g) = \begin{cases} \chi(g) & \text{if } g \in H \\ 0 & \text{if } g \notin H. \end{cases} \tag{116.4}$$

Now suppose that $|H| \in F^\times$, e.g., $\mathrm{char}(F) = 0$. Then equation (116.3) becomes

$$\mathrm{ind}_H^G(\chi_W)(g) = \frac{1}{|H|} \sum_{x \in G} \dot{\chi}_W(x^{-1} g x), \tag{116.5}$$

independent of the transversal of H in G. Moreover, in this case, we have

$$\deg\big(\mathrm{ind}_H^G(\chi_W)\big) = \deg\big(\mathrm{ind}_H^G(\chi_W)(1)\big) = [G : H]\chi_W(1).$$

Example 116.6. Let F be a field, G a finite group and H a subgroup of G of index n. Let 1_H be the trivial representation of H and $W = Fw$ afford 1_W. So $hw = w$ for all $h \in H$. In the notation of the construction above, the G-action of $g \in G$ on the summand $F(g_i \otimes w)$ that affords $\operatorname{ind}_H^G(1_W)$ is $g(g_i \otimes w) = g_j \otimes w$ with $gg_i \in g_j H$. It follows that the G-set of F-subspaces $\{F(g_i \otimes w)\}$ of $\operatorname{ind}_H^G(\chi_W)$ with this G-action is isomorphic to the G-set of left cosets G/H with action left translation on the basis $\mathcal{D} = \{w_i \mid i = 1, \dots, n\}$ with the w_i corresponding to $g_i \otimes w$. In particular, 1_H^G is equivalent to the permutation representation of G afforded by the G-set G/H by choosing an F-basis $\mathcal{D}$ defined by $gw_i = w_j$ if $gg_iH = g_jH$ for $i = 1, \dots, n$. In particular, the left regular representation $F[G]$ is isomorphic to $\operatorname{ind}_H^G(1_H)$ when $H = 1$.

Proposition 116.7. *Let F be a field and G a finite group, $H \subset G$ a subgroup of index n. Let V be an $F[G]$-module so that $\operatorname{res}_H^G(V)$ contains an $F[H]$-submodule W that satisfies $V = \bigoplus_{i=1}^n g_i W$, where $\{g_1, \dots, g_n\}$ is an transversal of H in G. Then $V = \operatorname{ind}_H^G(\chi_W)$ as an $F[H]$-module.*

Proof. The map $F[G] \times W \to V$ is checked to be balanced, so induces a surjective F-linear transformation $\varphi : F[G] \otimes_{F[H]} W \to V$ satisfying $\varphi(g_i \otimes_{F[H]} W) = g_i W$, $i = 1, \dots, n$. As $V = \bigoplus_{i=1}^n g_i W$, we have a linear transformation $V \to \operatorname{ind}_H^G(\chi_W)$ inverse to φ, so φ is an F-isomorphism. This map is, in fact, an $F[G]$-homomorphism, so an $F[G]$-isomorphism. $\square$

We leave the following as an exercise.

Proposition 116.8. *Let F be a field, G a finite group and $H_0 \subset H$ subgroups of G.*

(1) If W_1 and W_2 are $F[H]$-modules, then

$$\operatorname{ind}_H^G(W_1 \oplus W_2) \cong \operatorname{ind}_H^G(W_1) \oplus \operatorname{ind}_H^G(W_2).$$

(2) If W_0 is an $F[H_0]$-module, then

$$\operatorname{ind}_H^G\left(\operatorname{ind}_{H_0}^H(W_0)\right) \cong \operatorname{ind}_{H_0}^G(W_0).$$

(3) If W_0 is a finite dimensional $F[H_0]$-module, then

$$\operatorname{ind}_H^G\left(\operatorname{ind}_{H_0}^H(\chi_{W_0})\right) \cong \operatorname{ind}_{H_0}^G(\chi_{W_0}).$$

Theorem 116.9. (Frobenius Reciprocity – Module Form) *Let F be a field and G a finite group, $H \subset G$ a subgroup. If V is an $F[G]$- module and W is an $F[H]$-module, then*

$$\operatorname{Hom}_{F[H]}\left(W, \operatorname{res}_H^G(V)\right) \cong \operatorname{Hom}_{F[G]}\left(\operatorname{ind}_H^G(W), V\right).$$

Proof. By the Adjoint Associativity Theorem 129.9 (to be proven), we have a group isomorphism

$$\operatorname{Hom}_{F[H]}\left(W, \operatorname{Hom}_{F[G]}(F[G], V)\right) \cong \operatorname{Hom}_{F[G]}\left(\operatorname{ind}_H^G(W), V\right)$$

that is easy to see is an isomorphism of F-vector spaces. Therefore, it suffices to show that $\operatorname{Hom}_{F[G]}(F[G], V) \cong \operatorname{res}_H^G(V)$ as $F[H]$-modules where $\operatorname{Hom}_{F[G]}(F[G], V)$ is an $F[H]$-module by $(af)(x) := f(ax)$ for all $f \in \operatorname{Hom}_{F[G]}(F[G], V)$, $x \in F[G]$, and $a \in F[H]$. Let $\theta : \operatorname{Hom}_{F[G]}(F[G], V) \to V$ be the linear transformation defined by $f \mapsto f(1)$. It is easy to see that this is an F-vector space isomorphism. Then for all $x \in F[G]$, $a \in F[H]$, $f \in \mathrm{Hom}_{F[G]}(F[G], V)$, we have

$$\theta(af) = (af)(1) = f(a) = a\big(f(1)\big) = a\theta(f).$$

It follows that θ is an $F[H]$-isomorphism. $\square$

Remark 116.10. Let G be a finite group and F a subfield of $\mathbb{C}$. Suppose that $F[G]$ is F-split. Of course, G is $\mathbb{C}$-split also. Let K be the field composite of F and $\mathbb{C}$ in some larger field (with fixed embeddings). Then G is also K-split by extension of scalars. The basic sets of irreducible $F[G]$-modules and basic sets of irreducible $\mathbb{C}[G]$-modules yield isomorphic basic sets of irreducible $K[G]$-modules. Moreover the characters viewed as class functions on each are then all the same. Therefore, we can identify the irreducible characters on F and $\mathbb{C}$. In particular, Remark 113.13 holds for any such F in $\mathbb{C}$.

Therefore, by Remarks 113.13 and 116.10 we have:

Corollary 116.11. *Let F be a subfield of the complex numbers and G a finite group, $H \subset G$ a subgroup. Suppose that both $F[G]$ and $F[H]$ are F-split. If V is a finite dimensional $F[G]$-module and W is a finite dimensional $F[H]$-module, then*

(1) $\dim_F\big(\mathrm{Hom}_{F[H]}(W, \mathrm{res}^G_H(V))\big) = \langle \chi_W, \mathrm{res}^G_H(\chi_V)\rangle_H$.
(2) $\dim_F\big(\mathrm{Hom}_{F[G]}\big(\mathrm{ind}^G_H(W), V\big)\big) = \langle \mathrm{ind}^G_H(\chi_W), \chi_V\rangle_G$.

In particular,

$$\langle \chi_W, \mathrm{res}^G_H(\chi_V)\rangle_H = \langle \mathrm{ind}^G_H(\chi_W), \chi_V\rangle_G.$$

Recall the F-vector space $\mathrm{class}_F(G)$ of a finite group G has as an F-basis the irreducible characters under the conditions that $\mathrm{char}\, F = 0$ or $\mathrm{char} \not| \, |G|$ and $F[G]$ is F-split by Corollary 113.12. We generalize the character form of Frobenius Reciprocity without using the previous module form.

Theorem 116.12. (Frobenius Reciprocity – Class Function Case) *Let F be a subfield of the complex numbers, G a finite group, and H a subgroup of G. Let $\lambda \in \mathrm{class}_F(H)$ and $\mu \in \mathrm{class}_F(G)$. Define $\mathrm{ind}^G_H(\lambda) : G \to H$ by*

$$\mathrm{ind}^G_H(\lambda) := \frac{1}{|H|}\sum_{x\in G} \dot{\lambda}(x^{-1}gx),$$

where $\dot{\lambda}$ is the function λ extended to G (analogous to equation (116.4)). *Then $\mathrm{ind}^G_H(\lambda) \in \mathrm{class}_F(G)$ and*

$$\langle \mathrm{ind}^G_H(\lambda), \mu\rangle_G = \langle \lambda, \mathrm{res}^G_H(\mu)\rangle_H.$$

Proof. The first statement is clear. As for the equation, since μ is a class function and for each $y \in G$, there exist $|G|$ x's satisfying $y = x^{-1}gx$. Therefore, we have, using

Remark 113.13 and equation (116.5),

$$\begin{aligned}\langle \mathrm{ind}_H^G(\lambda), \mu\rangle_G &= \frac{1}{|G|}\frac{1}{|H|}\sum_{g\in G}\sum_{x\in G}\dot{\lambda}(x^{-1}gx)\overline{\mu(g)}\\ &= \frac{1}{|G|}\frac{1}{|H|}\sum_{g\in G}\sum_{x\in G}\dot{\lambda}(x^{-1}gx)\overline{\mu(x^{-1}gx)}\\ &= \frac{1}{|H|}\sum_{y\in G}\dot{\lambda}(y)\overline{\mu(y)} = \sum_{y\in H}\dot{\lambda}(y)\overline{\mu(y)}\\ &= \langle \lambda, \mathrm{res}_H^G(\mu)\rangle_H. \qquad \square\end{aligned}$$

Examples 116.13. 1. Let G be a finite group and $H \subset G$ a finite subgroup. Suppose that λ is a character of H over $\mathbb{C}$. Then the class function $\mathrm{ind}_H^G(\lambda)$ is character of G over $\mathbb{C}$ if and only if $\langle \mathrm{ind}_H^G(\lambda), \chi\rangle_G = \langle \lambda, \mathrm{res}_H^G(\chi)\rangle_H$ is a non-negative integer for all irreducible characters χ of G over $\mathbb{C}$ as $\mathrm{res}_H^G(\chi)$ is a character of H over $\mathbb{C}$.

2. Let D_m be the dihedral group of order $2m$ with the presentation $\langle a, b \mid a^m = b^2 = (ab)^2 = 1\rangle$. Let $A = \langle a\rangle$ and $\lambda : A \to \mathbb{C}$ be a linear character(hence irreducible). Let ζ be a primitive mth root of unity. So $\lambda(a) = \zeta^i$ for some i. Then $\mathrm{ind}_A^{D_m}(\lambda)(x) = \dot{\lambda}(x) + \dot{\lambda}(bxb^{-1})$ for $x \in A$ satisfying $bxb^{-1} = x^{-1}$ by equation (116.4). Therefore, $\mathrm{res}_A^{D_m}\big(\mathrm{ind}_A^{D_m}(\dot{\lambda})\big) = \lambda + \overline{\lambda}$. By Frobenius Reciprocity, we have

$$\langle \mathrm{ind}_A^{D_m}(\lambda), \mathrm{ind}_A^{D_m}(\lambda)\rangle_{D_m} = \langle \lambda, \mathrm{res}_A^{D_m}\big(\mathrm{ind}_A^{D_m}(\lambda)\big)\rangle_A = \langle \lambda, \lambda + \overline{\lambda}\rangle.$$

By Orthogonal Relations (Theorem 113.8), $\mathrm{ind}_A^{D_m}(\lambda)$ is irreducible if and only if $\langle \mathrm{ind}_A^{D_m}(\lambda), \mathrm{ind}_A^{D_m}(\lambda)\rangle = 1$. It follows that $\mathrm{ind}_A^{D_m}(\lambda)$ is irreducible if and only if $\lambda \neq \overline{\lambda}$. If $\mu : A \to \mathbb{C}$ is a character with $\mathrm{ind}_A^{D_m}(\lambda)$ and $\mathrm{ind}_A^{D_m}(\mu)$ both irreducible, then

$$\langle \mathrm{ind}_A^{D_m}(\lambda), \mathrm{ind}_A^{D_m}(\mu)\rangle_{D_m} = \langle \lambda, \mu + \overline{\mu}\rangle_A = \begin{cases} 0 & \text{if } \lambda \neq \mu \text{ and } \lambda \neq \overline{\mu}\\ 1 & \text{if } \lambda = \mu \text{ or } \lambda = \overline{\mu}.\end{cases}$$

There are two possible cases.

Case 1. m is odd:

In this case, there are two linear characters of D_m over $\mathbb{C}$ as the commutator of D_m has index two and $\frac{1}{2}(m-1)$ distinct characters of the form $\mathrm{ind}_A^{D_m}(\lambda)$ by the argument above. Therefore, these irreducible characters constitute $1^2 + 1^2 + 2^2\big(\frac{1}{2}(m-1)\big) = 2 + 2(m-1) = 2m$ in the sum of squares giving $|D_m|$. It follows that these must be all the irreducible characters.

Case 2. m is even:

In this case, there are four linear characters of D_m over $\mathbb{C}$ and $\frac{1}{2}(m-2)$ distinct irreducible characters of degree two. So these irreducible characters constitute $4+2^2(\frac{1}{2}(m-2) = 4 + 2(m-2) = 2m$ in the sum of squares giving $|D_m|$. It follows that these must be all the irreducible characters.

Exercises 116.14. 1. Let F be a subfield of $\mathbb{C}$ and G a finite group such that $F[G]$ is F-split. Let V and W be finite dimensional $F[G]$-modules. Show all of the following:

(i) $\chi_{V\otimes_F W} = \chi_V\chi_W$ where $V\otimes_F W$ is an $F[G]$-module with G-action induced by $g(v\otimes w) = gv\otimes gw$, for all $g\in G$, $v\in V$, and $w\in W$.
(ii) $\chi_{V^*} = \overline{\chi}_V$.
(iii) $\chi_{\mathrm{Hom}(V,W)} = \overline{\chi}_V\chi_W$.

2. Let F be a subfield of $\mathbb{C}$ and G a finite group such that $F[G]$ is F-split. Show the characters of G over F form a ring with multiplication given by the previous exercise.
3. Let F be a subfield of $\mathbb{C}$ and G a finite group such that FG is F-split. Suppose that H is a subgroup of G and χ an F-character of H. Show if $x\in G$, then
$$\mathrm{ind}_H^G(\chi)(x) = \frac{|Z_G(x)|}{|H|}\sum_{y\in C(x)\cap H}\chi(y)$$
where $\mathcal{C}(x)$ is the conjugacy class of x in G.
4. Let F be a subfield of $\mathbb{C}$ and G a finite group such that $F[G]$ is F-split. Suppose that H is a subgroup of G, $\psi\in\mathrm{class}_F(H)$ and $\psi'\in\mathrm{class}_F(G)$. Show if $\langle\psi',\lambda\rangle = \langle\psi,\mathrm{res}_H^G(\lambda)\rangle$ for all $\lambda\in\mathrm{class}_F(G)$, then $\psi' = \mathrm{ind}_H^G(\psi)$.
5. Let F be a subfield of $\mathbb{C}$ and G a finite group such that $F[G]$ is F-split. Suppose that N is a normal subgroup of G. Then using the notation of Exercise113.15(4), show that $1_N^G = \tilde{\chi}_{F[G/N]}$ (where $\chi_{F[G/N]}$ is the regular character of G/N).
6. Prove Proposition 116.8.
7. Let F be a field, G a finite group with $H\subset G$ a subgroup. Let V be an $F[G]$-module and W an $F[H]$-module. Then there exists and isomorphism $\mathrm{Hom}_{F[G]}\big(V,\mathrm{ind}_H^G(W)\big)\cong\mathrm{Hom}_{F[H]}\big(\mathrm{res}_H^G(V),W\big)$.

117. Torsion Linear Groups

In this section, we introduce further work of Burnside that focuses on the problem of conditions that force a group G to be finite. Of course, one would need that the group is a torsion group, i.e., every element has finite order. Burnside looked at the stronger condition on a group G, viz., one in which there exist a positive integer n such that $x^n = e$ for every element in the group G. Since the group $\bigtimes_{i=1}^{\infty}\mathbb{Z}/n\mathbb{Z}$ satisfies this condition yet is an infinite group, other conditions are necessary. Burnside showed if, in addition, such a group G was also a subgroup of $\mathrm{GL}_n(F)$, with F a field of characteristic zero, then G is a finite group. This is the principal result that we prove in this section. This led to a long research project on attempts to generalize it.

We begin with the following lemma:

Lemma 117.1. (Frobenius-Schur) *Let F be an algebraically closed field and G a (not necessarily finite) group, $\varphi: G\to\mathrm{GL}_n(F)$ an irreducible representation, and $a_{ij}: G\to F$, $1\le i,j\le n$, the coordinate functions of φ, i.e., $\varphi = \big(a_{ij}\big)$. Then*

(1) $\mathbb{M}_n(F) = \langle\varphi(x)\mid x\in G\rangle$.
(2) $\mathcal{B} = \{a_{ij}\mid 1\le i,j\le n\}$ is a set of F-linearly independent functions.

Proof. (1): Extending φ to the F-algebra map $\varphi: F[G]\to\mathbb{M}_n(F)$ yields (1) by Burnside's Theorem 110.4.

(2): If $\mathcal{B}$ is F-linearly dependent, then it is F-linearly dependent as a set of F-functionals $F[G] \to F$, say

$$0 = \sum_{i,j} b_{ij} a_{ij}, \quad b_{ij} \in F \text{ not all zero.}$$

Suppose that $b_{i_0 j_0} \neq 0$. Let $e_{i_0 j_0}$ be the $i_0 j_0$th matrix unit. By (1), there exists an x in $F[G]$ such that $\varphi(x) = e_{i_o j_o}$. Therefore, we have

$$0 = \sum_{i,j} b_{ij} a_{ij}(x) = \sum_{i,j} b_{ij} \delta_{ii_0} \delta_{jj_0} = b_{i_0 j_0},$$

a contradiction. Hence $\mathcal{B}$ is F-linearly independent in $\mathrm{Hom}_F(F[G], F)$. $\square$

Corollary 117.2. *Let F be an algebraically closed field and G a group. Suppose that $\varphi : G \to \mathrm{GL}_n(F)$ is an irreducible representation and $\varphi = \big(a_{ij}\big)$, $1 \leq i, j \leq n$, the coordinate functions of G. Then there exist $x_1, \dots, x_{n^2}$ elements in G such that $\{x_1, \dots, x_{n^2}\}$ is an F-basis for $\mathbb{M}_n(F)$. Moreover, if $v_k = \varphi(x_l)$, $1 \leq k \leq n^2$, is viewed in F^{n^2}, then $\{v_1, \dots, v_{n^2}\}$ is a basis for F^{n^2}.*

Definition 117.3. Let G be a group. We say that G is of *bounded period* if there exists a positive integer N such that $x^N = e$ for all elements x in G.

Theorem 117.4. (Burnside's Theorem on Linear Groups of Bounded Period) *Let F be a field of characteristic zero and G a subgroup of $\mathrm{GL}_n(F)$ (i.e., there exists a faithful representation of G of degree n). Suppose that G has bounded period N. Then G is a finite group. In fact, $|G| \leq N^{n^3}$.*

Proof. We may assume that F is algebraically closed, the inclusion $\iota : G \subset \mathrm{GL}_n(F)$ affords the $F[G]$-module $V = F^n$, and $\iota = \big(a_{ij}\big)$ with the a_{ij} the coordinate functions of ι. Let $\chi = \chi_V$.

Case 1. V is an irreducible $F[G]$-module.

By Corollary 117.2, there exists $\{x_1, \dots, x_{n^2}\} \subset G$, an F-basis for $\mathbb{M}_n(F)$ with $v_k = \big(a_{ij}(x_k)\big)$, $k = 1, \dots n^2$, (viewed in F^{n^2}) giving a basis $\mathcal{B} = \{v_1, \dots, v_{n^2}\}$ for F^{n^2}. As ι is a group homomorphism.

$$(*) \qquad \chi(x_k x) = \sum_{i=1}^{n} a_{ii}(x_k x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(x_k) a_{ji}(x), \quad k = 1, \dots, n^2$$

are n^2-linear equations in the unknowns $a_{ji}(x)$. As every $\chi(x)$, $x \in G$, is a sum of n roots of unity in μ_N, we have

$$s := \big|\{\chi(x) \mid x \in G\}\big| \leq N^n < \infty.$$

Since $\mathcal{B}$ is linearly independent in F^{n^2}, the matrix of coefficients of the system (*) is invertible, hence by Cramer's Rule each system (*) has a unique solution. It follows that

$$|G| = \big|\{a_{ij}(x) \mid x \in G\}\big| \leq s^{n^2} \leq N^{n^3}.$$

Case 2. V is not an irreducible $F[G]$-module.

As V is reducible, $n > 1$. Since G cannot act transitively on a basis for V, there exist representations $\varphi_i : G \to GL_{n_i}(F)$ with $1 = 1, 2$ each of degree less than n, satisfying

$$\iota(x) = \begin{pmatrix} \varphi_1(x) & * \\ 0 & \varphi_2(x) \end{pmatrix} = \begin{pmatrix} \varphi_1(x) & U(x) \\ 0 & \varphi_2(x) \end{pmatrix}$$

with $U : G \to F^{n_1 n_2}$. Set $G_i = \{\varphi_i(x) \mid x \in G\}$, $i = 1, 2$. Then G_i is a group of bounded period N of degree $n_i < n$, so finite with $|G_i| \leq N^{n_i^3}$ for $i = 1, 2$ by induction. Set $H_i = \ker \varphi_i$, then $G_i \cong G/H_i$ and $[G : H_i] < \infty$ for $i = 1, 2$. By Poincaré's Lemma $\big($Exercise 10.16(7)$\big)$, we have $[G : H_1 \cap H_2] < \infty$. Let $x \in H_1 \cap H_2$, then

$$\iota(x) = \begin{pmatrix} I & U(x) \\ 0 & I \end{pmatrix} \quad \textit{and} \quad 0 = \iota(x)^N = \quad \begin{pmatrix} I & NU(x) \\ 0 & I \end{pmatrix}.$$

Consequently, $NU(x) = 0$ and $U(x) = 0$ as F is a field of characteristic zero. Therefore, $H_1 \cap H_2 = 1$, hence G is a finite group and the representation $\psi : G \to G_1 \times G_2$ given by $x \mapsto \big(\varphi_1(x), \varphi_2(x)\big)$ is a monomorphism. It follows that

$$|G| \leq |G_1||G_2| \leq N^{n_1} N^{n_2} \leq N^{n^3}. \qquad \square$$

Remark 117.5. In the above proof we did not needed that F be a field of characteristic zero in Case 1; and in Case 2, we only needed that $\operatorname{char}(F) \nmid N$. In particular, the lemma holds if V is irreducible or if V is reducible and $\operatorname{char}(F) \nmid N$.

Using the proof above, we obtain another theorem of Burnside.

Theorem 117.6. (Burnside) *Let F be an arbitrary field and G a subgroup of $\mathrm{GL}_n(F)$. Then G is finite if and only if G has finitely many conjugacy classes.*

Proof. Certainly if G is finite, then it has finitely many conjugacy classes, so we need only show the converse. We use the notation as in the proof of the previous theorem. So ι is the inclusion map. If the character χ associated to ι is irreducible then the proof of Case 1 for Theorem 117.4 still works as $\{\chi(x) \mid x \in G\}$ is a finite set, which implies that G is a finite group. So we may assume that ι is not irreducible. As in the notation of Case 2 in the previous proof, we may assume that

$$\iota(x) = \begin{pmatrix} \varphi_1(x) & U(x) \\ 0 & \varphi_2(x) \end{pmatrix}$$

where $G_i = \{\varphi_i(x) \mid x \in G\}$, a representation of degree $n_i < n$, $\varphi : G_i \to \mathrm{GL}_{n_i}(F)$ and $U : G \to F^{n_1 n_2}$, $i = 1, 2$. Let $\psi : G \to G_1 \times G_2$ be given by $x \mapsto \big(\varphi_1(x), \varphi_2(x)\big)$ as before, and set $H = \ker \psi \triangleleft G$. Then $H \subset \ker \varphi_1 \cap \ker \varphi_2$. Suppose that $\left(\begin{smallmatrix} I & U(h) \\ 0 & I \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} I & U(h') \\ 0 & I \end{smallmatrix}\right)$ lie in H, then

$$\begin{aligned} \iota(x) = \begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix} \begin{pmatrix} I & U(h') \\ 0 & I \end{pmatrix} &= \begin{pmatrix} I & U(h + h')) \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & U(h') \\ 0 & I \end{pmatrix} \begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix}. \end{aligned}$$

In particular, H is abelian and $H \subset Z_G\Big(\left(\begin{smallmatrix} I & U(h) \\ 0 & I\end{smallmatrix}\right)\Big)$. It follows that the conjugacy class $\mathcal{C}\Big(\begin{pmatrix} I & U(h) \\ 0 & I\end{pmatrix}\Big)$ satisfies

$$\Big|\mathcal{C}\Big(\begin{pmatrix} I & U(h) \\ 0 & I\end{pmatrix}\Big)\Big| = \Big[G : Z_G\Big(\begin{pmatrix} I & U(h) \\ 0 & I\end{pmatrix}\Big)\Big] \leq [G : H].$$

As G has finitely many conjugacy classes, so do G_1 and G_2, hence both are finite by induction. Therefore, $[G : H] \leq |G_1|\,|G_2| < \infty$. As G has finitely many conjugacy classes and H is an abelian subgroup of G, we have H is also finite. It follows that G is a finite group. □

As mentioned in the introduction to this section, $\bigtimes_{i=1}^{\infty}\mathbb{Z}/p\mathbb{Z}$ is an infinite group of bounded period p, so certainly not finitely generated. Burnside conjectured in 1902 that any finitely generated torsion group G was finite. This is called the Burnside Conjecture. [Torsion groups are also called *periodic groups*.] We next improve the Burnside's theorems in the linear group case.

We need two lemmas.

Lemma 117.7. *Let G be a finitely generated torsion group. Suppose that G contains an abelian subgroup H of finite index. Then G is a finite group.*

Proof. Let $G = \bigvee_{1=1}^{m} g_i H$ be a coset decomposition. As G is finitely generated, there exists a finite subset

$$\mathcal{S} = \{g_1, \dots, g_m, g_{m+1}, \dots, g_n\} \subset G$$

generating G and closed under taking inverses. For each ordered pair (i, j), $1 \leq i, j \leq n$, there exists an integer $r = r(i, j)$, $1 \leq r \leq m$, satisfying

$$(*) \qquad g_i g_j = g_r h_{ij},\ \ h_{ij} \in H.$$

Set

$$H_0 = \langle h_{ij} \mid h_{ij} \ \text{ occurring as in } (*)\rangle.$$

Therefore, H_0 is a finitely generated group. As H_0 is a subgroup of the abelian torsion subgroup H, it is finite. Suppose that (i, j, r) is as in (*) and let $1 \leq s \leq n$. Then for each $m = 1, \dots, m$, there exists a positive integer $v = v(r, s)$ satisfying

$$g_s g_i g_j = g_s g_r h_{ij} = g_v h_{sr} h_{ij} \ \text{ lies in the coset } \ g_v H_0.$$

It follows by induction that the group

$$\langle g_1, \dots, \dots, g_n\rangle = \{\text{words in the } g_i\}$$

lies in $\bigcup_{i=1}^{m} g_i H_0$, hence $G \subset \bigcup_{i=1}^{m} g_i H_0$, so is finite. □

You should compare this proof with the proof of Theorem 16.3. We can use the lemma in the linear group case to see

Lemma 117.8. *Let F be a field and G a finitely generated torsion subgroup of $\mathrm{GL}_n(F)$. Then G has bounded period. In particular, if $\mathrm{char}(F) = 0$, then G is finite.*

Proof. Let Δ be the prime subfield of F. We may assume that $\Delta = \mathbb{Q}$ or $\Delta = \mathbb{Z}/p\mathbb{Z}$, depending on the characteristic of F. Since G is finitely generated, we may assume that F/Δ is a finitely generated field extension. In particular, there exists an intermediate field $F/F_0/\Delta$ with F_0/Δ a finitely generated purely transcendental field extension and F/F_0 a finite field extension. Let $r = [F : F_0]$. The subgroup G of $\mathrm{GL}_n(F)$ acts faithfully on F^n, so viewing F^n as F_0^{nr}, we have G acts faithfully on the finite dimensional F_0-vector space F_0^{nr}. In particular, we may assume that $F = F_0$ is a finitely generated purely transcendental extension.

If $x \in G \subset \mathrm{GL}_n(F)$, let q_x be the minimal polynomial of x in $F[t]$. In particular, $\deg q_x \leq \deg f_x = n$, where f_x is the characteristic polynomial of x in $\mathbb{M}_n(F)$.

Case 1. $\Delta = \mathbb{Q}$.

As x in G has finite order, the roots of q_x are roots of unity, so the coefficients of q_x lie in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap F$, where $\widetilde{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$ and $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is the integral closure of $\mathbb{Z}$ in $\widetilde{\mathbb{Q}}$ by Remarks 79.14. Since $F/\mathbb{Q}$ is purely transcendental, $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap F = \mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap \mathbb{Q} = \mathbb{Z}$ as $\mathbb{Z}$ is an integrally closed domain. It follows that q_x lies in $\mathbb{Z}[t]$ (as each irreducible factor of q_x does) by Proposition 80.4. The coefficients of q_x are elementary symmetric functions in the roots of q_x and these roots are roots of unity, hence the coefficients for each q_x must have bounded absolute value. So there can only be finitely many q_x with roots in a finite set of roots of unity.

Let ω be a root of q_x. Then $m_{\mathbb{Q}}(\omega) \mid q_x \mid f_x$, where $m_{\mathbb{Q}}(\omega)$ is the minimal polynomial of ω in $\mathbb{C}$. It follows that $\deg m_{\mathbb{Q}}(\omega) \leq n$ by the Cayley-Hamilton Theorem. As a primitive mth root of unity ε in $\mathbb{Q}$ satisfies $\deg m_{\mathbb{Q}}(\varepsilon) = \phi(m)$ and $\phi(m) \to \infty$ as $m \to \infty$, there can only be finitely many ω that are roots of the q_x, all uniformly bounded, i.e., $\{q_x \mid x \in G\}$ is a finite set. Therefore, there exists a positive integer N such that $x^N = 1$ for all x in G. In particular, the subgroup G of $GL_n(F)$ has bounded period. By Burnside's Theorem 117.4, G is a finite group.

Case 2. $\Delta = \mathbb{Z}/p\mathbb{Z}$ with $p > 0$.

As in Case 1, we have $q_x \in (\mathbb{Z}/p\mathbb{Z})[t]$ satisfies $\deg q_x \leq n$, so there exist finitely many such q_x, $x \in G$. The result follows by an analogous argument as in Case 1. □

Theorem 117.9. (Schur) *Let F be a field and G a finitely generated torsion subgroup of $\mathrm{GL}_n(F)$. Then G is a finite group.*

Proof. By Lemma 117.8, G has bounded period, so by Lemma 117.7, it suffices to show that G contains an abelian subgroup of finite index. By Remark 117.5, the only part of the proof of Burnside's Theorem on Linear Groups of Bounded Period 117.4 in the characteristic zero case that must be modified is the case that the inclusion $\iota : G \to GL_n(F)$ is not irreducible. But the proof of Burnside's Theorem 117.6 produces an abelian subgroup of finite index. □

In 1964, Golod-Shafarevich produced an infinite group on three generators in which every element has order a power of a prime, so the Burnside Conjecture is false. A modification of this conjecture was made to the conjecture that any finitely generated torsion group of bounded period is finite. Adrian-Novikov proved this too was false in 1968 (although Novikov claimed to have proven it in 1959). It was modified again to the so-called Restricted Burnside Problem which said: There exist only finitely many m-generator groups G of bounded period n. In particular, there exists an integer $N = N(m, n)$ such that $|G| < N$ for all such G. This was proven by Zelmanov in 1994 for which he won a Fields Medal.

Exercise 117.10. Complete the proof of Case 2 of Theorem 117.9.

Part 8

Homological Algebra and Category Theory

CHAPTER XX

Universal Properties and Multilinear Algebra

Given an algebraic object, it is usually very difficult to define a homomorphism from that object to another. Even if we have a well-defined set map, to show that it preserves the structure, i.e., is a homomorphism, we have to show that it preserves all the relations of that object under the map. We saw that free modules were those modules defined by the universal property that a homomorphism from a free module on a basis $\mathcal{B}$ was completely determined by where the basis $\mathcal{B}$ went. This meant that free modules had no nontrivial relations. Many structures in algebra can be defined by such universal properties, we shall call such definitions *categorical definitions*. For example, we also showed that the quotient field of a domain was defined by one (cf. Theorem 27.14). An algebraic object A is defined by a universal property (P) if it satisfies property (P) together with a map (or maps) and if C satisfies property (P), then there is a unique homomorphism(s) $A \to C$ (or $C \to A$) together with a commutative diagram (or commutative diagrams) respecting the given map(s). For example, F is a free R-module on a set $\mathcal{B}$ if there is a fixed set map $i : \mathcal{B} \to F$ and if M is an R-module and $j : \mathcal{B} \to M$ set map, then there exists a unique R-homomorphism $f : F \to M$ satisfying

$$\begin{tikzcd} \mathfrak{B} \arrow[r, "i"] \arrow[dr, "j"'] & F \arrow[d, "f"] \\ & M \end{tikzcd}$$

commutes. This also means that if F' is another free R-module on $\mathcal{B}$ with $i' : \mathcal{B} \to F'$ the given map, then there exists a unique R-isomorphism $f : F \to F'$. Thus the module F together with the set map i satisfies: there is a canonical bijection

$$\mathrm{Hom}_{\mathrm{sets}}(\mathcal{B}, N) \to \mathrm{Hom}_R(F, N)$$

relative to $i : \mathcal{B} \to F$, where the left hand side are set maps and canonical means the the map given uniquely by the commutative diagram.

In this chapter, we shall investigate other universal properties. Many of the details will be left to the reader. As an application we shall define the determinant of an R-homomorphism when R is a commutative ring and satisfying those properties of determinants of matrices that you know.

118. Some Universal Properties of Modules

Let R be a ring. Given an R-homomorphism $f : M \to N$ of R-modules, we know that we should determine its kernel and image. We also saw that cokernels existed. The kernel was a submodule of M on which f vanished. Of course, f can vanish on other submodules. Does $\ker f$ satisfy a universal property? The answer is yes. Formally we may define the kernel of an R-homomorphism as follows:

Definition 118.1. Let $f : A \to B$ be an R-homomorphism of R-modules. A *kernel* of f is an R-module K together with a R-homomorphism $i : K \to A$ satisfying

$$\begin{tikzcd} K \arrow[d, "i"'] \arrow[dr, "0"] & \\ A \arrow[r, "f"'] & B \end{tikzcd}$$

commutes (where 0 is the zero map), and satisfies the following universal property: If

$$\begin{tikzcd} M \arrow[d, "g"'] \arrow[dr, "0"] & \\ A \arrow[r, "f"'] & B \end{tikzcd}$$

is a commutative diagram of R-modules and R-homomorphisms, then there exists a unique R-homomorphism $h : M \to K$ satisfying

$$\begin{tikzcd} K \arrow[r, "i"] & A \arrow[r, "f"] & B \\ & M \arrow[ul, "h"] \arrow[u, "g"'] \arrow[ur, "0"'] & \end{tikzcd}$$

commutes. Moreover, if the kernel exists, it is unique up to a unique isomorphism.

Our original definition of kernel satisfies the above with $K = \ker f$ and i the inclusion map. We have just made a choice of a representation of a kernel object. Since it is unique relative to the inclusion map, there can be no other one. Note also that f is a monomorphism if and only if the kernel is the zero module (of which there is only one).

We can define a cokernel of an R-homomorphism $f : A \to B$ by reversing arrows (called *duality*), i.e.,

Definition 118.2. Let $f : A \to B$ be an R-homomorphism of R-modules. A *cokernel* of f is an R-module C together with a R-homomorphism $j : B \to C$ satisfying

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "0"'] & B \arrow[d, "j"] \\ & C \end{tikzcd}$$

commutes, and satisfying the following universal property: If

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "0"'] & B \arrow[d, "g"] \\ & N \end{tikzcd}$$

is a commutative diagram of R-modules and R-homomorphisms, then there exists a unique R-homomorphism $h : C \to N$ satisfying

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "0"'] & B \arrow[r, "j"] \arrow[d, "g"] & C \arrow[dl, "h"] \\ & N & \end{tikzcd}$$

commutes. Moreover, if the cokernel exists, it is unique up to a unique isomorphism.

Then our definition of cokernel is the unique quotient object $B/\operatorname{im} f$ relative to the canonical epimorphism. One problem here is that we want to define the image by a universal property, and it will depend on the definition of cokernel. Note that it is automatic that the map to a cokernel is surjective, so isomorphic to a quotient of B by some submodule. We define the image of an R-homomorphism as follows:

Definition 118.3. Let $f : A \to B$ be an R-homomorphism of R-modules. An *image* of f is kernel of $j : B \to C$, where C is a cokernel.

Remarks 118.4. By choice of nomenclature, a cokernel of a kernel of an R-homomorphism $f : A \to B$, should be called a *coimage* of $f : A \to B$. Therefore, $A/\ker f$ is the coimage of f. By the First Isomorphism Theorem, the quotient $A/\ker f$ is (canonically) isomorphic to the image of f, i.e., we can view the First Isomorphism Theorem as saying the coimage of f and the image of f are canonically isomorphic. These notions are important in category theory when they hold.

When we gave examples of modules, two that arose were the direct product and direct sum of modules. We now define these by categorical definitions.

Definition 118.5. Let A and B be R-modules. Then a *product* of A and B is an R-module P together with R-homomorphisms

$$\begin{tikzcd} P \arrow[r, "\pi_A"] \arrow[d, "\pi_B"'] & A \\ B & \end{tikzcd}$$

satisfying the following universal property: If

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "g"'] & A \\ B & \end{tikzcd}$$

are R-homomorphisms, then there exists a unique R-homomorphism $h : X \to P$ such that

$$
\begin{tikzcd}
X \arrow[drr, "f"] \arrow[dr, "h"] \arrow[ddr, "g"'] & & \\
& P \arrow[r, "\pi_A"'] \arrow[d, "\pi_B"] & A \\
& B &
\end{tikzcd}
$$

commutes. Moreover, if the product exists, it is unique up to a unique isomorphism and denoted by $A \prod B$.

Of course, this product is the external direct product previously defined via the projection maps.

The *coproduct* Y of two R-modules A and B is the dual notion of product defined by reversing arrows, i.e., it is a diagram

$$
\begin{tikzcd}
A \arrow[r, "\iota_A"] & Y \\
& B \arrow[u, "\iota_B"']
\end{tikzcd}
$$

satisfying the obvious universal property. It is unique up to a unique isomorphism and denoted by $A \coprod B$. Of course, this coproduct is the external direct sum previously defined via the inclusion maps of coordinates.

One defines a product and a coproduct of an arbitrary collection $\{M_i\}_I$ of R-modules in the obvious way and denotes it by $\prod_I M_i$ and $\coprod_I M_i$ respectively. For all j, they come equipped with both epimorphisms $\pi_j : \prod_I M_i \to M_j$ and monomorphisms $\iota_j : M_j \to \coprod_I M_i$ respectively which we identify as the ones that previously arose. If I is finite these are isomorphic and are identified as before.

As an example of how to use universal properties, we show:

Lemma 118.6. *An R-module A is a coproduct of the R-submodules $A_1, \dots, A_n$ if and only if there exist R-homomorphisms $\iota_j : A_j \to A$ and $\pi_j : A \to A_j$ for $j, k = 1, \dots, n$ satisfying $\pi_k \iota_j = \delta_{kj} 1_{A_j}$ and $\sum_j \iota_j \pi_j = 1_A$.*

PROOF. ($\Rightarrow$): The definition of coproduct gives rise to the ι_j's. For a fixed k and and each j, we have a diagram

$$
\begin{tikzcd}
A_j \arrow[r, "\delta_{kj} 1_{A_j}"] \arrow[d, "\iota_j"'] & A_k\ . \\
A &
\end{tikzcd}
$$

By the universal property of coproduct, there exists a unique R-homomorphism $\pi_k : A \to A_k$ such that

$$\begin{tikzcd} A_j \arrow[r, "\delta_{kj}1_{A_j}"] \arrow[d, "\iota_j"'] & A_k \\ A \arrow[ur, "\pi_k"'] & \end{tikzcd}$$

commutes. By distributivity, $(\iota_1\pi_1 + \cdots + \iota_n\pi_n)\iota_j = \iota_j$. Then

$$\begin{tikzcd} A_j \arrow[r, "\iota_j"] \arrow[dr, "\iota_j"'] & A \arrow[d, "1_A"] \\ & A \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} A_j \arrow[r, "\iota_j"] \arrow[dr, "\iota_j"'] & A \arrow[d, "\iota_1\pi_1+\cdots+\iota_n\pi_n"] \\ & A \end{tikzcd}$$

both commute for each j. Uniqueness gives $1_A = \iota_1\pi_1 + \cdots + \iota_n\pi_n$.
($\Leftarrow$): If $f_j : A_j \to B$ is an R-homomorphisms for each j, then defining $f : A \to B$ by $f = \sum_i f_i\pi_i$ works. $\square$

Suppose that we have R-homomorphisms of R-modules

$$A \xrightarrow{f} B \xrightarrow{h} D \quad \text{and} \quad A \xrightarrow{g} C \xrightarrow{j} D,$$

then by the universal properties of product and coproduct, there exist unique R-homomorphisms

$$A \xrightarrow{(f,g)} B \coprod C \xrightarrow{h\coprod j} D$$

given by

$$\begin{aligned} (f,g) &:= \iota_B f + \iota_C g \\ h\textstyle\coprod j &:= h\pi_B + j\pi_C, \end{aligned}$$

i.e., as we are identifying $A \prod B$ and $A \coprod B$, i.e., we have a diagram

$$\begin{tikzcd} & B \arrow[d, shift right, "\iota_B"'] \arrow[dr, "h"] & \\ A \arrow[ur, "f"] \arrow[r, "{(f,g)}"] \arrow[dr, "g"'] & B\coprod C \arrow[u, shift right, "\pi_B"'] \arrow[r, "h\coprod j"] \arrow[d, shift left, "\pi_C"] & D. \\ & C \arrow[u, shift left, "\iota_C"] \arrow[ur, "j"'] & \end{tikzcd}$$

In particular, we have

$$(h\textstyle\coprod g)(f,g) = hf + jg.$$

An interesting example of this is when $A = B = C$ and $f = g = 1_A$. Set $\Delta = (1_A, 1_A) : A \to A \coprod A$, called the *diagonal map*. Then the above shows

$$\begin{tikzcd} A \arrow[rr, "h+j"] \arrow[dr, "\Delta"'] & & D \\ & A \coprod A \arrow[ur, "h \coprod j"'] & \end{tikzcd}$$

commutes. Note that this commutative diagram categorically determines addition in $\operatorname{Hom}_R(A, D)$.

Exercises 118.7.

1. Define a free group, i.e., a group satisfying the obvious definition. Do you know of any examples of such an item.
2. Define the product and coproduct of a collection of R-modules $\{M_i \mid i \in I\}$ and show that these are the external direct product and external direct sum of R-modules.
3. Define the product and coproduct of two groups. Can you identify them?
4. Let $f : M \to N$ be an R-homomorphism of R-modules. Show that there exists a natural R-homomorphism $\operatorname{coim} f \to \operatorname{im} f$ arising from universal properties without using the First Isomorphism Theorem. Then show that it is an R-isomorphism.
5. Let R be a commutative ring, $S \subset R$ a multiplicative set, and M an R-module. Let $S^{-1}M := \{\frac{m}{s} \mid m \in M,\ s \in S\}$. Define an addition and R-action on $S^{-1}M$ making it an R-module and an $S^{-1}R$-module. This is called the *localization of M at S*. Determine the universal property that such a module satisfies.

119. Tensor Products

Let V be a vector space over F. If v and w are vectors in V, in general, there is no way to define a product of v and w in V. We want to rectify this, by creating a vector space in which a 'product' of v and w makes sense, called the tensor product. We shall do this in generality based upon a universal property. For vector spaces, as they are free modules, what we get is what one desires, but in general, unexpected results can occur.

Let $M_1, \dots, M_n$, and N be R-modules and $M_1 \times \cdots \times M_n$ the cartesian product of the M_i. A map

$$f : M_1 \times \cdots \times M_n \to N$$

is called *(R-) n-linear* or (*R*-) *n-multilinear* if f is linear in each variable, i.e.,

$$f(x_1, \dots, rx_i + x_i', \dots, x_n) = rf(x_1, \dots, x_i, \dots, x_n) + f(x_1, \dots, x_i', \dots, x_n)$$

for all $x_j \in M_j$, $x_{i'} \in M_i$, and $r \in R$. If $n = 2$ an n-linear form is called an *R-bilinear form*. If R is a commutative ring we would like to convert bilinear maps into R-homomorphisms. This is done as follows:

Definition 119.1. Let R be a commutative ring and M, N two R-modules. An R-module T is called a *tensor product* of M and N if there exists an R-bilinear map $\iota : M \times N \to T$

satisfying the following universal property: If A is an R-module and $j : M \times N \to A$ an R-bilinear map, then there exists a unique R-homomorphism $f : T \to A$ such that

$$\begin{tikzcd} M \times N \arrow[r, "\iota"] \arrow[dr, "j"'] & T \arrow[d, "f"] \\ & A \end{tikzcd}$$

commutes. If a tensor product $\iota : M \times N \to T$ exists, it is unique up to a unique isomorphism and T is denoted by $M \otimes_R N$.

Lemma 119.2. *Let R be a commutative ring and M, N two R-modules. Then a tensor product of M and N exists.*

Proof. Let P be the free R-module on basis $\mathcal{B} = M \times N$. Let W be the submodule of P generated by the following:

$$\begin{gathered} (m + m', n) - (m, n) - (m', n) \\ (m, n + n') - (m, n) - (m, n') \\ (rm, n) - r(m, n) \\ (m, rn) - r(m, n) \end{gathered}$$

for all m, m' in M, n, n' in N, and r in R. Let $^- : P \to P/W$ by $x \mapsto \overline{x} = x + W$ be the be the canonical R-epimorphism. Then the composition ι of the maps $M \times N \xrightarrow{inc} P \xrightarrow{-} P/W$ is R-bilinear. Set $T = P/W$ and $M \times N \to T$. We denote $\iota(m, n)$ by $m \otimes n$. Then every element in T is a finite sum

$$\sum_{\text{finite}} r_i(m_i \otimes n_i) = \sum_{\text{finite}} (r_i m_i) \otimes n_i = \sum_{\text{finite}} m_i \otimes (r_i n_i),$$

for appropriate elements m_i in M, n_i in N, and r_i in R, as $r(m \otimes n) = (rm) \otimes n = m \otimes (rn)$ for all m in M, n in N, and r in R. Suppose that $j : M \times N \to A$ is R-bilinear. Define $f : T \to A$ as follows: By the universal property of freeness on basis $M \times N$, there exists a unique R-homomorphism $g : P \to A$ satisfying

$$\begin{tikzcd} M \times N \arrow[r, "\iota"] \arrow[dr, "j"'] & P \arrow[d, "g"] \\ & A \end{tikzcd}$$

commutes. Now, with $^- : P \to P/W$ the canonical R-epimorphism,

$$\begin{tikzcd} M \times N \arrow[r] \arrow[dr, "j"'] & P \arrow[r, "-"] \arrow[d, "g"] & P/W \\ & A & \end{tikzcd}$$

and clearly, $g|_{\ker \bar{\ }} = 0$, i.e., $g|_W = 0$ as j is R-bilinear. Hence

$$\begin{tikzcd}
0 \arrow[r] & W \arrow[r, "\mathrm{inc}"] \arrow[dr, "0"'] & P \arrow[r, "\bar{\ }"] \arrow[d, "g"] & P/W \\
& & A &
\end{tikzcd}$$

is exact. But P/W is the cokernel of the inclusion map $\mathrm{inc} : W \to P$, so there exists a unique R-homomorphism $\overline{g} : P/W \to A$ satisfying

$$\begin{tikzcd}
P \arrow[r, "\bar{\ }"] \arrow[d, "g"'] & P/W \arrow[dl, "\overline{g}"] \\
A &
\end{tikzcd}$$

commutes. As

$$\begin{tikzcd}
M \times N \arrow[r, "\iota"] \arrow[dr, "j"'] & P/W \arrow[d, "\overline{g}"] \\
& A
\end{tikzcd}$$

commutes and $\overline{g}$ is unique, we have established existence.
[If you do not like using the universal property of cokernels, you can prove this directly, i.e., if A is a submodule of an R-module B and $f : B \to C$ is an R-homomorphism satisfying $A \subset \ker \varphi$, then there exists a unique R-homomorphism $\overline{\varphi} : B/A \to C$ satisfying

$$\begin{tikzcd}
B \arrow[r] \arrow[d, "\bar{\ }"'] & C. \\
B/A \arrow[ur] &
\end{tikzcd}$$

commutes.] We leave the proof of uniqueness to the reader. □

We adopt the notation above, i.e., elements of $M \otimes_R N$ are finite sums $\sum r_i(m_i \otimes n_i)$, etc.

Remark 119.3. Let R be a commutative ring.

1. If M, N, and A are R-modules, by the universal property of tensor products, the module structure of $M \otimes_R N \to A$ is determined by a R-bilinear map $M \times N \to A$ (hence on generators (m, n) in $M \times N$). In particular, an R-homomorphism $\varphi : M \otimes_R N \to A$, is completely determined by the $\varphi(m \otimes n)$.
2. Let M_i, N_i be R-modules, and $\varphi_i : M_i \to N_i$ an R-homomorphism for $i = 1, 2$. By the universal property of tensor products, there exists a unique R-homomorphism $\varphi_1 \otimes \varphi_2 : M_1 \otimes_R M_2 \to N_1 \otimes_R N_2$ induced by $m_1 \otimes m_2 \mapsto \varphi(m_1) \otimes \varphi(m_2)$ as $M_1 \times M_2 \to N_1 \otimes_R N_2$ given by $(m_1, m_2) \mapsto \varphi(m_1) \otimes \varphi(m_2)$ is R-bilinear.

If R is a commutative ring, then the lemma shows that there exists a bijection of sets

$$\{f : M \times N \to A \mid f \text{ an } R\text{-bilinear map}\}$$
$$\to \{f : M \otimes_R N \to A \mid f \text{ an } R\text{-homomorphism}\}$$

i.e., $\otimes_R$ converts R-bilinear maps to R-linear maps, i.e., R-homomorphisms, as desired.

Properties 119.4. Let R be a commutative ring and P, M, N R-modules.

1. $0 \otimes 0 = m \otimes 0 = 0 \otimes n$ is the zero of $M \otimes_R N$ for any m in M, any n in N.
2. The map $\iota_M : R \otimes_R M \to M$ induced by $r \otimes m \mapsto rm$ is a *canonical R-homomorphism*, where canonical means if $f : M \to N$ is an R-homomorphism, then we have commutative diagram
$$\begin{array}{ccc} R \otimes_R M & \xrightarrow{\iota_M} & M \\ {\scriptstyle 1_R \otimes \varphi}\downarrow & & \downarrow{\scriptstyle \varphi} \\ R \otimes_R N & \xrightarrow{\iota_N} & N. \end{array}$$
[Why is it well defined?] Check that $M \to R \otimes_R M$ given by $m \mapsto 1 \otimes m$ is the inverse.
3. Let M_i, $i \in I$, N be R-homomorphisms. Then there exists a canonical R-isomorphism
$$\coprod_I (M_i \otimes_R N) \to (\coprod_I M_i) \otimes_R N,$$
i.e, $\coprod_I$ and $\otimes_R$ commute. (Similarly, on the other side.):

   Since the map
$$(\coprod_I M_i) \times N \to \coprod_I (M_i \otimes_R N_i) \text{ given by } \{\{m_i\}_I, n\} \mapsto \{m_i \otimes n\}_I$$
is R-bilinear, the universal property of the tensor products induces a unique R-homomorphism $f : (\coprod_I M_i) \otimes_R N \to \coprod_I (M_i \otimes_R N_i)$. For each $j \in I$, we have an R-bilinear map $M_j \times N \to (\coprod_I M_i) \otimes_R N$ given by $(m, n) \mapsto \{\delta_{ij} m_i\} \otimes n$, inducing a unique R-homomorphism $M_j \otimes_R N \to (\coprod_I M_i) \otimes_R N$ by the universal property of tensor products. Hence, by the universal property of coproducts, there exists a unique R-homomorphism $g : \coprod_I (M_i \otimes_R N) \to (\coprod_I M_i) \otimes_R N$. Check that f and g are inverses of each other.
4. $M \otimes_R N$ is canonically isomorphic to $N \otimes_R M$ induced by the R-bilinear map $M \times N \to N \otimes_R M$ given by $(m, n) \mapsto n \otimes m$.
5. $(M \otimes_R N) \otimes_R P$ is canonically isomorphic to $M \otimes_R (N \otimes_R P)$
6. $(\coprod_I R) \otimes_R M \cong \coprod_I (R \otimes_R M) \cong \coprod_I M$ with all R-isomorphisms canonical.
7. Suppose that M is a free R-module on basis $\mathcal{B}$ and N is a free R-module on basis $\mathcal{C}$. Then $M \otimes_R N$ is a free R-module on basis $\{b \otimes c \mid b \in \mathcal{B}, c \in \mathcal{C}\}$. In particular,
$$\operatorname{rank}_R(M \otimes_R N) = \operatorname{rank}_R M \operatorname{rank}_R N :$$
As $R \cong Rx$ for all x in $\mathcal{B}$ or $\mathcal{C}$, we have $M \cong \coprod_{\mathcal{B}} R$ and $N \cong \coprod_{\mathcal{C}} R$. Therefore,
$$M \otimes_R N \cong (\coprod_{\mathcal{B}} R) \otimes_R (\coprod_{\mathcal{C}} R) \cong \coprod_{\mathcal{B}} \big(R \otimes_R (\coprod_{\mathcal{C}} R)\big) \cong \coprod_{\mathcal{B}} \coprod_{\mathcal{C}} R.$$
8. If M, N are not free R-modules, bad things can occur. For example, the abelian group $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} /3\mathbb{Z} = 0$. For if $[a]_n$ is the congruence class of the integer a modulo n, then $[1]_2 \otimes [1]_3 = [1]_2 \otimes 4[1]_3 = 4[1]_2 \otimes [1]_3 = 0$. Similarly, $[1]_2 \otimes [-1]_3 = 0$.

9. (Base Extension) Let $\varphi : R \to T$ be a ring homomorphism of commutative rings. We know that any T-module B becomes an R-module via the pullback, i.e., $rx := \varphi(r)x$ for all $r \in R$, $x \in B$, i.e., via φ. In particular, T becomes an R-module, hence so does $T \otimes_R M$. We can now make $T \otimes_R M$ into a T-module by defining
$$\alpha(\beta \otimes m) := (\alpha\beta) \otimes m,$$
for all α and β in T and m in M. We call the T-module $T \otimes_R M$ the *base extension* of the R-module M to a T-module (via φ).

Examples 119.5. Let K/F be a field extension. Then K is an F-vector space, say on basis $\mathcal{C}$. Let V be an F-vector space on basis $\mathcal{B}$. The $K \otimes_F V$ is an F-vector space on basis $\{c \otimes b \mid c \in \mathcal{C}, b \in \mathcal{B}\}$, so $\dim_F K \otimes_F V = \dim_F K \dim_F V$ and $K \otimes_F V$ is a K-vector space on basis $\{1 \otimes b \mid b \in \mathcal{B}\}$. For example, $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ is a four dimensional real vector space and a 2-dimensional complex vector space.

Question 119.6. Can you make $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ into a ring? If so, is it a field?, a domain?, commutative?

Proposition 119.7. *Let V and W be finite dimensional F-vector spaces, $V^* = \operatorname{Hom}_F(V, F)$, the dual space of V (or finitely generated R-free modules with R commutative). Then the natural map*
$$\varphi : V^* \otimes_F W \to \operatorname{Hom}_F(V, W)$$
induced by $f \otimes w \mapsto (v \mapsto f(v)w)$ is an F-linear isomorphism.

Proof. Although the map in the proposition is independent of any basis, we use bases to prove it. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be an F-basis for V and $\mathcal{B}^* = \{f_1, \dots, f_n\}$ the dual F-basis for V^*, i.e., $f_i(v_j) = \delta_{ij}$, so f_i is the coordinate function on v_i. Therefore, if v is an element of V, $v = \sum_i f_i(v)v_i$. Define
$$\psi : \operatorname{Hom}_F(V, W) \to V^* \otimes_F W \text{ by } T \mapsto \sum_i f_i \otimes Tv_i.$$

Check: φ and ψ are both F-linear.
Then we have
$$\varphi\psi T(v) = \varphi(\sum f_i \otimes Tv_i)(v) = \sum f_i(v)Tv_i = \sum(T(f_i(v)v_i) = T(v),$$
so $\varphi\psi = 1_{\operatorname{Hom}_F(V,W)}$. Hence φ is an epimorphism. As
$$\dim_F(V^* \otimes_F W) = \dim_F \operatorname{Hom}_F(V, W) < \infty,$$
φ is an F-isomorphism (as both are finite dimensional F-vector spaces). [If V and W are only assumed to be finitely generated free R-modules, one must show that $\psi\varphi = 1_{V^* \otimes_R W}$.] $\square$

Remarks 119.8. We look at a interesting special case of the proposition above. Let $V = W$ and $\mathcal{B}, \mathcal{B}^*$ as in the proposition. Then the proposition gives an F-linear transformation
$$\varphi : V^* \otimes_F V \to \operatorname{End}_F(V).$$
Moreover, the inverse map ψ satisfies $\psi(1_{\operatorname{End}_F(V)}) = \sum f_i \otimes v_i$. Let
$$\operatorname{End}_F(V) \xrightarrow{\psi} V^* \otimes_F V \xrightarrow{\rho} F,$$

where ρ is induced by $f \otimes v \mapsto f(v)$. It follows that for $T \in \operatorname{End}_F(V)$, we have

$$\rho\psi(T) = \rho(\sum f_i \otimes Tv_i) = \sum f_i(Tv_i).$$

In particular, if $Tv_i = \sum \alpha_{ji}v_j$, then

$$\begin{aligned}\rho\psi(T) &= \sum_i f_i(\sum_j \alpha_{ji}v_j) = \sum_{i,j} \alpha_{ji}f_i(v_j) \\ &= \sum_i \alpha_{ii} = \operatorname{trace}\,[T]_{\mathcal{B}}.\end{aligned}$$

Since φ is defined independently of $\mathcal{B}$ and ψ is its inverse, ψ is also independent of $\mathcal{B}$. therefore, $\rho\psi(T)$ is independent of $\mathcal{B}$, i.e., the *trace* of T defined by

$$\operatorname{trace} T := \rho\psi(T) = \operatorname{trace}\,[T]_{\mathcal{B}}$$

is independent of $\mathcal{B}$. Since $\psi(1_{\operatorname{End}_F(V)}) = \sum f_i \otimes v_i$, we have $\varphi(f_i \otimes v_i) = 1_{\operatorname{End}_F(V)}$. The element $\sum f_i \otimes v_i$ is called the *Casimir element*. It too is independent of $\mathcal{B}$ and $\mathcal{B}^*$

Exercises 119.9.

1. Prove that the tensor product of two modules over a commutative ring is unique up to a unique isomorphism.
2. Show $0 \otimes 0 = m \otimes 0 = 0 \otimes n$ is the zero of $M \otimes_R N$ for any m in M, any n in N.
3. Prove Property 119.4(5).
4. Let m and n be positive integer and d the greatest common divisor of m and n. Then $\mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$ (but not canonically).
5. In the proof of Proposition 119.7, show that $\psi\varphi = 1_{V^* \otimes_R W}$ if V and W are finitely generated free R-modules.
6. Let R be a commutative ring, M an R-module, and
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
a short exact sequence of R-modules. Show that
$$A \otimes_R M \xrightarrow{f \otimes 1_M} B \otimes_R M \xrightarrow{g \otimes 1_M} C \otimes_R M \to 0$$
and
$$M \otimes_R A \xrightarrow{1_M \otimes f} M \otimes_R B \xrightarrow{1_M \otimes g} M \otimes_R C \to 0$$
are exact.
7. Let R be a commutative ring, $S \subset R$ a multiplicative set, and M an R-module. Show that there exists a natural $S^{-1}R$-isomorphism $S^{-1}M \to S^{-1}R \otimes_R M$. (Cf. Exercise 118.7(5).) Show further that if
$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$
is a short exact sequence of R-modules, then so is
$$0 \to S^{-1}R \otimes_R M' \xrightarrow{1 \otimes f} S^{-1}R \otimes_R M \xrightarrow{1 \otimes g} S^{-1}R \otimes_R M'' \to 0.$$

8. Let R be a commutative ring and N an R-module. We say that N is a *flat* R-module if whenever
$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$
is exact, so is
$$0 \to M' \otimes_R N \xrightarrow{f \otimes 1_N} M \otimes_R N \xrightarrow{g \otimes 1_N} M'' \otimes_R N \to 0.$$
Show that if N is a projective R-module, then it is R-flat.

9. Let R be a commutative ring and V and W be finitely generated free R-modules with ordered R-bases $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ respectively. Let $\mathcal{D} = \{v_1 \otimes w_1, \dots, v_1 \otimes w_m, v_2 \otimes w_1, \ \dots, v_n \otimes w_m\}$, an ordered R-basis for $V \otimes_R W$. Let $T \in \operatorname{End}_R(V)$ and $S \in \operatorname{End}_R(W)$. Write $A = [T]_{\mathcal{B}}$ and $B = [S]_{\mathcal{C}}$. Show that the matrix representation $[T \circ S]_{\mathcal{D}}$ is the *Kronecker product*
$$\begin{pmatrix} A_{11}B & \cdots & A_{1m}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nm}B \end{pmatrix}$$
i.e.,
$$\big([T \circ S]_{\mathcal{D}}\big)_{n(i-1)+r, m(j-1)+s} = \big([T]_{\mathcal{B}}\big)_{ij}\big([S]_{\mathcal{C}}\big)_{rs}.$$

120. Tensor, Symmetric, and Exterior Algebras

In the last section, we defined the tensor product of modules over a commutative ring. In this section, we show how to construct algebras from given algebras and given modules.

First let us recall a ring A over a commutative ring R is called an *R-algebra* if there exists a ring homomorphism $\varphi_A : R \to Z(A)$ (where $Z(A)$ is the center of A) and an *R-algebra homomorphism* of R-algebras is a ring homomorphism $\psi : A \to B$ of R-algebras such that
$$\begin{tikzcd} A \arrow[rr, "\psi"] & & B \\ & R \arrow[ul, "\varphi_A"] \arrow[ur, "\varphi_B"'] & \end{tikzcd}$$
commutes. The algebra structure on A makes A into a left and right R-module. For example, every ring is a $\mathbb{Z}$-algebra. If R is a field F and A a nonzero F-algebra, we often identify F and $F1_A$, i.e., view φ_A as an inclusion as before. An R-algebra A is called a *graded R-algebra* if $A = \coprod_{i=0}^{\infty} A_i$ as an additive group with $\varphi_A(R) \subset A_0$ and $A_iA_j \subset A_{i+j}$ for all i, j. In particular, A_0 is a ring. Elements in A_i are called *homogeneous elements of degree i*. We write $\deg(a) = i$ if $a \in A_i$. If $B = \coprod_{i \geq 0} B_i$ is another graded R-algebra, an R-algebra homomorphism $\psi : A \to B$ is called *graded* if $\varphi(A_i) \subset B_i$ for all $i \geq 0$. We want to construct some graded R-algebras.

Let R be a commutative ring and S a set. Then the *free commutative R-algebra* on S is the R-algebra satisfying the appropriate universal property — what is it? It exists and is unique up to a unique isomorphism, i.e., if two R-algebras satisfy the universal property there is a unique R-algebra isomorphism between them. Indeed, $R[t_s]_S$, the polynomial ring on $|S|$-variables, works.

Construction 120.1. Let R be a commutative ring and A, B two R-algebras. If $(a, b) \in A \times B$, then

$$m_{a,b} : A \times B \to A \otimes_R B \text{ given by } (a', b') \mapsto aa' \otimes bb'$$

is R-bilinear, so induces a unique R-homomorphism

$$\mu_{a,b} : A \otimes_R B \to A \otimes_R B \text{ given by } a' \otimes b' \mapsto aa' \otimes bb'$$

and hence

$$(A \otimes_R B) \times (A \otimes_R B) \to A \otimes_R B \text{ given by } (a \otimes b, a' \otimes b') \mapsto aa' \otimes bb'$$

is a well defined map that makes $A \otimes_R B$ into a ring and into an R-algebra by $R \to A \otimes_R B$ by $r \mapsto r \otimes 1$, where $r \otimes 1 = r1_A \otimes 1_B = 1_A \otimes r1_B = 1 \otimes r = \varphi_A(r)1_A \otimes 1_B = 1_A \otimes \varphi_B(r)1_B$ for all r in R. We also call this algebra the *tensor product* of A and B.

Definition 120.2. Let R be a commutative ring and M, N be R-modules. For each positive integer n, let $T^n(M)$ be the R-module $\underbrace{M \otimes_R \cdots \otimes_R M}_{n}$. Then $T^n(M)$ is generated by $m_1 \otimes \cdots \otimes m_n$ with $m_i \in M$. Define $T^0(M) := R$. If $f : M \to N$ is an R-homomorphism, then there exists a unique R-homomorphism $T^n f : T^n(M) \to T^n(N)$ induced by the (R-)n-linear map $M_1 \times \cdots \times M_n \to T^n(N)$ given by $(m_1, \cdots, m_n) \mapsto f(m_1) \otimes \cdots \otimes f(m_n)$ (why?).

Example 120.3. Let n be a positive integer, V an m-dimensional F-vector space on basis $\mathcal{B} = \{v_1, \ldots, v_m\}$. Then $T^n(V)$ is an F-vector space basis $\{v_{i_1} \otimes \cdots \otimes v_{i_n} \mid 1 \leq i_j \leq m\}$ (repetitions are allowed), a vector space of dimension m^n. [The same is true if we have M a free R-module of on a basis $\mathcal{B}$ with R a commutative ring.] We also define $T^{-n}(V) := T^n(V^*)$, with V^* the dual space $\mathrm{Hom}_F(V, F)$.

Construction 120.4. Let R be a commutative ring and M an R-module. Associativity of the tensor products implies that

$$T^m(M) \times T^n(M) \to T^{m+n}(M) \text{ given by } (x, y) \mapsto x \otimes y$$

is R-bilinear, so induces a (non-commutative) ring structure on the R-module

$$T(M) := \coprod_{i=0}^{\infty} T^n(M)$$

called the *tensor algebra* of M. If $x, y \in T(M)$, we write xy for the product. Hence if $x \in T^m(M)$ and $y \in T^n(M)$, then $xy = x \otimes y$. Under this ring structure, we have

$$\begin{aligned}
0_{T(M)} &:= \coprod_{n=0}^{\infty} 0_{T^n(M)} \\
1_{T(M)} &:= 1_R \\
R &:= T^0(M), \text{ a subring of } T(M) \\
T(M) &\text{ is a } R\text{-module} \\
T(M) &\text{ is an } R\text{-algebra.}
\end{aligned}$$

The R-algebra $T(M)$ also satisfies the following universal property: If $\varphi : M \to A$ is a R-homomorphism of R-algebras, then there exists a unique R-algebra homomorphism $\Phi : T(M) \to A$ satisfying $\Phi|_M = \varphi$.

Let R be a commutative ring and M an R-module. As the tensor algebra $T(M)$ is a ring, it has ideals, hence quotient rings. We use this to construct other algebras. But $T(M)$ has an additional structure, it is a graded R-algebra by $T(M) = \coprod T^n(M)$ as the ring multiplication on $T(M)$ satisfies $T^m(M) \times T^n(M) \to T^{m+n}(M)$. We want to construct graded R-algebras from $T(M)$. The first is the *symmetric algebra*.

Construction 120.5. Let R be a commutative ring and M an R-module. For each positive integer n, let $\mathfrak{A}_n$ be the submodule of $T^n(M)$ generated by all elements $m_1 \otimes \cdots \otimes m_n - m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(n)}$ with $m_1, \dots, m_n$ in M and σ in S_n. Define $\mathfrak{A}_0 = 0$. Let

$$\begin{aligned} \mathfrak{A} &:= \coprod_{i \geq 0} \mathfrak{A}_n \\ S^n(M) &:= T^n(M)/\mathfrak{A}_n \\ S(M) &:= \coprod_{i \geq 0} S^n(M) = \coprod_{i \geq 0} T^n(M)/\mathfrak{A}_n. \end{aligned}$$

Check that $\mathfrak{A}$ is a ideal in $T(M)$ and $S(M) = T(M)/\mathfrak{A}$ is a graded R-algebra, i.e., the ring multiplication on the algebra $S(M)$ satisfies $S^m(M) \times S^n(M) \to S^{m+n}(M)$ for all $m, n \geq 0$. The R-algebra $S(M)$ is a commutative algebra called the *symmetric algebra* of M and we have ring homomorphisms $R \xrightarrow{\varphi_{T(M)}} T(M) \xrightarrow{-} S(M)$ with $^-$ the canonical epimorphism.

Example 120.6. Let V be an n-dimensional F-vector space. Then $S(V) \cong F[t_1, \dots, t_n]$ (An analogous statement holds if V is an R-free module of rank n with R commutative): Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be an F-basis for V and $^- : T^m(V) \to S^m(V)$ the canonical R-module epimorphism. Then

$$\{\overline{v_{i_1} \otimes \cdots \otimes v_{i_m}} \mid 1 \leq i_1 \leq \cdots \leq i_m \leq n\}$$

is an F-basis for $S^m(V)$. In particular,

$$\dim_F S^m(V) = \binom{m+n-1}{n-1} = \binom{m+n-1}{m},$$

the number of (*homogeneous*) monic monomials of total degree m in $t_1, \dots, t_n$. (Why?) In particular, $\dim_F S(V) = \infty$.

In the construction above, the ideal $\mathfrak{A} := \coprod_{n \geq 0} \mathfrak{A}_n$ satisfies

$$\mathfrak{A} \cap T^n(M) = \mathfrak{A}_n \text{ and } \mathfrak{A}_m \mathfrak{A}_n \subset \mathfrak{A}_{mn} \text{ for all } m, n \geq 0.$$

We say $\mathfrak{A} = \coprod_{n \geq 0} \mathfrak{A}_n$ is a *graded* or *homogeneous ideal* in the graded ring $T(M)$.

We want universal properties for $S^n(M)$ and $S(M)$. Call a map $f : \underbrace{X \times \cdots \times X}_{n} \to Y$ of R-modules *symmetric* if $f(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ for all $\sigma \in S_n$.

Let R be a commutative ring and M an R-module. Then we have a canonical n-linear map $f : \underbrace{M \times \cdots \times M}_{n} \to S^n(M)$ given by the composition $\underbrace{M \times \cdots \times M}_{n} \to T^n(M) \to T^n(M)/\mathfrak{A}_n$. Then $S^r(M)$ satisfies the following universal property:

Let $g : M \times \cdots \times M \to N$ be an n-linear and symmetric map of R-modules. Then there exists a unique R-homomorphism $\varphi : S^n(M) \to N$ such that

$$\begin{tikzcd} M \times \cdots \times M \arrow[r, "f"] \arrow[dr, "g"'] & S^n(M) \arrow[d, "\varphi"] \\ & N \end{tikzcd} \tag{120.7}$$

commutes and the commutative R-algebra $S(M)$ satisfies the following universal property:

If A is a commutative R-algebra and $\varphi : M \to A$ is an R-homomorphism, then there exists a unique R-algebra homomorphism

$$\Phi : S(M) \to A \text{ satisfying } \Phi|_M = \varphi. \tag{120.8}$$

We next construct another quotient of the tensor algebra, this time a non-commutative algebra.

Construction 120.9. Let R be a commutative ring and M an R-module. Define $\mathfrak{B}_n$ to be the submodule of $T^n(M)$ generated by all $m_1 \otimes \cdots \otimes m_n$ with $m_1, \ldots, m_n$ in M and $m_i = m_j$ for some $i \neq j$. Set $\bigwedge^n(M) := T^n(M)/\mathfrak{B}_n$. The map $f : \underbrace{M \times \cdots \times M}_{n} \to \bigwedge^n(M)$ defined by the composition $\underbrace{M \times \cdots \times M}_{n} \to T^n(M) \xrightarrow{-} T^n(M)/\mathfrak{B}_n$ is n-linear and *alternating*, i.e., satisfies $f(x_1, \ldots, x_n) = 0$ if there exists an $i \neq j$ such that $x_i = x_j$. It satisfies the following universal property:

If N is an R-module and $g : M \times \cdots \times M \to N$ is n-linear and alternating, then there exists a unique R-homomorphism $\varphi : \bigwedge^n(M) \to N$ such that

$$\begin{tikzcd} M \times \cdots \times M \arrow[r, "f"] \arrow[dr, "g"'] & \bigwedge^n(M) \arrow[d, "\varphi"] \\ & N \end{tikzcd}$$

commutes. This follows as there exists a unique R-homomorphism $\psi : T^n(M) \to N$ such that

$$\begin{tikzcd} M \times \cdots \times M \arrow[r, "f'"] \arrow[dr, "g"'] & T^n(M) \arrow[d, "\psi"] \\ & N \end{tikzcd}$$

commutes. Then

$$\begin{tikzcd} 0 \arrow[r] & \mathfrak{B}_n \arrow[r, "h"] \arrow[dr, "0"'] & T^n(M) \arrow[r, "-"] \arrow[d, "\psi"] & \bigwedge^n(M) \arrow[r] & 0 \\ & & N & & \end{tikzcd}$$

is exact and commutes, so φ exists and is unique by the universal property of the cokernel. We denote the image of $(m_1, \dots, m_n)$ in $\bigwedge^n(M)$ by $m_1 \wedge \cdots \wedge m_n$.

If $\varphi : M \to N$ is an R-homomorphism, then

$$\underbrace{M \times \cdots \times M}_{n} \to \bigwedge^n(N)$$

defined by $m_1, \dots, m_n \mapsto \varphi(m_1) \wedge \cdots \wedge \varphi(m_n)$ is n-linear and alternating, so induces a unique R-homomorphism $\wedge^n \varphi : \bigwedge^n(M) \to \bigwedge^n(N)$. Set $\bigwedge^0(M) := R$, $\bigwedge^1(M) = M$ and $\bigwedge(M) := \coprod_{n \geq 0} \bigwedge^n(M)$.

Check. $\mathfrak{B} := \coprod_{n \geq 0} \mathfrak{B}_n$ is a graded ideal in $T(M)$ by using

$$T^m(M) \times \mathfrak{B}_n \text{ and } \mathfrak{B}_m \times T^n(M) \text{ lie in } \mathfrak{B}_{m+n},$$

so induces an R-bilinear map $\bigwedge^n(M) \times \bigwedge^m(M) \to \bigwedge^{m+n}(M)$ and (as we have associativity) an R-homomorphism

$$\bigwedge^m(M) \otimes_R \bigwedge^n(M) \to \bigwedge^{m+n}(M).$$

This defines a graded R-algebra structure $\bigwedge(M)$ with multiplication $\bigwedge(M) \times \bigwedge(M) \to \bigwedge(M)$ defined by

$$(x_1 \wedge \cdots \wedge x_m) \cdot (y_1 \wedge \cdots \wedge y_n) := x_1 \wedge \cdots \wedge x_m \wedge y_1 \wedge \cdots \wedge y_n.$$

[Why does it suffice to define multiplication on generators?] This R-algebra is called the *exterior algebra* of M.

Remark 120.10. Let R be a commutative ring and M an R-module.

1. $\bigwedge(M)$ can be defined to be the R-algebra generated by M with defining relations

   $$m \wedge m = 0 \text{ for all } m \in M:$$

   If m, n are elements of M, then

   $$mn + nm = (n+m)^2 - n^2 - m^2 \text{ in } T(M),$$

   so

   $$m \wedge n = -n \wedge m \text{ in } \bigwedge(M)$$

   and it follows that the graded ideal $\mathfrak{B}$ is generated in the algebra $T(M)$ by m^2, $m \in M$.
2. Using (1), one checks $\bigwedge(M)$ satisfies the following universal property: If A is an R-algebra and $\varphi : M \to A$ an R-homomorphism satisfying $\varphi(m)^2 = 0$ for all $m \in M$, then there exists a unique R-algebra homomorphism $\Phi : \bigwedge(M) \to A$ satisfying $\Phi|_M = \varphi$. In particular, any R-homomorphism $\varphi : M \to N$ extends to a unique graded R-algebra homomorphism $\wedge\varphi : \bigwedge(M) \to \bigwedge(N)$, i.e., $\wedge\varphi\big(\bigwedge^n(M)\big) \subset \bigwedge^n(N)$. For example, $\varphi(1_M) = 1_{\bigwedge(M)}$.

3. If $\varphi : M \to N$ and $\psi : N \to P$ are R-homomorphisms, then $\wedge(\psi\varphi) = \wedge\psi \circ \wedge\varphi$.

In the next section, we shall show how the exterior algebra on a finitely generated free R-module V leads to the determinant of a R-homomorphism $f : V \to V$ when R is a commutative ring. We end this section, with an example of an exterior algebra that is useful in commutative algebra. We shall leave details and proofs as exercises.

Recall that a sequence

$$(*) \qquad \cdots \xrightarrow{d_{i+1}} M_i \xrightarrow{d_i} M_{i-1} \xrightarrow{d_{i-1}} \cdots$$

of R-modules is called a chain complex if it is a zero sequence, i.e., $d_i d_{i+1} = 0$ for all i. The maps d_i are called *differentials of degree* -1 and (*) is denoted by (M, d) with $M = \{M_i\}$ and $d = \{d_i\}$. We let $Z_i(M) := \ker d_i$, the abelian group of *i-cycles* and $B_i(M) := \operatorname{im} d_{i+1}$, the abelian group of *$i$-boundaries* which is a subgroup of $Z_i(M)$. The quotient $H_i(M)$ is called the ith *homology group* of M. If R is a commutative ring, these are all R-modules (proof?). We also denote $Z_*(M), B_*(M)$, and $H_*(M)$ for $\{Z_i(M)\}, \{B_i(M)\}, \{H_*(M)\}$, respectively.

Construction 120.11. Let R be a commutative ring, V a free R-module of rank n and $f : L \to V$ an R-homomorphism. Then for all $x_1, \dots, x_n \in L$,

$$(x_1, \dots, x_n) \mapsto \sum_{i=1}^{n} f(x_i) x_1 \wedge \cdots \wedge \widehat{x_i} \wedge \cdots x_n,$$

where $\widehat{\ }$ means omit, defines an alternating R-module homomorphism $L^n \to \bigwedge^{n-1} L$. By the universal property of $\bigwedge^n$, this induces a homomorphism

$$d_f^{(n)} : \bigwedge^n L \to \bigwedge^{n-1} L$$

by

$$d_f^{(n)}(x_1 \wedge \cdots \wedge x_n) = \sum_{i=1}^{n} (-1)^{i+1} f(x_i) x_1 \wedge \cdots \widehat{x_i} \wedge x_n$$

for all $x_1, \dots x_n \in L$. This in turn induces a graded R-homomorphism $d_f : \bigwedge L \to \bigwedge L$ (of degree -1), with $d_f = \{d_f^{(n)}\}$. In particular, we have a chain complex $(\bigwedge L, d_f)$ satisfying

$$d_f(x \wedge y) = d_f(x) \wedge y + (-1)^{\deg x} x \wedge d_f(y)$$

for all homogeneous elements $x, y \in \bigwedge L$. The map d_f is called an *anti-derivation of degree* -1. This complex is called the *Koszul complex* of f and denoted by $K_*(f)$. If M is an R-module, we define $K_*(f, M) := K_*(f) \otimes_R M$. This complex is called the *Koszul complex* with *coefficients* in M. The differential of this complex is denoted by $d_{f,M}$.

We leave the proofs of the following two propositions as exercises.

Proposition 120.12. *Let R be a commutative ring, L an R-module and $f : L \to R$ an R-homomorphism. Then*

(1) $K_*(f)$ is a graded alternating algebra.
(2) d_f is an anti-derivation of degree -1.
(3) If M is an R-module, then $K_*(f, M)$ is a $K_*(f)$-module (in a natural way).

(4) If M is an R-module, then $d_{f,M}(x, y) = d_f(x) \cdot y + (-1)^{\deg x} x \cdot d_{f,M}(y)$ for all homogeneous elements $x \in K_*(f)$ and all elements $y \in K_*(f, M)$.

As $K_*(f) \cong K_*(f, R)$ (naturally), $Z_*(f)$ is a graded subalgebra of $K_*(f)$ and $B_*(f) \subset Z_*(f)$ is a graded 2-sided ideal. If M is an R-module, $Z_*(f, M)$ is a graded subalgebra of $K_*(f, M)$ and $B_*(f, M) \subset Z_(f, M)$ is a graded 2-sided ideal. In particular, $H_*(f) := H_*(Z_*(R))$ and $H_*(f, M) := H_*(Z_*(M))$ are graded R-algebras. The algebras $H_*(f)$ $H_*(f, M)$ is called the *Koszul homologies*, respectively.

Proposition 120.13. *Let R be a commutative ring, L an R-module, $f : L \to R$ an R-homomorphism.*

(1) The Koszul homology $H_*(f)$ is a graded alternating algebra induced by $K_*(f)$.
(2) If M is an R-module, then $H_*(f, M)$ is an $H_*(f)$-module (in the natural way).
(3) $H_*(f, M)$ is an $(R/\operatorname{im} f)$-module.
(4) $H_0(f) = R/\operatorname{im} f$ and if M is an R-module $H_0(f, M) = M/(\ker(f)\, M)$.

This proposition is quite useful. For example, one can show that if $(R, \mathfrak{m})$ is a nonzero local Noetherian ring, M a finitely generated R-module, then an M-sequence of length n (cf. Remark 98.13) exists if and only if $H_i(M \otimes R/\mathfrak{m}) = 0$ for $r = 1, \ldots, n-1$. This allows the use of homology to study the integer *depth* introduced in Remark 98.13, a very important integer in studies of commutative algebra and algebraic geometry.

Exercises 120.14.

1. Let V be an n-dimensional F-vector space. Show
$$\dim_F S^m(V) = \binom{m+n-1}{n-1} = \binom{m+n-1}{m}.$$
2. Let R be a commutative ring. Define the free commutative R-algebra on a set S and show it is a polynomial ring on $|S|$ variables.
3. Let R be a commutative ring. Define the free R-algebra on a set S and show it exists.
4. Let R be a commutative ring and A and B commutative R-algebras. Show that $A \otimes_R B$ is the coproduct of A and B as R-algebras [definition?]. [If A and B are not commutative, this is not true.]
5. Let K_i/F be (arbitrary) field extensions for $i = 1, 2$. A *composite* of K_1, K_2 over F is a field L such that there exist (field) homomorphisms $\varphi_i : K_i \to L$ fixing F, i.e., F-homomorphisms, such that L is generated by $\varphi_1(K_1)\varphi_2(K_2)$. Show that there is a bijection between $\operatorname{Spec}(K_1 \otimes_F K_2)$ and F-isomorphism classes of composites of K_1 and K_2 over F.
6. If R is a commutative ring and M an R-module, prove that the tensor algebra $T(M)$ satisfies the universal property in Construction 120.4.
7. Fill in the details needed to justify the construction of the symmetric algebra in Construction 120.5.
8. Prove $S^r(M)$ satisfies the universal property (120.7).
9. Prove $S(M)$ satisfies the universal property (120.8).
10. Prove $\bigwedge(M)$ satisfies the universal property in Remark 120.10(2).

11. Let $\varphi : R \to S$ be a homomorphism of commutative rings and M an R-module. Show that φ induces a natural isomorphism $(\bigwedge M) \otimes_R S \to \bigwedge(M \otimes_R S)$ of graded S-modules.
12. Let R be a commutative ring and M, N R-modules. Define multiplication on $(\bigwedge M \otimes_R (\bigwedge N)$ as follows: For any homogeneous elements $x, x' \in M$, $y, y' \in N$, let

$$(x \otimes y)(x' \otimes y') = (-1)^{(\deg y)(\deg x')}(x \wedge x') \otimes (y \wedge y') \tag{*}$$

Show that $(\bigwedge M) \otimes_R (\bigwedge N)$ is an alternating graded R-algebra via (*) with first homogeneous component $(M \otimes R) \otimes_R (R \otimes N) \cong M \otimes_R N$.
13. Let V be an n-dimensional F-vector space. If $y_1, \ldots, y_n$ are elements in V, show $\{y_1, \ldots, y_n\}$ is linearly dependent if and only if $y_1 \wedge \cdots \wedge y_n = 0$.
14. Let R be a commutative ring and

$$0 \to M' \to M \to M'' \to 0$$

an exact sequence of free R-modules of ranks r, n, s, respectively. Show that there exists a natural isomorphism

$$\varphi : \bigwedge^r(M') \otimes_R \bigwedge^s(M'') \to \bigwedge^n(M).$$

15. Let $M = M \oplus M''$ be a direct sum of free R-modules of finite rank. Then for all positive integers n, show that

$$\bigwedge^n(M) \cong \coprod_{r+s=n} \bigwedge^r(M') \otimes_R \bigwedge^s(M'').$$

16. Let R be a commutative ring and A an R-algebra. A *derivation* $\delta : A \to A$ is an R-homomorphism satisfying $\delta(ab) = \delta(a)b + a\delta(b)$ for all $a, b \in A$. (E.g., the derivative on rings of real differentiable functions.) Show that if M is an R-module, then any R-homomorphism $\varphi : M \to S(M)$ extends to a derivation $S(M) \to S(M)$.
17. Fill in the details in Construction 120.11.
18. Prove Proposition 120.12.
19. Prove Proposition 120.13.

121. The Determinant

In this section, we show that the determinant exists for matrices over a commutative ring using the exterior algebra of a finitely generated free module M. We construct the determinant of an R-endomorphism of M that gives rise to the determinant of its matrix representations, so is more intrinsic than the more elementary matrix proof. Recall over a commutative ring all bases for a finitely generated free module have have the same number of elements (the rank of the module). Unlike the tensor algebra over a finitely free module that has infinite rank, the exterior algebra has finite rank. More precisely, we have the following:

Theorem 121.1. *Let R be a commutative ring and M a free R-module of rank n on basis $\{x_1, \ldots, x_n\}$. Then $\bigwedge(M)$ is a free R-module of rank 2^n. More precisely,*

$$\mathcal{B}_r := \{x_{i_1} \wedge \cdots \wedge x_{i_r} \mid 1 \leq i_1 < \cdots < i_r \leq n\}$$

is a basis for $\bigwedge^r(M)$ *for* $r \leq n$. *In particular,*

$$\operatorname{rank} \bigwedge^r(M) = \begin{cases} \binom{n}{r} & \text{if } r \leq n \\ 0 & \text{otherwise.} \end{cases}$$

To prove this we need a preliminary idea and its development in the case of exterior algebras.

Construction 121.2. Let R be a commutative ring, M an R-module, and $A = \coprod_{i \geq 0} A_i$ a graded R-algebra.

For each $i \geq 0$, define an R-homomorphism $\sigma : A_i \to A_i$ by $z \mapsto (-1)^i z$. Then σ induces a graded R-algebra homomorphism $\sigma : A \to A$, since $(-1)^i(-1)^j = (-1)^{i+j}$. Note that $\sigma^2 = 1_A$.

An R-homomorphism $\delta : A \to A$ is called an *anti-derivation* if

$$\delta(ab) = \delta(a)b + \sigma(a)\delta(b) \text{ for all } a, b \in A.$$

If A is $T(M)$, $S(M)$, or $\bigwedge(M)$, then M generates A as an R-algebra, hence δ is determined by what it does to $A_1 = M$.

Case. $A = T(M)$:

Let $\widetilde{\varphi} : M \to T(M)$ be an R-homomorphism. Then $M \times M \to T(M)$ defined by

$$(a, b) \mapsto \widetilde{\varphi}(a)b - a\widetilde{\varphi}(b) = \widetilde{\varphi}(a)b + \sigma(a)\widetilde{\varphi}(b) \tag{*}$$

is R-bilinear, so induces an R-homomorphism $T^2(M) \to T(M)$. Inductively, we obtain R-homomorphisms $T^n(M) \to T(M)$, i.e., $\widetilde{\varphi}$ induces an anti-derivation $T(M) \to T(M)$. We also call this map $\widetilde{\varphi}$.

Case. $A = \bigwedge(M)$:

Let φ be the composition

$$T(M) \xrightarrow{\widetilde{\varphi}} T(M) \xrightarrow{-} \bigwedge(M),$$

with $\widetilde{\varphi}$ the anti-derivation of the previous case.

Claim. If φ satisfies

$$\varphi(x) \wedge x - x \wedge \varphi(x) = 0 \text{ for all } x \in M, \tag{121.3}$$

then φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$ that we shall also call φ:

As $\widetilde{\varphi} : T(M) \to T(M)$ is an anti-derivation, to show that φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$, it suffices to show that

$$\widetilde{\varphi}(\mathfrak{B}) \subset \mathfrak{B} = \ker\left(^- : T(M) \to \bigwedge(M)\right).$$

Let $x \in M$. By (*), we have

$$\widetilde{\varphi}(x^2) = \widetilde{\varphi}(x \cdot x) = \widetilde{\varphi}(x)x + \sigma(x)\widetilde{\varphi}(x) = \widetilde{\varphi}(x)x - x\widetilde{\varphi}(x). \tag{†}$$

Applying the map $^{-}$ to equation (†), shows that $\widetilde{\varphi}(x^2)$ lies in $\mathfrak{B}$. Let a and b lie in $T(M)$ and x in M. Then

$$\begin{aligned}\widetilde{\varphi}(ax^2b) &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2b)\\ &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2)b + \sigma(a)\sigma(x^2)\widetilde{\varphi}(b)\\ &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2)b + \sigma(a)\sigma(x)^2\widetilde{\varphi}(b).\end{aligned}$$

Since $x^2, \sigma(x)^2$ lie in $\mathfrak{B}$ and $\widetilde{\varphi}(x^2)$ lies in $\mathfrak{B}$ by (†) with $\mathfrak{B}$ a 2-sided ideal in $T(M)$, we conclude that $\widetilde{\varphi}(ax^2b)$ lies in $\mathfrak{B}$ for all a, b in $T(M)$ and for all x in M. Since $\mathfrak{B}$ is generated by x^2, $x \in M$, by Remark 120.10(1), we have

$$\mathfrak{B} = \langle ax^2b \mid a, b \in T(M), x \in M\rangle.$$

Therefore, $\widetilde{\varphi}(\mathfrak{B}) \subset \mathfrak{B}$, and the claim is established.

Examples 121.4. Let R be a commutative ring and M an R-module. Suppose that we are in the case $A = \bigwedge(M)$ above together with its notation.

1. If $\varphi(x)$ lies in M for all x in M, then equation (121.3) becomes
$$2\varphi(x) \wedge x = 0 \text{ for all } x \in M.$$
2. If $\varphi(x)$ lies in R for all x in M, then equation (121.3) always holds. In particular, as $R = \bigwedge^0(M) \subset \bigwedge(M)$, if φ lies in $M^* = \mathrm{Hom}_R(M, R)$, then φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$ that we also call φ.
3. Suppose that M is R-free on basis $\mathcal{B} = \{x_1, \dots, x_n\}$ and $\mathcal{B}^* = \{f_1, \dots, f_n\}$, the dual basis for $M^* = \mathrm{Hom}_R(M, R)$. Then each f_i defines an anti-derivation $f_i : \bigwedge(M) \to \bigwedge(M)$.

We now proceed to the proof of Theorem 121.1.

PROOF. We use the notation of Example 121.4(3). As

$$x_{\sigma(i_1)} \wedge \cdots \wedge x_{\sigma(x_{i_r})} = \pm x_{i_1} \wedge \cdots \wedge x_{i_r}, \text{ for all } \sigma \in S_r,$$

we have

$$\mathcal{B}_r = \{x_{i_1} \wedge \cdots \wedge x_{i_r} \mid 1 \le i_1 < \cdots < i_r \le n\}$$

generates $\bigwedge^r(M)$. As $\bigwedge(M) = \coprod_{r\ge 0} \bigwedge^r(M)$ is graded, it suffices to show that $\mathcal{B}_r$ is R-linearly independent. Suppose this is false. Since $\mathcal{B} = \mathcal{B}_1$ is linearly independent, there exists a mimimal r such that $\mathcal{B}_r$ is linearly dependent. Let

$$(*) \qquad \sum_{1\le i_1<\cdots<i_r\le n} a_{i_1,\dots,i_r} x_{i_1} \wedge \cdots \wedge x_{i_r} = 0 \text{ in } \bigwedge^r(M),$$

not all $a_{i_1,\dots,i_r}$ zero. Changing notation, we may assume that $a_{1,j_2,\dots j_r}$ is nonzero with $1 < j_2 < \cdots < j_r \le n$. Applying the anti-derivation arising from f_1 in $\mathcal{B}^*$ to $(*)$, we get

$$\sum_{1\le i_2<\cdots<i_r\le n} a_{1,i_2,\dots,i_r} x_{i_2} \wedge \cdots \wedge x_{i_r} = 0,$$

since $f_1(x_i) = \delta_{i1}$. This implies that $\mathcal{B}_{r-1}$ is linearly dependent, a contradiction. □

Corollary 121.5. *The sign map* $\operatorname{sgn} : S_n \to \{\pm 1\}$ *given by*

$$\operatorname{sgn}\sigma = \begin{cases} 1 & \text{if } \sigma \text{ is a product of an even number of transpositions} \\ -1 & \text{otherwise.} \end{cases}$$

is a well-defined group homomorphism.

Proof. Let R be a commutative ring and M a free R-module on basis $\{x_1, \dots, x_n\}$, then $\bigwedge^n(M)$ is R-free on basis $\{x_1 \wedge \cdots \wedge x_n\}$. Hence if $\sigma \in S_n$. there exists a unique $\operatorname{sgn}(\sigma)$ in $\{\pm 1\}$ satisfying

$$x_{\sigma(1)} \wedge \cdots \wedge x_{\sigma(n)} = \operatorname{sgn}(\sigma) x_1 \wedge \cdots \wedge x_n. \tag{121.6}$$

It is easily checked that sgn is a group homomorphism. □

Construction 121.7. Let R be a commutative ring and M an R-free module on basis $\{x_1, \dots, x_n\}$. Let $f : M \to M$ be an R-homomorphism. Write

$$f(x_i) = \sum_{j=1}^{n} \alpha_{ji} x_j \text{ for } i = 1, \dots, n$$

and $A = (\alpha_{ij}) = [f]_{\mathcal{B}}$, the matrix representation of f relative to the (ordered) basis $\mathcal{B}$. Then f induces a unique graded R-algebra homomorphism $\wedge f : \bigwedge(M) \to \bigwedge(M)$. Since $\{x_1 \wedge \cdots \wedge x_n\}$ is a basis for $\bigwedge^n(M)$, there exists a unique element λ in R satisfying $\wedge^n f = \lambda 1_{\bigwedge^n(M)}$, where $\wedge f|_{\bigwedge^n(M)} = \wedge^n f$. In particular, λ is an eigenvalue of $\wedge^n f$ on the rank one free R-module $\bigwedge^n(M)$. Using equation (121.6) and the alternating property, we see that

$$\begin{aligned} \lambda 1_{\bigwedge^n(M)} &= \wedge^n f(x_1 \wedge \cdots \wedge x_n) \\ &= f(x_1) \wedge \cdots \wedge f(x_n) = \sum_{j=1}^{n} \alpha_{j1} x_j \wedge \cdots \wedge \sum_{j=1}^{n} \alpha_{jn} x_j \\ &= \sum_{S_n} \operatorname{sgn}(\sigma) \alpha_{\sigma(1)1} \cdots \alpha_{\sigma(n)n} x_1 \wedge \cdots \wedge x_n. \end{aligned}$$

The element

$$\lambda = \sum_{S_n} \operatorname{sgn}(\sigma) \alpha_{\sigma(1)1} \cdots \alpha_{\sigma(n)n}$$

is called the *determinant* of f and denoted $\det f$. As $\det f$ is an eigenvalue for $\wedge^n f$ on a rank one free R-module, it is unique and independent of $\mathcal{B}$. The map $\det : \operatorname{End}_R(M) \to R$ defined by $\det f = \wedge^n f$ is n-multilinear, alternating and satisfies $\det 1_V = 1_R$. It is called the *determinant (function)*.

Remark 121.8. The above shows the existence and uniqueness of the usual determinant (function) on matrices $\mathbb{M}_n(R)$ over a commutative ring R, i.e., it is the unique alternating, n-multilinear function $D : \mathbb{M}_n(R) \to R$ on columns (or rows) of $n \times n$ matrices over R satisfying $D(I) = 1$. Indeed, if $\mathcal{S}$ is the standard basis for R^n, then $D(A) = \det(A)$, viewing A as a linear transformation on R^n. In particular, $\det(A) = 0$ if and only if A is not a unit in $\mathbb{M}_n(R)$.

Next let $g : M \to M$ be another R-homomorphism and $B = [g]_{\mathcal{B}}$. Then

$$\wedge^n(fg) = \wedge^n f \circ \wedge^n g \;\text{ so }\; \det fg = \det f \det g,\; i.e., \det AB = \det A \det B.$$

More generally, f, g induce $\wedge^r f, \wedge^r g : \bigwedge^r(M) \to \bigwedge^r(M)$ for each $r = 1, \dots, n$. Set $A^{(r)} = [\wedge^r f]_{\mathcal{B}_r}$ and $B^{(r)} = [\wedge^r g]_{\mathcal{B}_r}$ in $\mathbb{M}_{\binom{n}{r}}(R)$. Then

$$(AB)^{(r)} = A^{(r)} B^{(r)}$$

which gives identities of degree r in the entries of A and B called the *Binet-Cauchy Equations*. Note that $A^{(r)}$ is an $\big(\binom{n}{r} \times \binom{n}{r}\big)$-matrix whose entries are the rth order minors of A, etc.

To get additional formulae, let $A = (a_{ij})$ be an $n \times n$-matrix over a commutative ring R. Then we view A as the matrix representation of the appropriate $f : R^n \to R^n$ in the standard basis $\{e_1, \dots, e_n\}$. Set $S = \{1, \dots, n\}$. If H is a subset of S consisting of r elements, let $\underline{H}$ be the ordered r-tuple $(i_1, \dots, i_r)$ where $1 \leq i_1 < \cdots < i_r \leq n$ with all $i_j \in H$. and $e_{\underline{H}} = e_{i_1} \wedge \cdots \wedge e_{i_r}$. If K is another subset of S, then

$$e_{\underline{H}} \wedge e_{\underline{K}} = \begin{cases} \varepsilon_{\underline{H},\underline{K}} e_{\underline{H \cup K}} & \text{if } H \cap K = \emptyset \\ 0 & \text{if } H \cap K \neq \emptyset, \end{cases} \tag{121.9}$$

where $\varepsilon_{\underline{H},\underline{K}} = (-1)^m$, if $H \cap K = \emptyset$ and m is the number of transpositions modulo two to get $\underline{H \cup K}$ from $\underline{H} \cup \underline{K}$. If H is a subset of S, we denote by H' the complement of H in S. Fix r, $1 \leq r \leq n$, and let $S(r) = \{I \mid I \in P(S) \text{ with } |I| = r\}$, where $P(S)$ is the power set of S. i.e., the set of subsets of S. Let $A_{\underline{I},\underline{H}}$ denote the submatrix of A with rows of A associated to $\underline{I}$, $I \in S(r)$, and columns of A associated to $\underline{H}$ i.e, $\det A_{\underline{I},\underline{H}}$ is the r-order minor associated to $\underline{I'}$ and $\underline{H'}$.

Fix $H \in S(r)$. Then check we have

$$\wedge^r(f)(e_{\underline{H}}) = \sum_{I \in S(r)} \det(A_{\underline{I},\underline{H}})\, e_{\underline{I}}. \tag{121.10}$$

Therefore, for such a fixed H, we have

$$\begin{aligned} (\det A \cdot \varepsilon_{\underline{H},\underline{H'}})\, e_{\underline{S}} &= \big(\wedge^r(f)\big)(e_{\underline{H}}) \wedge \big(\wedge^{n-r}(f)\big)(e_{\underline{H'}}) \\ &= \sum_{I \in S(r)} \sum_{J \in S(n-r)} \det(A_{\underline{I},\underline{H}}) \det(A_{\underline{J},\underline{H'}})\, e_{\underline{I}} \wedge e_{\underline{J}} \\ &= \sum_{I \in S(r)} \sum_{J \in S(n-r)} \det(A_{\underline{I},\underline{H}}) \det(A_{\underline{J},\underline{H'}})\, \varepsilon_{\underline{I},\underline{J}} e_{\underline{S}} \\ &= \sum_{I \in S(r)} \varepsilon_{\underline{I},\underline{I'}} \det(A_{\underline{I},\underline{H}}) \det(A_{\underline{I'},\underline{H'}})\, e_{\underline{S}} \end{aligned} \tag{†}$$

by (121.9). In particular,

$$\det A = \sum_{I \in S(r)} \varepsilon_{\underline{H},\underline{H'}} \varepsilon_{\underline{I},\underline{I'}} \det(A_{\underline{I},\underline{H}}) \det(A_{\underline{I'},\underline{H'}}).$$

This is called the *Laplace expansion* of $\det A$. For $r = 1$, this is the well-known expansion by minors along rows. By symmetry, we have an analogous result on columns by using right modules. If we let $A(i, j)$ be the $(n-1)$-order minor of A, i.e., the determinant of the submatrix of A by deleting the ith row and jth column, then $(-1)^{i+j}A(i, j)$ is called the (i, j)*th cofactor* of A. Define the *classical adjoint* $\operatorname{adj}(A)$ of A to be the transpose of the matrix of cofactors of A, i.e., the (i, j) entry of $\operatorname{adj}(A)$ is $(-1)^{i+j}A(j, i)$. By Laplace expansion, we check that

$$\sum_{i=1}^{n} a_{i1}(-1)^{i+1}A(j, 1) + \cdots + a_{in}(-1)^{i+1}A(j, n)$$

is the expansion of a determinant whose jth row equals the ith row of A while the other rows are as in A. If $i \neq j$ this means this matrix has two rows the same so is zero, hence is zero off the main diagonal. It follows that

$$A \cdot \operatorname{adj}(A) = \det A \cdot I$$

with I the $n \times n$ identity matrix. Similarly, $\operatorname{adj}(A) \cdot A = \det A \cdot I$.

Let $M = \sum_{i=1}^{n} Rx_i$ be an R-module. If we have a system of equations $\sum_{j=1}^{n} a_{ji}x_j = 0$ for $i = 1, \ldots, n$, then we must have $\det(A)M = 0$. In particular, $\det A$ is in the annihilator of M. It follows that if $\operatorname{ann}_R(M) = 0$ and $M \neq 0$, then $\det A = 0$. Next suppose that A is invertible, i.e., $\det A$ is a unit (respectively, R is a domain and $M = R$) and we want to solve a system of equations $\sum_{j=1}^{n} a_{ji}x_j = b_i$ for $i = 1, \ldots, n$ in M (respectively the quotient field of R), i.e., if $x = (x_1 \cdots x_n)^t$ and $b = (b_1 \cdots b_n)^t$ column matrices, then we want to solve the matrix equation $Ax = b$. Then the solution is $x = (\det A)^{-1}\operatorname{adj}(A)b$. This is called *Cramer's Rule*.

Exercises 121.11.

1. Show equation (121.10) holds.
2. Show equation (†) holds (in the equation right after equation (121.10)).
3. Check that Cramer's Rule in the text is the usual one.
4. Let R be a commutative ring and $A \in \mathbb{M}_n(R)$. Let $(A(i|j) \in \mathbb{M}_{n-1}(R)$ be the matrix deleting the ith row and jth column of A and $C_{iJ} = (-1)^{i+j}\det(A(i|j)$, the (i, j)th *cofactor* of A. Show $\sum_{i=1}^{n} A_{ik}C_{ij} = \delta_{jk}\det A$.
5. Let R be a commutative ring, $A \in R^{m\times n}$, and $B \in R^{n\times m}$ with $m \leq n$. Set $S = \{1, \ldots, n\}$ and $S(m) = \{I \mid I \in P(S) \text{ with } |I| = m\}$. Let $\underline{M}$ be the ordered tuple $(1, 2, \ldots, m)$ where $M = \{1, \ldots, m\} \subset S$. For each $I \in S(m)$, let $A_{\underline{M},\underline{I}}$ denote the $m \times m$ matrix formed from A using the columns I of A and $B_{\underline{I},\underline{M}} \in \mathbb{M}_m(R)$ denote the $m \times m$ matrix formed from B using the rows I of B. Prove

$$\det(AB) = \sum_{I\in S(m)} \det(A_{\underline{M},\underline{I}})\det(B_{\underline{I},\underline{M}}).$$

   This is also called the Binet-Cauchy Equation.

6. Let V be a finite dimensional real inner product space with inner product $\langle\ ,\ \rangle$. Show that if $v = (v_1, \dots, v_n)$ and $w = (w_i, \dots, w_n)$ are elements of V, then
$$\sum_{1 \le i < j \le n} \det \begin{pmatrix} v_i & v_j \\ w_i & w_j \end{pmatrix}^2 = \det \begin{pmatrix} \langle v, v\rangle^2 & \langle v, w\rangle \\ \langle v, w\rangle & \langle w, w\rangle^2 \end{pmatrix}.$$
In particular, the Cauchy-Schwarz Inequality holds.
7. Let R be a commutative ring. Suppose that $E \in M_n(R)$ has the form
$$E = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
with $A \in GL_r(R)$, $B \in R^{r\times(n-r)}$, $C \in R^{(n-r)\times r}$, and $D \in \mathbb{M}_{n-r}(R)$.
Let $D' = D - CA^{-1}B \in \mathbb{M}_{n-r}(R)$, called the *Schur complement* of A in E and denoted by E/A. Show that
$$\det E = \det A \det(E/A).$$
In particular, if $C = 0$, then $\det E = \det A \det D$.
8. Let R be a commutative ring, $A \in R^{m\times n}$ and $B \in R^{n\times m}$. Show that $\det(I_m + AB) = \det(I_n + BA)$, where I_m and I_n are the appropriate identity matrices.
9. Let R be a commutative ring, $M \in R^{n\times(n-r)}$, and $a_1, \dots, a_{2r} \in R^{n\times 1}$. Show
$$\sum_{\sigma} (-1)^{k_1+\cdots+k_r} \det\left(M_{\underline{\{1,\dots,r\}},\underline{\{a_{k_1},\dots,a_{k_r}\}}}\right) \det\left(M_{\underline{\{a_{k_{r+1}},\cdots,a_{k_{2r}}\},\_\{1,\dots,r\}}}\right) = 0,$$
where $k_1, \cdots, k_{2r}$ is a permutation of $\{1, 2, \dots, 2r\}$ and $\sigma \in S_{2r}$ is the permutation with $1 \le k_1 < \cdots < k_r \le 2r$ and $k_{r+1} < \cdots < k_{2r}$.

CHAPTER XXI

Introduction to Homological Algebra

Homology and cohomology are groups that are essential in measuring algebraic properties of groups, rings, and modules. It is also used in geometry and topology. In this chapter, we set up the basics of this theory working in the special case of modules. [The proper setting is abelian categories.] Since our rings are not necessarily commutative, as usual R-module will mean left R-module. Obvious definitions and results for right R-modules will be left to the reader. In addition, many results have "dual" statements, i.e., reversing all maps give the analogous results, and if so, the proofs will be omitted. There are a lot of definitions and many of the proofs follow via diagram chasing as you learned if you proved the Five Lemma, so some details will be left to the reader. In addition, there is a lot of notation involved. We will also introduce collection of modules that are very useful. These are the injective modules that arise from abelian divisible groups and whose generalization itself is crucial in algebraic geometry, projective modules that generalize free modules (and are dual to injective modules) that generalize to vector bundles, and flat modules that generalize projective modules and are important when extending a base ring. Reading this chapter can be omitted unless needed, since there will be little application outside of the theory itself.

122. Homology

Often a *chain complex* of R-modules, i.e., a zero sequence of R-modules and R-homomorphisms, is written as

$$\cdots \xrightarrow{d_{n+2}} A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \xrightarrow{d_{n-2}} \cdots$$

and the R-homomorphisms d_i are called *differentials*. One usually denotes such a chain complex by (A_*, d_*). If (A_*, d_*) is a chain complex, then we define the quotient of R-modules

$$H_n(A) := \ker d_n / \operatorname{im} d_{n+1}$$

and call it the *nth homology of* (A_*, d_*). If (A_*, d_*) is a chain complex, the set A_n is called the set of *n-chains*. We call

$$Z_n(A) := \ker d_n \text{ the set of } n\text{-cycles of } A_n$$
$$B_n(A) := \operatorname{im} d_{n+1} \text{ the set of } n\text{-boundaries of } A_n.$$

So $H_n(A) = Z_n(A)/B_n(A)$ and every element in $H_n(A)$, called a *homology class*, is represented by an n-cycle. This agrees with its historical topological interpretation.

If we write the indices of the A_* and the d_n to go up, i.e.,

$$\cdots \xrightarrow{d_{n-2}} A_{n-1} \xrightarrow{d_{n-1}} A_n \xrightarrow{d_n} A_{n+1} \xrightarrow{d_{n+1}} \cdots,$$

then (A_*, d_*) is called a *cochain complex* with the differentials written as d^n. We use the notation (A^*, d^*) for a cochain complex. Analogously, the elements in A^n are called *n-cochains* and

$$Z^n(A) := \ker d^n \text{ is called the set of } n\text{-}cocycles \text{ of } A_n$$
$$B^n(A) := \operatorname{im} d^{n-1} \text{ is called the set of } n\text{-}coboundaries \text{ of } A_n$$

and

$$H^n(A) := \ker d^n / \operatorname{im} d^{n-1} = Z^n(A)/B^n(A)$$

called the *nth-cohomology* of A_n with elements called *cohomology classes.*

Therefore, a (co)chain complex is exact, i.e., *acyclic*, if and only if its (co)homology is trivial, i.e., (co)homology measures the obstruction from a chain complex to be exact.

A *chain homomorphism* (or *chain homomorphism of degree zero*) of chain complexes $f_* : (A_*, d_*) \to (A'_*, d'_*)$ consists of R-modules and R-homomorphisms is $f_n : A_n \to A'_n$ for all n satisfying

$$\begin{tikzcd} A_{n+1} \arrow[r, "f_{n+1}"] \arrow[d, "d_{n+1}"'] & A'_{n+1} \arrow[d, "d'_{n+1}"] \\ A_n \arrow[r, "f_n"'] & A'_n \end{tikzcd}$$

commutes for all n.

Let 0 denote the trivial chain complex $(0_*, 0_*)$. A sequence of chain complexes and chain homomorphisms

$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$

is called an *exact* sequence of chain complexes of R-modules if, writing out the chain complexes vertically and horizontally, we have a commutative diagram with exact rows.

The obvious definitions are given for *cochain homomorphisms.*

The study of (co)homology, in its general context, is a major subject in mathematics. We begin with a fundamental lemma that will be used to turn a chain complex into exact sequence of their corresponding homology groups. As mentioned above some details will be left to the reader.

Lemma 122.1. (Snake Lemma) *Let*

$$\begin{tikzcd} & A \arrow[r, "f"] \arrow[d, "\alpha"'] & B \arrow[r, "g"] \arrow[d, "\beta"'] & C \arrow[r] \arrow[d, "\gamma"'] & 0 \\ 0 \arrow[r] & A' \arrow[r, "f'"] & B' \arrow[r, "g'"] & C' & \end{tikzcd}$$

be a commutative diagram of R-modules and R-homomorphisms with exact rows. Then there exists an R-homomorphism $\partial : \ker\gamma \to \operatorname{coker}\alpha$ so that the following sequence is exact:

$$\ker\alpha \xrightarrow{f|_{\ker\alpha}} \ker\beta \xrightarrow{g|_{\ker\beta}} \ker\gamma \xrightarrow{\partial} \operatorname{coker}\alpha \xrightarrow{\overline{f}'} \operatorname{coker}\beta \xrightarrow{\overline{g}'} \operatorname{coker}\gamma,$$

with $\overline{f}'$ and $\overline{g}'$ the induced maps. Moreover, in this diagram, we have $f|_{\ker\alpha}$ is a monomorphism if f is a monomorphism and $\overline{g}$ is an epimorphism if g is an epimorphism.

Proof. Except for the definition of ∂, the exactness at all the positions is easily checked by diagram chasing. So we shall only indicate what the map ∂ is. Although the inverse of a map is not a function, one checks that if $x \in \ker\gamma \subset C$, that

$$\partial : \ker\gamma \to \operatorname{coker}\alpha \ \text{ by } \ x \mapsto f'^{-1}\beta g^{-1}(x) + \operatorname{im}\alpha$$

is well-defined, i.e., independent of the choices of the preimages. □

The snake lemma implies:

Corollary 122.2. *If*

$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$

is an exact sequence of chain complexes of R-modules, then for all n the following are exact:

$$0 \to \ker(d_A)_{n-1} \xrightarrow{(f_{n-1})|_{\ker d_A}} \ker(d_B)_{n-1} \xrightarrow{(g_{n-1})|_{\ker d_B}} \ker(d_C)_{n-1}$$

and

$$\operatorname{coker}(d_A)_{n+1} \xrightarrow{\overline{f}_{n+1}} \operatorname{coker}(d_B)_{n+1} \xrightarrow{\overline{g}_{n+1}} \operatorname{coker}(d_C)_{n+1} \to 0$$

are exact.

Corollary 122.3. *Let*

$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$

be an exact sequence of chain complexes of R-modules. Then the following diagram is commutative and has exact rows:

$$\begin{tikzcd}
 & \operatorname{coker}(d_A)_{n+1} \arrow[r, "\overline{f}_{n+1}"] \arrow[d, "\overline{d}_A"'] & \operatorname{coker}(d_B)_{n+1} \arrow[r, "\overline{g}_{n+1}"] \arrow[d, "\overline{d}_B"'] & \operatorname{coker}(d_C)_{n+1} \arrow[r] \arrow[d, "\overline{d}_C"'] & 0 \\
0 \arrow[r] & \ker(d_A)_{n-1} \arrow[r, "f_{n-1}"] & \ker(d_B)_{n-1} \arrow[r, "g_{n-1}"] & \ker(d_C)_{n-1} &
\end{tikzcd}$$

where the maps $\overline{f}$, $\overline{g}$, $\overline{d}_A$, $\overline{d}_B$, $\overline{d}_C$ are the induced maps and the bottom maps are restrictions to the appropriate kernels.

Proof. By the Snake Lemma, for all n, we have a commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & \ker(d_A)_n \arrow[r, "f_n"] \arrow[d] & \ker(d_B)_n \arrow[r, "g_n"] \arrow[d] & \ker(d_C)_n \arrow[d] & \\
0 \arrow[r] & A_n \arrow[r, "f_n"] \arrow[d, "(d_A)_n"'] & B_n \arrow[r, "g_n"] \arrow[d, "(d_B)_n"'] & C_n \arrow[r] \arrow[d, "(d_C)_n"'] & 0 \\
0 \arrow[r] & \to A_{n-1} \arrow[r, "f_{n-1}"] \arrow[d] & B_{n-1} \arrow[r, "g_{n-1}"] \arrow[d] & C_{n-1} \arrow[r] \arrow[d] & 0 \\
 & \dfrac{A_{n-1}}{\operatorname{im}(\overline{d}_A)_n} \arrow[r, "\overline{f}_{n-1}"] & \dfrac{B_{n-1}}{\operatorname{im}(\overline{d}_B)_n} \arrow[r, "\overline{g}_{n-1}"] & \dfrac{C_{n-1}}{\operatorname{im}(\overline{d}_C)_n} \arrow[r] & 0
\end{tikzcd}$$

where the maps on the top row are restricted to the kernels. It follows that the induced maps given by the Isomorphism Theorem produces a commutative diagram with exact rows

$$\begin{tikzcd}
 & \dfrac{A_{n-1}}{\operatorname{im}(\overline{d}_A)_n} \arrow[r, "\overline{f}_{n-1}"] \arrow[d, "(\overline{d}_A)_{n-1}"'] & \dfrac{B_{n-1}}{\operatorname{im}(\overline{d}_B)_n} \arrow[r, "\overline{g}_{n-1}"] \arrow[d, "(\overline{d}_B)_{n-1}"'] & \dfrac{C_{n-1}}{\operatorname{im}(\overline{d}_C)_n} \arrow[r] \arrow[d, "(\overline{d}_C)_{n-1}"'] & 0 \\
0 \arrow[r] & \ker(d_A)_{n-2} \arrow[r, "f_{n-1}"] & \ker(d_B)_{n-2} \arrow[r, "g_{n-1}"] & \ker(d_C)_{n-2} &
\end{tikzcd}$$

with the bottom horizontal row with maps restricted to the appropriate kernels. By the previous corollary, the rows are exact. The result follows by the Snake Lemma. $\square$

Theorem 122.4. (The Long Exact Sequence in Homology) *Let*

$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$

be a an exact sequence of chain complexes of R-modules. Then there exists an R-homomorphism $\partial_{n+1} : H_{n+1}(C) \to H_n(A)$ for all n called the connecting homomorphism of the sequence such that we have a long exact sequence in homology

$$\cdots \to H_{n+1}(C) \xrightarrow{\partial_{n+1}} H_n(A) \xrightarrow{\overline{f}_n} H_n(B) \xrightarrow{\overline{g}_n} H_n(C) \xrightarrow{\partial_n} H_{n-1}(A) \to \cdots$$

where $\overline{f}_n$ and $\overline{g}_n$ are the induced maps for all n.

Proof. The proceeding corollary gives a diagram that satisfies the hypothesis of the Snake Lemma. Let $D_* = A_*, B_*,$ or C_*. So $\ker(d_D)_n = Z_n(D)$ and $\operatorname{im}(d_D)_{n+1}$ are the n-boundaries. It follows that the induced maps imply that

$$H_n(D) = \ker\big(D_n/\operatorname{im}(d_D)_{n+1} \to Z_{n-1}(D)\big)$$
$$H_{n-1}(D) = \operatorname{coker}\big(D_{n-1}/\operatorname{im}(\overline{d}_D)_n \to Z_{n-1}(D)\big).$$

Therefore, applying the Snake Lemma to the diagram in the conclusion of the last corollary, we have

$$H_n(A) \xrightarrow{\overline{f}_n} H_n(B) \xrightarrow{\overline{g}_n} H_n(C) \xrightarrow{\partial_n} H_{n-1}(A) \xrightarrow{\overline{f}_n} H_{n-1}(B) \xrightarrow{\overline{g}_n} H^{n-1}(C)$$

where ∂_n is the connecting homomorphism. Glueing these sequences together produces the desired result.

□

Remark 122.5. It is common to write the long exact sequence in homology as

$$\begin{tikzcd} H_*(A) \arrow[rr, "f_*"] & & H_*(B) \arrow[dl, "g_*"] \\ & H_*(C) \arrow[ul, "\partial_*"] & \end{tikzcd}$$

called the *exact triangle* in homology.

Computation 122.6. The construction of the connecting homomorphism in the above theorem can be explicitly defined as follows: Suppose that $z \in H_n(C)$. Choose a cycle in $c \in Z_n(C)$ whose homology class is z. By hypothesis, there exists an element $b \in B_n$ satisfying $g_n(b) = c$ and $(d_B)_n(b) \in B_{n-1}$. As $(d_B)_{n-1}(d_B)_n(b) = 0$, there exists $a \in Z_{n-1}(A)$ satisfying $f_{n-1}(a) = (d_B)_n(b)$. Then $\partial_n(z)$ is equal to the homology class of a in $H_{n-1}(A)$.

We must also show any R-homomorphisms between short exact sequences of chain complexes induces a homomorphism of exact triangles in homology.

Theorem 122.7. (Naturality of the Long Exact Sequence in Homology) *Let*

$$\begin{tikzcd} 0 \arrow[r] & (A_*, d_A) \arrow[r, "f"] \arrow[d, "\alpha_*"'] & (B_*, d_B) \arrow[r, "g"] \arrow[d, "\beta_*"'] & (C_*, d_C) \arrow[r] \arrow[d, "\gamma_*"'] & 0 \\ 0 \arrow[r] & (A'_*, d_{A'}) \arrow[r, "f'"] & (B'_*, d_{B'}) \arrow[r, "g'"] & C'_*, d_{C'}) \arrow[r] & 0 \end{tikzcd}$$

be a commutative diagram of chain complexes of R-modules. Then there exists a commutative diagram with exact rows

$$\begin{tikzcd} \cdots \arrow[r] & H_n(A) \arrow[r, "f_n"] \arrow[d, "\overline{\alpha}_n"'] & H_n(B) \arrow[r, "g_n"] \arrow[d, "\overline{\beta}_n"'] & H_n(C) \arrow[r, "\partial_n"] \arrow[d, "\overline{\gamma}_n"'] & H_{n-1}(A) \arrow[r] \arrow[d, "\overline{\alpha}_{n-1}"'] & \cdots \\ \cdots \arrow[r] & H_n(A') \arrow[r, "f'_n"] & H_n(B') \arrow[r, "g'_n"] & H_n(C') \arrow[r, "\partial'_n"] & H_{n-1}(A') \arrow[r] & \cdots , \end{tikzcd}$$

where ∂_n and ∂'_n are the corresponding connecting homomorphisms and $\overline{f}_n$, $\overline{g}_n$ $\overline{\alpha}_n$, $\overline{\beta}_n$, and $\overline{\gamma}_n$ the induced maps for all n.

Proof. Viewing the construction of the homology sequence (excluding the connecting homomorphisms) for each exact sequences of chains and the maps connecting them as well as the ones induced by the construction show the first two squares commute. (We do not write down the three dimensional diagram.) The commutativity of the last square is a somewhat more complicated diagram chase. Let $z \in H_n(C)$ and choose $c \in C_n$ with z as its homology class. Let $z' = \overline{\gamma}_n(z)$ and choose $c' \in C'_n$ with z' as its homology class. Now choose $b' \in B'_n$ such that $b' \mapsto c'$. Then the image of b' of $b \in B'_n$ maps to c' by commutativity. By Computation 122.6, we have $\partial_n(z') \in H_{n-1}(A')$ is the homology class of $a' \in Z_{n-1}(A')$ such that $a' = (d_{B'})_n$, i.e., by the image of a representative of $\partial'_n(z')$. Consequently, $\partial'_n(z) \mapsto \partial_n(z')$ $\square$

The term naturality roughly means that when we have a commutative diagram and act on it, say by Hom_R, take quotients, etc, the induced diagram is still commutative.

Next, we wish to see when two short exact sequences of chains of R-modules give rise to the same homology. We first need a definition (arising from topology).

Definition 122.8. Let (A_*, d) and (A'_*, d') be chain complexes of R-modules. A map $s_* : (A_*, d) \to (A'_*, d')$ is called a *chain map of degree i* if for all n we have a commutative diagram

$$\begin{tikzcd}
A_n \arrow[r, "f_n"] \arrow[d, "d_n"'] & A'_{n+i} \arrow[d, "d'_{n+i}"] \\
A_{n-1} \arrow[r, "f_{n+1}"'] & A'_{n-1+i}
\end{tikzcd}$$

of R-modules and R-homomorphisms. So a chain map is a chain map of degree 0. Let $f_*, g_* : (A_*, d_A) \to (B_*, d_B)$ be chain maps of chain complexes of R-modules. We say that f and g are *chain homotopic* if there exists a chain map $s_* : (A_*, d_A) \to (B_*, d_B)$ of degree $+1$ that satisfies

$$f_n - g_n = d_{B_{n+1}} s_n + s_{n-1} d_{A_n},$$

i.e., we have a commutative diagram

$$\begin{tikzcd}
{} \arrow[r] & A_{n+1} \arrow[r, "(d_A)_{n+1}"] \arrow[d, shift right, "f_{n+1}"'] \arrow[d, shift left, "g_{n+1}"] & A_n \arrow[r, "(d_A)_n"] \arrow[dl, "s_n"'] \arrow[d, shift right, "f_n"'] \arrow[d, shift left, "g_n"] & A_{n-1} \arrow[r] \arrow[dl, "s_{n-1}"'] \arrow[d, shift right, "f_{n-1}"'] \arrow[d, shift left, "g_{n-1}"] & {} \\
{} \arrow[r] & B_{n+1} \arrow[r, "(d_B)_{n+1}"'] & B_n \arrow[r, "(d_B)_n"'] & B_{n-1} \arrow[r] & {}\,.
\end{tikzcd}$$

for all n (viewing the vertical maps as $f_n - g_n$).

Chain homotopies are useful because of the following observation:

Lemma 122.9. *Let $f_*, g_* : (A_*, d_A) \to (B_*, d_B)$ be chain maps of chain complexes of R-modules. If f_* and g_* are chain homotopic, then the maps $\overline{f}_*$ and $\overline{g}_*$ induced in homology are equal, i.e.,*

$$\overline{f}_n = \overline{g}_n : H_n(A) \to H_n(B)$$

for all n. In particular, (A_*, d_A) is acyclic if there a chain homotopy between 1_{A_*} : $(A_*, d_*) \to (A_*, d_*)$ and the zero map on (A_*, d_*).

Proof. By definition, $\mathrm{im}(f_n - g_n)$ are boundaries. $\square$

In later sections, we will be given an R-module M and will have a method for constructing a chain complex for M that we shall want to be independent of the construction up to chain homotopy. This will lead to homology groups that will be independent of the construction.

Of course, analogous results of all the above hold for cochains and cohomology. We leave it to the reader to write the obvious results down. It is also a good exercise to prove these results (at least to see if you can keep the notation and maps correct).

Exercises 122.10. 1. Fill in the details of the proof of the Snake Lemma 122.1.

2. (Nine Lemma) Suppose the following diagram is commutative and all its columns are exact.

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & A' \arrow[r, "f'"] \arrow[d, "\alpha'"] & B' \arrow[r, "g'"] \arrow[d, "\beta'"] & C' \arrow[r] \arrow[d, "\gamma'"] & 0 \\
0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & A'' \arrow[r, "f''"] \arrow[d] & B'' \arrow[r, "g''"] \arrow[d] & C'' \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd}$$

Show all of the following:
   - (i) If the bottom two rows are exact, then so is the top row.
   - (ii) If the top two rows are exact, so is the bottom row.
   - (iii) If the top and bottom rows are exact, and the middle sequence is a zero sequence, then the middle row is exact.

3. Let
$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$
be a short exact sequence of chain complexes. Show if two of the chain complexes are exact, then so is the third.

4. Verify Computation 122.6.

123. Hom

We want to define homology that arises in the study of R-modules. To do so, we first look at the set of homomorphisms from one R-module to another.

Let R be a ring and M, N be R-modules. [The case for right R-modules has the obvious notational modifications.] Then $\mathrm{Hom}_R(M, N)$ is an abelian group. If R is commutative, then it is an R-module by Example 38.6(9).

We view at $\mathrm{Hom}_R(\quad,\quad)$ as a "function" of two variables (called a *functor* in category theory). We start with $\mathrm{Hom}_R(M,\quad)$, i.e., fix the first variable.

Let $h : A \to B$ be an R-homomorphism of R-modules. For each R-module M, define

$$h_* : \mathrm{Hom}_R(M, A) \to \mathrm{Hom}_R(M, B) \text{ by } f \mapsto h \circ f.$$

Then h_* is an abelian group homomorphism and an R-homomorphism if R is commutative.

We have the following 'naturality" of Hom that we leave as an easy exercise.

Lemma 123.1. *(*Naturality of Hom*) Let*

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ g\downarrow & & \downarrow h \\ A' & \xrightarrow[f']{} & B' \end{array}$$

be a commutative diagram of R-modules and R-homomorphisms. Then for all R-modules M,

$$\begin{array}{ccc} \mathrm{Hom}_R(M, A) & \xrightarrow{f_*} & \mathrm{Hom}_R(M, B)) \\ g_*\downarrow & & \downarrow h_* \\ \mathrm{Hom}_R(M, A') & \xrightarrow[f'_*]{} & \mathrm{Hom}_R(M, B') \end{array}$$

is a commutative diagram of abelian groups (and of R-modules if R is commutative).

A key fact about Hom is the following:

Proposition 123.2. *Let*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

be a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(M, A) \xrightarrow{f_*} \mathrm{Hom}_R(M, B) \xrightarrow{g_*} \mathrm{Hom}_R(M, C)$$

is an exact sequence of abelian groups (R-modules if R is commutative).

Note the missing 0 on the right.

Proof. Certainly, $g_* f_* = (gf)_*$, so the sequence is a zero sequence. Suppose that $h \in \mathrm{Hom}_R(M, A)$ satisfies $f_*(h) = 0$. Then $f\big(h(x)\big) = 0$ for all $x \in M$. As f is monic, $h(x) = 0$ for all $x \in M$, i.e., $h = 0$ and f_* is monic.

We show exactness at $\mathrm{Hom}_R(M, B)$, i.e., $\ker g_* \subset \operatorname{im} f_*$. Let $h \in \ker g_*$. Then $gh(y) = 0$, i.e., $h(y) \in \ker g$. Hence for each $y \in M$, there exists an $x \in A$ satisfying $f(x) = h(y)$ by exactness. Since f is monic, x is unique. So this defines a function $k : M \to A$ such that $k(x) = h(y)$. Since h and f are R-homomorphisms so is k. Hence $f_*(k) = h$. $\square$

Remark 123.3. The proof above shows that the result holds even if we do not assume that g is onto.

We say that $\mathrm{Hom}_R(M,\quad)$ is *left exact* (for all R-modules M).

Proposition 123.4. *Let M and A_i, $i \in I$, be R-modules. Denote the projections maps of $\prod_I A_i$ by $\pi_j : \prod_I A_i \to A_j$, $j \in I$. Then the map*

$$\varphi_I : \mathrm{Hom}_R(M, \prod_I A_i) \to \prod_I \mathrm{Hom}_R(M, A_i)$$

given by $h \mapsto (\pi_i h)_I$ is an isomorphism of abelian groups (and R-modules if R is commutative).

Proof. By the Universal Property of Direct Products (cf. Exercise 38.18(10)), given $h_i : M \to A_i$, for all $i \in I$, there exists a unique map $h : M \to \prod_I A_i$ such that $h(m) = \big(h_i(m)\big)_I$. In particular φ_I is a well-defined surjection. It is easily checked that φ_I is an abelian group homomorphism (and R-homomorphism if R is commutative). That φ is injective follows since $\varphi_I(h) = \varphi_I(h')$ if and only if $\pi_i h(m) = \pi_i h'(m)$ for all $m \in M$. Hence φ_I is an isomorphism. $\square$

We also have a "naturality" result, viz., the commutative diagram given by the next result.

Proposition 123.5. *Let A_i, $i \in I$, and B_j, $j \in J$ be R-modules. Suppose for each $i \in I$, there exist a $j \in J$ and an R-homomorphism $f_{ij} : A_i \to B_j$. Then for all R-modules M, we have a commutative diagram of abelian groups*

$$\begin{tikzcd}
\mathrm{Hom}_R(M, \prod_I A_i) \arrow[r, "f_*"] \arrow[d, "\varphi_I"'] & \mathrm{Hom}_R(M, \prod_J B_j) \arrow[d, "\varphi_J"] \\
\prod_I \mathrm{Hom}_R(M, A_i) \arrow[r, "\widetilde{f}"'] & \prod_J \mathrm{Hom}_R(M, B_j)
\end{tikzcd}$$

(of R-modules if R is commutative), where φ_I and φ_J are the isomorphism in Proposition 123.4, $f : \prod_I A_i \to \prod_J B_j$ by $(a_i)_I \mapsto \big(f_{ij}(a_i)\big)_J$ and $\widetilde{f} : (h_i)_I \mapsto (f_{ij} h_i)_J$.

Proof. Let $\pi_i : \prod_I A_k \to A_i$, $i \in I$, and $\pi'_j : \prod_J B_k \to B_j$, $j \in J$, be the projections. For all $h \in \mathrm{Hom}_R(M, \prod_I A_i)$, we have

$$\varphi_J f_*(h) = (\pi'_j f h)_J \text{ and } \widetilde{f} \varphi_I(h) = \widetilde{f}(\pi_i h)_I$$

Let $m \in M$ and $h(m) = (a_i)_I$. Then $\pi_i h(m) = a_i$. Consequently,

$$\pi'_j h(m) = f_{ij} h(m) = f_{ij}(a_i) \text{ and } \pi'_j(\widetilde{f} h)(m) = \pi'_j (f_{ij} h)_J(m) = f_{ij} a_i.$$

$\square$

Next, we turn to $\mathrm{Hom}_R(\ \ , N)$. Let $h : A \to B$ is an R-homomorphism of R-modules. For each R-module N, define

$$h^* : \mathrm{Hom}_R(B, N) \to \mathrm{Hom}_R(A, N) \text{ by } f \mapsto f \circ h.$$

Then h^* is an abelian group homomorphism and an R-homomorphism if R is commutative. Note if $A \xrightarrow{f} B \xrightarrow{g} C$ is a sequence of R-modules that $(gf)^* = f^* g^*$. This reverses arrows in the following analogues above and whose proofs we leave as exercises.

Lemma 123.6. (Naturality of Hom) *Let*

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle h} \\ A' & \xrightarrow[f']{} & B' \end{array}$$

be a commutative diagram of R-modules and R-homomorphisms. Then for all R-modules N,

$$\begin{array}{ccc} \mathrm{Hom}_R(B,N) & \xrightarrow{f^*} & \mathrm{Hom}_R(A,N)) \\ {\scriptstyle h^*}\uparrow & & \uparrow{\scriptstyle g^*} \\ \mathrm{Hom}_R(B',N) & \xrightarrow[f'^*]{} & \mathrm{Hom}_R(A',N) \end{array}$$

is a commutative diagram of abelian groups (and of R-modules if R is commutative).

Proposition 123.7. *Let*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

be a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(C,N) \xrightarrow{g^*} \mathrm{Hom}_R(B,N) \xrightarrow{f^*} \mathrm{Hom}_R(A,N)$$

is an exact sequence of abelian groups (R-modules if R is commutative).

Note the missing 0 on the right.

We say that that $\mathrm{Hom}_R(\ \ ,N)$ is *left exact*.

Proposition 123.8. *Let A_i, $i \in I$, and N be R-modules. Denote the injection maps into $\coprod_I A_i$ by $\iota_j : A_j \to \coprod_I A_i$ for all $j \in I$. Then*

$$\psi_I : \mathrm{Hom}_R(\coprod_I A_i, N) \to \prod_I \mathrm{Hom}_R(A_i, N)$$

by $h \mapsto (h\iota_i)_I$ is an isomorphism of abelian groups (and R-modules if R is commutative).

[Recall that if I is finite, then the coproduct and product are equal.]

As before, we have a "naturality" result.

Proposition 123.9. *Let A_i, $i \in I$, and B_j, $j \in J$, be R-modules. Suppose for each $j \in J$, there exist an R-homomorphism $f_{ji} : B_j \to A_i$. Then for all R-modules N, we have a commutative diagram of abelian groups*

$$\begin{array}{ccc} \mathrm{Hom}_R(\coprod_J B_j, N) & \xleftarrow{g^*} & \mathrm{Hom}_R(\coprod_I A_i, N) \\ {\scriptstyle \psi_J}\downarrow & & \downarrow{\scriptstyle \psi_I} \\ \prod_J \mathrm{Hom}_R(B_j, N) & \xleftarrow[\widetilde{g}]{} & \prod_I \mathrm{Hom}_R(A_i, N) \end{array}$$

(of R-modules if R is commutative), where ψ_I, ψ_J are the isomorphisms given by Proposition 123.8, $g : \coprod_J B_j \to \coprod_I A_i$ by $(b_j)_J \mapsto (g_{ji}b_j)_I$ and $\widetilde{g} : (h_j)_J \mapsto (h_j g_{ij})_I$.

Exercises 123.10. 1. Prove Lemma 123.1

2. Prove Lemma 123.6
3. Prove Proposition 123.7.
4. Prove Proposition 123.8.
5. Prove Proposition 123.9.
6. Let M and N be R-modules. Show that $\mathrm{Hom}_R(M,\ )$ and $\mathrm{Hom}_R(\ ,N)$ take split exact sequences to split exact sequences.

124. Injective Modules

We now apply our homological constructions to a chain complex that arises in Module Theory and whose generalization is very useful in algebraic geometry. Therefore, instead of generalizing the concept of free modules (that we shall do Section §126), we use a type of module that arises from the notion of divisible groups that we studied before and will be applicable to studying $\mathrm{Hom}_R(\ ,N)$ of the last section.

Free modules have the property that given any R-module M, there is a free R-module F mapping epimorphically onto it. We are interested in a "dual" notion of this property of R-free modules, i.e., to find a collection of R-modules Q such that any R-module M embeds into one of them. We also want this collection to satisfy the property that for such a Q, any short exact sequence $0 \to Q \xrightarrow{f} B \xrightarrow{g} C \to 0$ of R-modules splits, i.e., there exists an R-epimorphism $h : B \to Q$ satisfying $hg = 1_Q$. This property is one shared by a direct summand of a module (and is the true dual of what we want). This will lead to a long exact sequence in cohomology associated to short exact sequences $0 \to M' \to M \to M'' \to 0$ of R-modules. One difficulty of doing this case first is that we have no easy examples of this type of module. But as we shall see, they they do arise from the study of divisible abelian groups.

Definition 124.1. An R-module Q is called an *injective* R-module if for any R-monomorphism $f : A \to B$ and R-homomorphism $g : A \to Q$, there exists an R-homomorphism $h : B \to Q$ such that the diagram

(124.2)
$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, "h"] \\ Q & \end{tikzcd}$$

commutes.

An easy consequence of the definition is the following:

Lemma 124.3. *Let Q be an R-module. Then Q is an R-injective if and only if, whenever*

$$0 \to A \xrightarrow{f} B \xrightarrow{k} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(C,Q) \xrightarrow{k^*} \mathrm{Hom}_R(B,Q) \xrightarrow{f^*} \mathrm{Hom}_R(A,Q) \to 0$$

is exact.

Proof. By Proposition 123.7, $\mathrm{Hom}_R(\quad, Q)$ is left exact for all R-modules Q, so it suffices to show that f^* is a surjective. If $g \in \mathrm{Hom}_R(A,Q)$, then by equation (124.2), $f^*(h) = hf = g$ if and only if one of the conditions holds. $\square$

We say that $\mathrm{Hom}_R(\quad, N)$ is *exact* if $\mathrm{Hom}_R(\quad, N)$ takes short exact sequences to short exact sequences by "Hom_R"ing them. So the lemma says $\mathrm{Hom}_R(\quad, N)$ is exact if and only if N is an injective R-module.

Corollary 124.4. *Let A be an injective R-module. Then any short exact sequence of R-modules $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ splits. In particular, if A is a submodule of M, then there exists an submodule M' of M such that $M = A \oplus M'$.*

Proof. If A is injective, then given the exact sequence

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "1_A"'] & B \arrow[r, "g"] & C \arrow[r] & 0 \\ & A, & & & \end{tikzcd}$$

there exists an R-homomorphism $h : B \to A$ with $hf = 1_A$ showing that f is a split monomorphism. So the sequence splits by Exercise 38.18(11). $\square$

Lemma 124.5. *Let Q_i, $i \in I$, be R-modules. Then $\prod_I Q_i$ is R-injective if and only Q_i is R-injective for all $i \in I$. In particular, if I is finite, then $\coprod_I Q_i$ is R-injective if and only Q_i is R-injective for all $i \in I$ (since then the coproduct and product of the Q_I's are the same).*

Proof. Let $Q = \prod_I Q_i$. Set $\pi_i : Q \to Q_i$ to be the projection map and $\iota_i : Q_i \to \coprod Q_i$ the (natural) injection into a direct sum for all $i \in I$. So $\pi_i \iota_i = 1_{Q_i}$ for all $i \in I$. Suppose that Q is an injective R-module. Consider the diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, dashed, "h_i"] \arrow[d, dashed, "h"] \\ & Q_i \arrow[r, "\iota_i"'] & Q \end{tikzcd}$$

with f monic. The R-homomorphism $h : A \to Q$ such that $hf = \iota_i g$ exists as Q is injective. Let $h_i = \pi_i h : B \to Q_i$. We have $g = 1_Q g = \pi_i \iota_i g = \pi_i h f = h_i f$. Therefore, Q_i is an injective R-module for all $i \in I$.

Suppose that Q_i is an injective R-module for all $i \in I$. Consider the diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, dashed, "h"] \arrow[d, dashed, "h_i"] \\ & Q \arrow[r, "\pi_i"'] & Q_i \end{tikzcd}$$

for all $i \in I$ with f monic. The R-homomorphism $h_i = \pi_i g : B \to Q_i$ exists for all $i \in I$, as Q_i is an injective R-module. By the Universal Property of Direct Products (Exercise 38.18(9)), there exists an R-homomorphism $h : B \to Q$ such that $hf = g$. This shows that Q is an injective R-module. $\square$

Corollary 124.6. *Let Q be an injective R-module. Then any direct summand of Q is an injective R-module.*

We next prove a necessary and sufficient condition for an R-module to be R-injective. The proof is similar to that we gave for Proposition 28.10, the case of divisible abelian groups.

Theorem 124.7. (Baer Criterion) *Let Q be an R-module. Then Q is an injective R-module if and only if given any left ideal $\mathfrak{A}$ in R and an R-homomorphism $p : \mathfrak{A} \to Q$, there exists an R-homomorphism $q : R \to Q$ such that the diagram*

$$\begin{tikzcd} \mathfrak{A} \arrow[r, hook, "\mathrm{inc}"] \arrow[d, "p"'] & R \arrow[dl, "q"] \\ Q & \end{tikzcd}$$

commutes where inc *is the inclusion.*

Proof. Given a diagram

(124.8)
$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, dashed, "h"] \\ & Q & \end{tikzcd}$$

with exact row, we must construct h so that the diagram commutes. Let

$$\mathcal{S} = \{(A', g') \mid A \subset A' \subset B,\ g' : A' \to Q \text{ an } R\text{-homomorphism with } g_{A'}|_A = g\}.$$

Partially order $\mathcal{S}$ by

$$(A', g') \leq (A'', g'') \text{ if } A' \subset A'' \text{ is a submodule and } g''|_{A'} = g'.$$

The set $\mathcal{S}$ is not empty, since $(A, g) \in \mathcal{S}$. Let $\mathcal{C}$ be a chain in $\mathcal{S}$ and set $A_0 = \bigcup_{(A',g') \in \mathcal{C}} A'$. Define $\widetilde{g} : A_0 \to Q$ by $x \mapsto g'(x)$ if $x \in A'$. Since $\mathcal{C}$ is a chain, $\widetilde{g}$ is well-defined. Therefore, $\mathcal{C}$ has an upper bound in $\mathcal{S}$. By Zorn's Lemma, there exist a maximal element $(A_0, g_0) \in \mathcal{S}$.

Claim. $A_0 = B$:

Suppose the claim does not hold. Then there exists an $x \in B \setminus A_0$. Let $A' = A_0 + Rx$ and set $\mathfrak{A} = \{r \in R \mid rx \in A_0\}$, a left ideal in R. Define an R-homomorphism $p : \mathfrak{A} \to Q$ by $p(r) = g_0(rx) \in Q$ for all $r \in \mathfrak{A}$. By assumption, this extends to an R-homomorphism $q : R \to Q$. Define $g' : A' \to Q$ by $g'(a + sx) = g_0(a) + sq(1)$ for all $a \in A_0$ and $s \in R$. We must show that g' is well-defined. Suppose that $a + sx = a' + s'x$ with $a, a' \in A_0$ and $s, s' \in R$. Then we have $(s - s')x = a' - a$ lies in A_0, so $s - s' \in \mathfrak{A}$. We then have

$$\begin{aligned} g_0(a') - g_0(a) =& g_0\big((s - s')x\big) = q(s - s') \\ & q\big((s - s')1)\big) = (s - s')q(1), \end{aligned}$$

as q is an R-homomorphism. Therefore, g' well-defined and extends g_0. So $(A', g') \in \mathcal{S}$. As $(A_0, g_0) < (A', g')$, this is a contradiction. $\square$

A ring R is called *left Noetherian* (respectively, *right Noetherian*) if every left (respectively, right) ideal in R is finitely generated. The usual properties in the commutative case hold. In particular, every finitely generated R-module over a left Noetherian ring is a Noetherian R-module.

Corollary 124.9. *Let R be a left Noetherian ring and $\{Q_i\}_I$ a set of injective R-modules. The $\coprod_I Q_i$ is an injective R-module.*

Proof. By the Baer Criterion, it suffices to show that given any left ideal $\mathfrak{A}$ in R and R-homomorphism $p : \mathfrak{A} \to Q$, there exists an R-homomorphism $q : R \to Q$ such that the diagram

$$\begin{tikzcd} \mathfrak{A} \arrow[r, hook, "\text{inc}"] \arrow[d, "p"'] & R \arrow[dl, "q"] \\ \coprod_I Q_i & \end{tikzcd}$$

commutes. Since $\mathfrak{A}$ is finitely generated, there exists $J \subset I$ with J finite and $p(\mathfrak{A}) \subset \coprod_J Q_j \subset \coprod_I Q_i$. Since $\coprod_J Q_j = \prod_J Q_j$, p extends to $q : R \to \coprod_J Q_j$ by Lemma 124.5. The result follows. $\square$

The converse of this corollary is also true as we shall prove in Corollary 124.14 below.

We now show that any R-module can be embedded into an injective R-module. This is the key reason that injective R-modules are important. This will take a few steps. This is where we use our knowledge of (abelian) divisible groups. The statement of Proposition 28.10 in this new language says:

Lemma 124.10. *A divisible abelian group is an injective $\mathbb{Z}$-module.*

We will need the following remark to prove the next lemma.

Remark 124.11. Let R be a ring and D an abelian group. Then $\operatorname{Hom}_{\mathbb{Z}}(R, D)$ is a (left) R-module with R-action given by

$$(rf)(x) = f(xr) \text{ for all } f \in \operatorname{Hom}_{\mathbb{Z}}(R, D) \text{ and } r, x \in R.$$

Lemma 124.12. *Let D be a divisible abelian group. Then* $\operatorname{Hom}_{\mathbb{Z}}(R, D)$ *is an injective R-module.*

Proof. By the Baer Criterion 124.7, it suffices to show given

$$\begin{tikzcd} \mathfrak{A} \arrow[r, hook, "inc"] \arrow[d, "p"'] & R \\ \mathrm{Hom}_R(R, D) & \end{tikzcd}$$

with $\mathfrak{A}$ a left ideal in R that p extends to a group homomorphism $h : R \to \mathrm{Hom}_R(R, D)$ such that the resulting diagram commutes. Write $p_r = p(r)$ for $r \in R$.

Let $g : \mathfrak{A} \to D$ be defined by $g(a) = p_a(1_R)$ for $a \in \mathfrak{A}$. It is a group homomorphism. Since D is a divisible group, it is an injective $\mathbb{Z}$-module. In particular, there exists a group homomorphism $\widetilde{g} : R \to D$ such that $\widetilde{g}|_{\mathfrak{A}} = g$.

Define $h : R \to \mathrm{Hom}_{\mathbb{Z}}(R, D)$ by $r \mapsto h_r$, where $h_r(x) = \widetilde{g}(xr)$ for $x \in R$. (Check that h_r is a group homomorphism.) It follows easily that h is a group homomorphism. We show that h is an R-homomorphism. If s, r, x lie in R, we have

$$h_{sr}(x) = \widetilde{g}\big(x(sr)\big) = \widetilde{g}\big((xs)r\big) = h_r(xs),$$

and, by the R-structure on $\mathrm{Hom}_{\mathbb{Z}}(R, D)$,

$$h_r(xs) = sh_r(x) \text{ for all } r, s, x \in R.$$

It follows that h is an R-homomorphism. Now suppose that $r \in \mathfrak{A}$ and $x \in R$. Then $xr \in \mathfrak{A}$, so

$$h_r(x) = \widetilde{g}(xr) = g(xr) = p_{xr}(1_R)$$

Since $\mathrm{Hom}_{\mathbb{Z}}(R, D)$ is an R-module, we have

$$p_{xr}(1_R) = xp_r(1_R) = p_r(1_Rx) = p_r(x).$$

Consequently, $h_r = p_r$ for all $r \in \mathfrak{A}$, hence h extends p as needed. □

Proposition 124.13. *Let A be an R-module. Then there exists an injective R-module Q and an R-monomorphism $\varepsilon : A \to Q$.*

Proof. As A is an abelian group, we know by Theorem 15.17 that there exists a divisible group D and a group monomorphism $g : A \to D$. We know that $\mathrm{Hom}_{\mathbb{Z}}(R, D)$ is an injective R-module by Lemma 124.12. We have $g_* : \mathrm{Hom}_{\mathbb{Z}}(R, A) \to \mathrm{Hom}_{\mathbb{Z}}(R, D)$ is an R-monomorphism by Proposition 123.2 and $\mathrm{Hom}_R(R, A) \subset \mathrm{Hom}_{\mathbb{Z}}(R, A)$ is an R-submodule. Let $\rho : A \to \mathrm{Hom}_R(R, A)$ be given by $a \mapsto \rho_a$ where $\rho_a(r) = ra$. Then ρ is not only a $\mathbb{Z}$-monomorphism but also an R-monomorphism (in fact, an R-isomorphism), as

$$\rho_{sa}(r) = r(sa) = (rs)(a) = \rho_a(rs) = s\rho_a(r)$$

for all $r, s \in R$, $a \in A$, using the module structure on $\mathrm{Hom}_{\mathbb{Z}}(R, A)$. The composition

$$A \xrightarrow{\rho} \mathrm{Hom}_R(R, A) \xrightarrow{inc} \mathrm{Hom}_{\mathbb{Z}}(R, A) \xrightarrow{g_*} \mathrm{Hom}_{\mathbb{Z}}(R, D)$$

is an R-monomorphism. This proves the Proposition. □

We can now prove the converse of Corollary 124.9.

Corollary 124.14. *Let R be a ring in which every coproduct of injective R-modules is an injective R-module. Then R is a left Noetherian ring.*

Proof. Suppose that R is not left Noetherian. Then there exists an infinite chain of left ideals $\mathfrak{A}_1 < \mathfrak{A}_2 < \cdots$. Let $\mathfrak{A} = \cup_{i=1}^{\infty}\mathfrak{A}_i$. Then $\mathfrak{A} < R$. For each i, there exists an injective R-module Q_i satisfying $\mathfrak{A}/\mathfrak{A}_i \subset Q_i$ by Proposition 124.13. By hypothesis, $Q = \coprod_{i=1}^{\infty} Q_i$ is an injective R-module. Let $f_i : \mathfrak{A} \to \mathfrak{A}/\mathfrak{A}_i$ be the canonical epimorphism and $f : \mathfrak{A} \to \prod_{i=1}^{\infty} Q_i$ (the map defined by the Universal Property of Direct Products (Exercise 38.18(9)). For all $a \in \mathfrak{A}$, there exists an i such that $a \in \mathfrak{A}_i$, hence $f_n(a) = 0$ for all $n \geq i$. Therefore, $\operatorname{im} f \subset \coprod_{i=1}^{\infty} Q_i = Q$ and we may view $f : \mathfrak{A} \to Q$. Since Q is an injective R-module, there exists an extension of f to $g : R \to Q$ by the Baer Criterion 124.7. Let $g(1) = (x_1, x_2, \dots)$. Then there exists an N such that $x_i = 0$ for all $i \geq N$. If $a \in \mathfrak{A}$, we have $g(a) = ag(1) = (ax_1, ax_2, \dots)$ with $g_N(a) = 0$. But if $a \in \mathfrak{A}_{N+1} \setminus \mathfrak{A}_N$, then $g_N(a) \neq 0$, a contradiction. The result follows. $\square$

We can now also prove the converse of Corollary 124.4

Corollary 124.15. *Let A be an R-module. Then A is an injective R-module if and only if any short exact sequence of R-modules $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ splits.*

Proof. If $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ splits for any such exact sequence, then it does so for some injective R-module B by Proposition 124.13. But this means that $B \cong A \coprod C$. It follows that A is an injective R-module by Lemma 124.5. $\square$

Exercises 124.16. 1. Let Q be an injective R-module and (A_*, d_*) an acyclic chain complex of R-modules. Show that the chain complex $\big(\operatorname{Hom}_R\big((A_i, Q), (d_*)_*\big)\big)$ is acyclic.

2. Let R be a domain with K its quotient field. Use the Baer Criterion to show that K is R-injective. In particular, $\mathbb{Q}$ is $\mathbb{Z}$-injective.
3. Let R be a domain. Show every vector space over $qf(R)$ is an injective R-module.
4. Let R be a domain and M and R-module. We say that $x \in M$ is *divisible* by $r \in R$ if there exists a $y \in M$ satisfying $x = ry$. We say that M is a *divisible* R-module if every $x \in M$ is divisible by every $0 \neq r \in R$. Show all of the following:
   (i) $qf(R)$ is a divisible R-module.
   (ii) The direct sum and direct product of divisible R-modules are divisible. In particular, every vector space over $qf(F)$ is a divisible R-module
   (iii) Every quotient of a divisible R-module is divisible. In particular, every direct summand of a divisible R-module is divisible.
5. If R is a domain, show that every injective R module is a divisible R-module.
6. Let R be a domain and M an R-torsion-free R-module, i.e., if $rm = 0$, with $r \in R$ and $0 \neq m \in M$, then $r = 0$, (equivalently, $\lambda_r : M \to M$ given by $m \mapsto rm$ is injective for all nonzero $R \in R$. Show that an R-torsion-free R-module is a divisible R-module if and only if it is an injective R-module.
7. Let R be a PID and M an R-module. Show that M is an injective R-module if and only if it is divisible R-module. In particular, if R is a PID, then whenever M is a injective R-module, so is any quotient of M.

125. Ext

We now turn to developing a cohomology theory arising from short exact sequences of R-modules.

A cochain complex (A^*, d^*) is called *positive* if $A^n = 0$ for all $n < 0$. Let R be a ring and N an R-module. Let (A^*, d^*) be a positive cochain complex of R-modules. If there exists an R-monomorphism $\varepsilon : N \to A_0$ and

$$0 \to N \xrightarrow{\varepsilon} A^0 \xrightarrow{d^0} A^1 \xrightarrow{d^1} \cdots$$

is exact, we call this an *acyclic resolution of* of N with *augmentation* ε. We write this as $0 \to N \xrightarrow{\varepsilon} A^*$ is an acyclic resolution of N. If, in addition, A^i, is an injective R-module for all $i \geq 0$, we call the acyclic resolution $0 \to N \xrightarrow{\varepsilon} A^*$ an *injective resolution* of N.

Let $0 \to N \xrightarrow{\varepsilon} A^*$ be an injective resolution of N. We shall develop and investigate the cohomology of the cochain complex $\mathrm{Hom}_R(M, A^*)$ for M an R-module. Note that $d^0 : A^0 \to A^1$ is not a monomorphism. But also note that this will induce an injective resolution of $\operatorname{coker}\varepsilon$ given by $0 \to \operatorname{coker}\varepsilon \xrightarrow{\varepsilon^1} A^1 \xrightarrow{d_1} A^2 \xrightarrow{d^2} \cdots$ for the induced map ε^1. This will allow us to use induction.

Lemma 125.1. *Let M be an R-module. Then an injective resolution of M exists.*

PROOF. There exists an injective R-module I^0 and a R-monomorphism $\varepsilon : M \to I^0$ by Proposition 124.13. Let I^1 be an injective R-module such that the induced map $\varepsilon^1 : \operatorname{coker}\varepsilon \to I^1$ is an R-monomorphism and I^2 be an injective R-module such that the induced map $\varepsilon_2 : \operatorname{coker}\varepsilon^1 \to I^2$ is an R-monomorphism. Then we have a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & M \arrow[r, "\varepsilon"] & I^0 \arrow[rr, dashed, "d^0"] \arrow[dr, "-"'] & & I^1 \arrow[rr, dashed, "d^1"] \arrow[dr, "-"'] & & I^2 \\
& & & \operatorname{coker}(\varepsilon) \arrow[ur, "\varepsilon^1"'] \arrow[dr] & & \operatorname{coker}(\varepsilon^1) \arrow[ur, "\varepsilon^2"'] \arrow[dr] & \\
& & 0 \arrow[ur] & & 0\,0 \arrow[ur] & & 0
\end{tikzcd}$$

with $^-$ the canonical R-epimorphisms and where d^0, d^1 are defined to be the obvious compositions. This yields an exact sequence. Continuing by induction, gives an injective resolution of M. $\square$

We want to use injective resolutions of modules to obtain a cohomology theory after we apply $\mathrm{Hom}_R(M, \quad)$ in an appropriate way. For such a theory to be useful, we need it to be independent of the injective resolution that we take. The key reason that this is true will follow from the the next result.

Theorem 125.2. (Comparison Theorem) *Let $f : M \to M'$ be an R-homomorphism and*

$$\begin{tikzcd}
0 \arrow[r] & M' \arrow[r, "\varphi"] \arrow[d, "f"'] & X^* \\
0 \arrow[r] & M \arrow[r, "\varepsilon"] & I^*
\end{tikzcd}$$

be a diagram of positive cochain complexes with the top complex exact and each I_n injective in the bottom complex. Then there exists a cochain homomorphism

$$f^* : (X^*, d_X^*) \to (I^*, d_I^*)$$

such that the diagram

$$\begin{tikzcd} 0 \arrow[r] & M' \arrow[r, "\varphi"] \arrow[d, "f"'] & X^* \arrow[d, "f^*"'] \\ 0 \arrow[r] & M \arrow[r, "\varepsilon"] & I^* \end{tikzcd}$$

commutes. Moreover, the cochain map f^* is unique up to cochain homotopy.

Proof. The map f^0 exists by the definition of I_0 being an injective R-module. We proceed by induction. Given the diagram

$$\begin{tikzcd} X^{n-1} \arrow[r, "d_X^{n-1}"] \arrow[d, "f^{n-1}"'] & X^n \arrow[r, "d_X^n"] \arrow[d, "f^n"'] & X^{n+1} \\ I^{n-1} \arrow[r, "d_I^{n-1}"] & I^n \arrow[r, "d_I^n"] & I^{n+1}, \end{tikzcd}$$

we have $d_I^n f^n d_X^{n-1} = d_I^n d_I^{n-1} f^{n-1} = 0$ by the commutativity of the diagram. Since $\ker d_X^n = \operatorname{im} d_X^{n-1}$, restricting f^n to $\operatorname{coker} d_X^{n-1}$, we can extend f^n to f^{n+1} as I^{n+1} is an injective R-module.

Next we have to show if we if we are given another cochain map that satisfies the theorem, i.e., a cochain homomorphism $g^* : (X^*, d_X^*) \to (I^*, d_I^*)$ such that

$$\begin{tikzcd} 0 \arrow[r] & M' \arrow[r, "\varphi'"] \arrow[d, "f"'] & X^* \arrow[d, "g^*"'] \\ 0 \arrow[r] & M \arrow[r, "\varepsilon"] & I^* \end{tikzcd}$$

commutes, then f^* and g^* are cochain homotopic. Therefore,we must construct an R-homomorphism $s^* : (X^*, d_X^*) \to (I^*, d_I^*)$ satisfying $s^{n+1} d_X^n + d_I^{n-1} s^n = f^n - g^n$.

Let $d_I^{-2} = 0 = d_X^{-2}$, $d_I^{-1} = \varepsilon$, and $d_X^{-1} = \varphi'$. Now let $f^{-1} = f = g^{-1}$ and $s^{-1} = 0$. Then in the diagram

$$\begin{tikzcd} 0 \arrow[r, "d_X^{-2}"] & M' \arrow[r, "d_X^{-1}"] \arrow[dl, "s^{-1}"'] \arrow[d, "f"'] & X^0 \arrow[r, "d_X^0"] \arrow[dl, "s^0"'] \arrow[d, shift right, "f^0"'] \arrow[d, shift left, "g^0"] & X^1 \arrow[r, "d_{X^1}"] \arrow[dl, "s^1"] & {} \\ 0 \arrow[r, "d_I^{-2}"'] & M \arrow[r, "d_I^{-1}"'] & I^0 \arrow[r, "d_I^0"'] & I^1 \arrow[r, "d_I^1"'] & {}, \end{tikzcd}$$

set $s^0 = 0$. Consequently, $f^{-1} - g^{-1} = f - f = 0 = d_I^{-2} s^{-1} + s^0 d_X^{-1}$. To finish, we proceed by induction. Consider the diagram

$$\begin{tikzcd}
X^{n-2} \arrow[r, "d_X^{n-2}"] \arrow[d, shift right, "f^{n-2}"'] \arrow[d, shift left, "g^{n-2}"] & X^{n-1} \arrow[r, "d_X^{n-1}"] \arrow[d, shift right, "f^{n-1}"'] \arrow[d, shift left, "g^{n-1}"] \arrow[dl, "s^{n-1}"'] & X^{n} \arrow[r, "d_X^{n}"] \arrow[d, shift right, "f^{n}"'] \arrow[d, shift left, "g^{n}"] \arrow[dl, "s^{n}"'] & X^{n+1} \arrow[dl, dashed, "s^{n+1}"] \\
I^{n-2} \arrow[r, "d_I^{n-2}"'] & I^{n-1} \arrow[r, "d_I^{n-1}"'] & I^{n} \arrow[r, "d_I^{n}"'] & I^{n+1}
\end{tikzcd}.$$

We must define s^{n+1}. We have

$$\begin{aligned}
(f^n - g^n - d_I^{n-1}s^n)d_X^{n-1} &= f^n d_X^{n-1} - g^n d_X^{n-1} - d_I^{n-1}s^n d_X^{n-1} \\
&= f^n d_X^{n-1} - g^n d_X^{n-1} - d_I^{n-1}s^n d_X^{n-1} - d_I^{n-1}d_I^{n-2}s^{n-1} \\
&= (f^n - g^n)d_X^{n-1} - d_I^{n-1}(f^{n-1} - g^{n-1}) = 0
\end{aligned}$$

by the given commutativity of the diagram and induction. Therefore, $f^n - g^n - d_I^{n-1}s^{n-1} : X^{n-1} \to I^n$ vanishes on $\operatorname{im} d_X^{n-1} = \ker d_X^n$. So we have a factorization

$$\begin{tikzcd}
X^n \arrow[r, "d_X^{n-1}"] \arrow[dr] & X^n/\ker(d_X^n) \arrow[r, hook] \arrow[d, "f^n - g^n - d_I^n s^n"] & X^{n+1} \\
& I^n &
\end{tikzcd}$$

As I^{n+1} is an injective R-module, we get an extension $s^{n+1} : X^{n+1} \to I^n$ satisfying $s^{n+1}d_X^n + d_I^{n-1}s^n = f^n - g^n$ as needed. □

ho

Construction 125.3. Let M be an R-module and N an R module with a given injective resolution $0 \to N \xrightarrow{\varepsilon} I_N^*$. [If $N = 0$, then $0 \to N \xrightarrow{\varepsilon} 0^*$ is an injective resolution of N.] Apply $\operatorname{Hom}_R(M, \quad)$ to the cochain $(I_N^*, d_{I_N}^*)$ (that is not usually exact at the 0th term.) to get a cochain complex $\big(\operatorname{Hom}_R(M, I_N^*), d_N^*\big)$ of abelian groups (of R-modules if R is commutative) with $d_N^n = d_{I_N^*}^n$ for all n. In general, this new cochain complex is not exact, so we take its cohomology $H^n(\operatorname{Hom}_R(M, I_N^*)$. If $0 \to N \xrightarrow{\varphi} I_N'^*$ is another injective resolution, then applying the Companion Theorem 125.2 to the map $1_N : N \to N$, we get two chain maps

$$f^* : (I_N^*, d_{I_N}^*) \to (I_N^*, d_{I_N'}^*) \text{ and } g^* : (I_N'^*, d_{I_N'}^*) \to (I_N^*, d_{I_N}^*)$$

whose compositions are cochain homotopy to the identity. This means when we do the above construction that they induce maps

$$\begin{aligned}
(f^*)_* :& \big(\operatorname{Hom}_R(M, I_N^*), \overline{d}_{I_N}^*\big) \to \big(\operatorname{Hom}_R(M, I_N'^*), \overline{d}_{I_N'}^*\big) \\
(g^*)_* :& \big(\operatorname{Hom}_R(M, I_N'^*), \overline{d}_{I_N'}^*\big) \to \big(\operatorname{Hom}_R(M, I_N^*), \overline{d}_{I_N}^*\big)
\end{aligned}$$

Notation 125.4. Since maps between cochains always have upper indices, but maps of Hom's have lower and upper indices depending on the variable fixes, we shall always write upper stars for maps in cohomology.

We see that these maps in turn induce maps

$$\begin{aligned}\overline{f}^n :& H^n\big(\operatorname{Hom}_R(M, I_N^*)\big) \to H^n\big(\operatorname{Hom}_R(M, I_N'^*)\big)\\ \overline{g}^n :& H^n\big(\operatorname{Hom}_R(M, I_N'^*)\big) \to H^n\big(\operatorname{Hom}_R(M, I_N^*)\big).\end{aligned}$$

And to simplify notation, we shall write these induced maps simply as f^n and g^n when no confusion arises.

Since the original maps where inverse to each other up to cochain homotopy, these maps are inverse group isomorphisms of abelian groups (of R-isomorphisms if R is commutative) in cohomology. Also note that in cohomology, the isomorphism is "natural", i.e., dependent only on N. We set

$$\operatorname{Ext}_R^n(M, N) := H^n(\operatorname{Hom}_R(M, I_N^*).$$

Properties 125.5. of $\operatorname{Ext}_R^n(\quad,\quad)$. Let M and N be R-modules. Then we have

1. $\operatorname{Ext}_R^n(M, N)$ is independent of an injective resolution $0 \to N \xrightarrow{\varepsilon} I_N^*$.
2. $\operatorname{Ext}_R^0(M, N) = \operatorname{Hom}_R(M, N)$.
3. Suppose that $f : N \to N'$ is an R-homomorphism. Then f induces an abelian group homomorphism (R-homomorphisms if R is commutative)
$$f^n : \operatorname{Ext}_R^n(M, N) \to \operatorname{Ext}_R^n(M, N')$$
depending only on f.
4. Let A_i, $i \in I$, and B_j, $j \in J$ be R-modules. Suppose for each $i \in I$, there exist an R-homomorphism $f_{ij} : A_i \to B_j$. Then for all R-modules M, we have a commutative diagram of abelian groups (R-homomorphisms if R is commutative)
$$\begin{array}{ccc}\operatorname{Ext}_R^n(M, \prod_I A_i) & \xrightarrow{f^*} & \operatorname{Ext}_R^n(M, \prod_J B_j)\\ \downarrow{\scriptstyle\varphi_I} & & \downarrow{\scriptstyle\varphi_J}\\ \prod_I \operatorname{Ext}_R^n(M, A_i) & \xrightarrow[\widetilde{f}]{} & \prod_J \operatorname{Ext}_R^n(M, B_j).\end{array}$$
where the maps are induced by the maps in Proposition 123.5.
5. Suppose that $g : M \to M'$ is an R-homomorphism. Then g induces an abelian group homomorphism (R-homomorphism if R is commutative)
$$g^n : \operatorname{Ext}_R^n(M', N) \to \operatorname{Ext}_R^n(M, N)$$
depending only on g.
6. Let A_i, $i \in I$, and B_j, $j \in J$, be R-modules. Suppose for each $j \in J$, there exist an R-homomorphism $f_{ji} : B_j \to A_i$. Then for all R-modules N, we have a commutative diagram of abelian groups (R-homomorphisms if R is commutative)

$$\begin{tikzcd}
\operatorname{Ext}^n_R(\coprod_J B_j, N) \arrow[d, "\psi_J"'] & \operatorname{Ext}^n_R(\coprod_I A_i, N) \arrow[l, "g^*"'] \arrow[d, "\psi_I"] \\
\prod_J \operatorname{Ext}^n_R(B_j, N) & \prod_I \operatorname{Ext}^n_R(A_i, N). \arrow[l, "\widetilde{g}"]
\end{tikzcd}$$

where the maps are induced by from the maps in Proposition 123.9. [If I is finite, we the coproduct and product are the same.]

Next we want to construct a long exact sequence for Ext given a short exact sequence of R-modules. There are two possibilities, depending on which variable that we fix when we apply Hom. The case when the second variable is fixed is the harder case. We start with it. We need to construct compatible injective resolutions for the modules given in a short exact sequence. This is the content of the next result.

Lemma 125.6. (Horseshoe Lemma) *Let*

$$0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$$

be a short exact sequence of R-modules and

$$0 \longrightarrow M' \xrightarrow{\varepsilon'} I^*_{M'} \quad \text{and} \quad 0 \longrightarrow M'' \xrightarrow{\varepsilon''} I^*_{M''}$$

injective resolutions. Then there exist an injective resolution

$$0 \longrightarrow M \xrightarrow{\varepsilon} I^*_M$$

such that

$$0 \longrightarrow I^*_{M'} \xrightarrow{f^*} I^*_M \xrightarrow{g^*} I^*_{M''} \longrightarrow 0$$

is a split exact sequence of injective cochain complexes, i.e., the diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & M' \arrow[r, "f"] \arrow[d, "\varepsilon'"'] & M \arrow[r, "g"] \arrow[d, "\varepsilon"'] & M'' \arrow[r] \arrow[d, "\varepsilon''"'] & 0 \\
0 \arrow[r] & I^0_{M'} \arrow[r, "f^0"] \arrow[d, "d^0_{M'}"'] & I^0_{M'} \arrow[r, "g^0"] \arrow[d, "d^0_M"'] & I^0_{M''} \arrow[r] \arrow[d, "d^0_{M''}"'] & 0 \\
0 \arrow[r] & I^1_{M'} \arrow[r, "f^1"] & I^1_M \arrow[r, "g^1"] & I^1_{M''} \arrow[r] & 0 \\
 & \vdots & \vdots & \vdots &
\end{tikzcd}$$

commutes and has exact columns and split exact rows for all $i \geq 0$ (i.e., except for the top row).

PROOF. Let $I_M^0 = I_{M'}^0 \prod I_{M''}^0 = I_{M'}^0 \coprod I_{M''}^0$, an injective R-module by Lemma 124.5. Let

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & M' \arrow[r, "f"] \arrow[d, "\varepsilon'"'] & M \arrow[r, "g"] \arrow[dl, "h_{M'}"] \arrow[d, dashed, "\varepsilon"'] \arrow[dr, "h_{M''}"'] & M'' \arrow[r] \arrow[d, "\varepsilon''"'] & 0 \\
0 \arrow[r] & I_{M'}^0 \arrow[r, "\iota^0_{I_{M'}}"'] & I_M^0 \arrow[r, "\pi^0_{I_{M''}}"'] & I_{M''}^0 \arrow[r] & 0
\end{tikzcd}$$

be the diagram where the bottom sequence is a split exact sequence of injective R-modules, $f^0 = \iota^0_{I_{M'}}$ the R-monomorphism and $g^0 = \pi^0_{I_{M''}}$ the R-projection given by the splitting. The map $h_{M'}$ is the lift of ε' that exists since f is a monomorphism and I_M^0 is injective and the map $h_{M''}$ is the composition $\varepsilon'' g$. Now define $\varepsilon : M \to I_M^0$ by $\varepsilon(x) = \big(h_{M'}(x), h_{M''}(x)\big)$. By the Snake Lemma 122.1, the map ε is a monomorphism. The Snake Lemma also leads to a commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & M' \arrow[r, "f"] \arrow[d, "\varepsilon'"'] & M \arrow[r, "g"] \arrow[d, "\varepsilon"'] & M'' \arrow[r] \arrow[d, "\varepsilon''"'] & 0 \\
0 \arrow[r] & I_{M'}^0 \arrow[r, "f^0"] \arrow[d, "-"'] & I_M^0 \arrow[r, "g^0"] \arrow[d, "-"'] & I_{M''}^0 \arrow[r] \arrow[d, "-"'] & 0 \\
0 \arrow[r] & \operatorname{coker}(d_{M'}^0) \arrow[r, "\overline{f}^0"] \arrow[d] & \operatorname{coker}(d_M^0) \arrow[r, "\overline{g}^0"] \arrow[d] & \operatorname{coker}(d_{M''}^0) \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0. &
\end{tikzcd}$$

with exact columns and with the top two rows are exact (with the usual induced maps). Therefore, the bottom row is also exact by the Nine Lemma (Exercise 122.10(2)). As we also have the commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & \operatorname{coker}(d^0_{M'}) \arrow[r, "\overline{f}^0"] \arrow[d, "d^1_{M'}"'] & \operatorname{coker}(d^0_M) \arrow[r, "\overline{g}^0"] & \operatorname{coker}(d^0_{M''}) \arrow[r] \arrow[d, "d^1_{M''}"'] & 0 \\
0 \arrow[r] & I^1_{M'} \arrow[r, "\iota^1_{I_{M'}}"'] & I^1_M \arrow[r, "\pi^1_{I_{M''}}"'] & I^1_{M''} \arrow[r] & 0
\end{tikzcd}$$

with $I^1_M = I^1_{M'} \coprod I^1_{M''}$, we can repeat the argument. Continuing gives the result. $\square$

Theorem 125.7. *Let N be an R-module and*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

a short exact sequence of R-modules and R-homomorphisms. Then the exact sequence

$$0 \to \operatorname{Hom}_R(M, A) \xrightarrow{f_*} \operatorname{Hom}_R(M, B) \xrightarrow{g_*} \operatorname{Hom}_R(M, C)$$

extends to a long exact sequence in cohomology

$$\cdots \to \operatorname{Ext}^{n-1}_R(M, C) \xrightarrow{\partial^{n-1}} \operatorname{Ext}^n_R(M, A) \xrightarrow{\overline{f}^n} \operatorname{Ext}^n_R(M, B)$$
$$\xrightarrow{\overline{g}^n} \operatorname{Ext}^n_R(M, C) \xrightarrow{\partial^n} \operatorname{Ext}^{n+1}_R(M, A) \to \cdots$$

Proof. By the Horseshoe Lemma 125.6, we have a commutative exact diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "\varepsilon'"'] & B \arrow[r, "g"] \arrow[d, "\varepsilon"'] & C \arrow[r] \arrow[d, "\varepsilon''"'] & 0 \\
0 \arrow[r] & I^*_A \arrow[r, "f^*"] & I^*_B \arrow[r, "g^*"] & I^*_C \arrow[r] & 0
\end{tikzcd}$$

with the columns injective resolutions. We know that

$$0 \longrightarrow I^*_A \xrightarrow{f_*} I^*_B \xrightarrow{g_*} I^*_C \longrightarrow 0$$

is split exact as the I^i_A are injective. Since $\operatorname{Hom}_R(M, \ )$ takes split exact sequences to split exact sequences, we get an exact sequence of cochains

$$0 \to \operatorname{Hom}_R(M, I_{A^*}) \xrightarrow{g_{**}} \operatorname{Hom}_R(M, I_{B^*}) \xrightarrow{f_{**}} \operatorname{Hom}_R(M, I_{C^*}) \to 0.$$

Taking the long exact sequence of this short exact sequence of cochain complexes yields the result by Theorem 122.4 $\square$

By Theorem 125.7 and Theorem 122.7, we also have the following:

Theorem 125.8. *Let M be an R-module and*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{f} & B & \xrightarrow{g} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta_*} & & \downarrow{\scriptstyle\gamma} & & \\ 0 & \longrightarrow & A' & \xrightarrow{f} & B & \xrightarrow{g} & C'' & \longrightarrow & 0 \end{array}$$

an exact sequence of R-modules. Then there exists a commutative diagram

$$\begin{array}{ccc} \operatorname{Ext}^n_R(M,C) & \xrightarrow{\partial^n} & \operatorname{Ext}^{n+1}_R(M,A) \\ \downarrow{\scriptstyle\gamma_n} & & \downarrow{\scriptstyle\alpha_{n+1}} \\ \operatorname{Ext}^n_R(M,C') & \xrightarrow{\partial'^n} & \operatorname{Ext}^{n+1}_R(M,A'). \end{array}$$

We also have a long exact sequence in the second variable and the naturality using $\operatorname{Hom}_R(\quad, I^i)$ is exact when $0 \to N \to I^*$ is an injective resolution. The details are left to the reader.

Theorem 125.9. *Let N be an R-module. If*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms. Then the exact sequence

$$0 \to \operatorname{Hom}_R(C,N) \xrightarrow{g^*} \operatorname{Hom}_R(B,N) \xrightarrow{f^*} \operatorname{Hom}_R(A,N)$$

extends to a long exact sequence in cohomology

$$\cdots \to \operatorname{Ext}^{n-1}_R(A,N) \xrightarrow{\partial^{n-1}} \operatorname{Ext}^n_R(C,N) \xrightarrow{\overline{g}^n} \operatorname{Ext}^n_R(B,N)$$
$$\xrightarrow{\overline{f}^n} \operatorname{Ext}^n_R(A,N) \xrightarrow{\partial^n} \operatorname{Ext}^{n+1}_R(C,N) \to \cdots$$

Corollary 125.10. *Let N be an injective R-module. Then $\operatorname{Ext}^n_R(M,N) = 0$ for all R-modules M and all $n > 0$.*

Proof. The exact sequence $0 \to N \xrightarrow{1_N} N \to 0$ is an an injective resolution of N. $\square$

Theorem 125.11. *Let N be an R-module and*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{f} & B & \xrightarrow{g} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\alpha^*} & & \downarrow{\scriptstyle\beta^*} & & \downarrow{\scriptstyle\gamma^*} & & \\ 0 & \longrightarrow & A' & \xrightarrow{f} & B' & \xrightarrow{g} & C' & \longrightarrow & 0 \end{array}$$

an exact sequence of R-modules Then there exists a commutative diagram

$$\begin{array}{ccc} \operatorname{Ext}^n_R(A,N) & \xrightarrow{\partial^n} & \operatorname{Ext}^{n+1}_R(C,N) \\ \downarrow{\scriptstyle\alpha^n} & & \downarrow{\scriptstyle\gamma^{n+1}} \\ \operatorname{Ext}^n_R(A',N) & \xrightarrow{\partial'^n} & \operatorname{Ext}^{n+1}_R(C',N). \end{array}$$

We will now show how to "shift dimension" of the cohomology. To do this, we wish to show that if two R-modules N and N' satisfy $\mathrm{Ext}^n_R(M,N) \cong \mathrm{Ext}^n_R(M,N')$ for all $n > 0$ (i.e., have the same cohomology for all $n > 0$) if they are *injective equivalent*, i.e., there exist injective R-modules I and J satisfying $N \coprod I \cong N' \coprod J$. This is clearly an equivalence relation on the collection of R-modules. (Unfortunately, this collection is not a set.) Let $[N]$ denote the injective equivalence class of N in the collection of R-modules. So we have $[N] = [0]$ if and only if N is an injective R-module. The equivalence classes $\{[N] \mid N \text{ an } R\text{-module}\}$ looks like a semi-group with identity $[0]$ with addition defined by $[M] + [N] := [M \coprod N]$.

It is convenient to define the module analog of a free product of groups with amalgamation.

Definition 125.12. Suppose that we are given a diagram of R-modules and R-homomorphisms

$$(*) \qquad \begin{tikzcd} M \arrow[r, "f_1"] \arrow[d, "f_2"'] & N_1 \\ N_2 \end{tikzcd}$$

Then (X, g_1, g_2), with X an R-module and g_i an R homomorphism for $i = 1, 2$, is called the *pushout* or *cofiber product* of (*) if we have a commutative diagram

$$(\dagger) \qquad \begin{tikzcd} M \arrow[r, "f_1"] \arrow[d, "f_2"'] & N_1 \arrow[d, "g_1"] \\ N_2 \arrow[r, "g_2"'] & X. \end{tikzcd}$$

and if (Y, h_1, h_2) is another such triple satisfying $(\dagger)$, then there exists an R-homomorphism $\alpha : X \to Y$, unique up to isomorphism, satisfying the following commutative diagram

$$\begin{tikzcd} M \arrow[r, "f_1"] \arrow[d, "f_2"'] & N_1 \arrow[d, "g_1"'] \arrow[ddr, "h_1"] & \\ N_2 \arrow[r, "g_2"] \arrow[drr, "h_2"'] & X \arrow[dr, "\alpha"] & \\ & & Y. \end{tikzcd}$$

We leave it as an exercise to show the pushout exists.

The key is the following lemma.

Lemma 125.13. (Schanuel's Lemma) *Suppose that*

$$0 \to M \xrightarrow{f_1} I_1 \xrightarrow{g_1} N_1 \to 0 \quad \textit{and} \quad 0 \to M \xrightarrow{f_2} I_2 \xrightarrow{g_2} N_2 \to 0$$

are two short exact sequences of R-modules with I_1 and I_2 injective R-modules. Then $N_1 \coprod I_2 \cong N_2 \coprod I_1$, In particular, $[N_1] = [N_2]$.

Proof. Let $I = I_1 \coprod I_2 = I_2 \prod I_2$ and $i_i : I_i \to I$ be the map $x \to (x, 0)$ for $i = 1, 2$. Let

$$X = \{(x_1, x_2) \mid x_i \in I_i,\ i = 1, 2, \text{ with } f_1(x) = f_2(x)\} \subset I.$$

Then

$$\text{(125.14)} \qquad \begin{array}{ccc} M & \xrightarrow{f_1} & I_1 \\ {\scriptstyle f_2}\downarrow & & \downarrow{\scriptstyle \iota_1} \\ I_2 & \underset{\iota_2}{\longrightarrow} & X \end{array} \quad \text{is the pushout of} \quad \begin{array}{ccc} M & \xrightarrow{f_1} & I_1 \\ {\scriptstyle f_2}\downarrow & & \\ I_2. & & \end{array}$$

We have $\operatorname{coker} \iota_1 = \{(x_1, x_2) \in X \mid f_2(x_2) = 0\} \subset I_2$. Therefore, $\operatorname{coker} \iota_1 \cong N_2$. Similarly, $\operatorname{coker} \iota_2 \cong N_1$. The two short exact sequences

$$0 \to I_1 \to X \to \operatorname{coker} \iota_2 \to 0$$
$$0 \to I_2 \to X \to \operatorname{coker} \iota_1 \to 0.$$

split, as I_i is an injective R-module for $i = 1, 2$. The result follows. $\square$

Definition 125.15. Define the *injective shift operator* $\mathcal{I}$ on the collection of R-modules by $\mathcal{I}(M) := [N]$ if there exists an exact sequence

$$0 \to M \xrightarrow{f} I \xrightarrow{g} N \to 0$$

with I an injective R-module. This is well-defined by Schanuel's Lemma and $\mathcal{I}(M)$ only depends on $[M]$. In addition, $\mathcal{I}(M_1 \coprod M_2) = \mathcal{I}(M_1) + \mathcal{I}(M_2)$. Let $\mathcal{I}^0(M) = [0]$ and $\mathcal{I}^n = \mathcal{I}(\mathcal{I}^{n-1})$. In particular, if $0 \to M \to I_*$ is an injective resolution, then $\mathcal{I}^n(M) = [\operatorname{coker}(d_n)]$.

Lemma 125.16. *Let N be an R-module. then the following are equivalent:*

(1) N is an injective R-module.
(2) $\operatorname{Ext}^n_R(M, N) = 0$ for all $n \geq 1$ and all R-modules M.
(3) $\operatorname{Ext}^1_R(M, N) = 0$ for all R-modules M.
(4) $\operatorname{Ext}^1_R(M, N) = 0$ for all cyclic R-modules M.

Proof. (1) $\Rightarrow$ (2): Let $I_0 = N$ and $I_n = 0$ for $n > 0$. Then the sequence $0 \to N \xrightarrow{1_N} N \to 0 \to 0 \cdots$ is an injective resolution. Therefore, $0 \to \operatorname{Hom}_R(M, I_0) \to \operatorname{Hom}_R(M, I_1) \to \operatorname{Hom}_R(M, I_2) \to \cdots$ is just $0 \to \operatorname{Hom}_R(M, I_0) \to 0 \to 0 \to \cdots$ which has $H_n(M, N) = 0$ for $n > 0$.

(2) $\Rightarrow$ (3) $\Rightarrow$ (4) are immediate.

(4) $\Rightarrow$ (1): We show N is an injective R-module using the Baer Criterion 124.7. Let $0 \to \mathfrak{A} \xrightarrow{\text{inc}} R \xrightarrow{f} M \to 0$ be an exact sequence of R-modules. Then M is a cyclic R-module. Taking the long exact sequence in the first variable Theorem 125.7 yields the result in view of Theorem 125.16

$$0 \to \operatorname{Hom}_R(M, N) \to \operatorname{Hom}_R(R, N) \xrightarrow{f^*} \operatorname{Hom}_R(\mathfrak{A}, N) \to \operatorname{Ext}^1_R(M, N).$$

As $\operatorname{Ext}^1_R(M, N) = 0$ by hypothesis, $\operatorname{Hom}_R(R, N) \xrightarrow{f^*} \operatorname{Hom}_R(\mathfrak{A}, N)$ is surjective as needed. $\square$

Corollary 125.17. *Let M and N be R-modules. Then $\operatorname{Ext}_R^n(M,N)$ depends only on $[N]$ for $n \geq 1$.*

Proof. If I is an injective R-module, then $\operatorname{Ext}_R^n(M, N \coprod I) = \operatorname{Ext}_R^n(M,N)$ $\square$

Abusing notation, we shall write $\big(\operatorname{Ext}_R^n(M, \mathcal{I}(N)\big)$ for $\operatorname{Ext}_R^n(M,N')$ if $\mathcal{I}(N) = [N']$.

Theorem 125.18. (Dimension Shifting) *Suppose that M and N are R-modules. Then $\operatorname{Ext}_R^n\big(M,\mathcal{I}(N)\big) = \operatorname{Ext}_R^{n+1}(M,N)$ for all $n \geq 1$. In particular, we have $\operatorname{Ext}_R^{n+1}(M,N) = \operatorname{Ext}_R^1\big(M,\mathcal{I}^n(N)\big)$ for all $n \geq 1$.*

Proof. Let $0 \to N \xrightarrow{\varepsilon} I_*$ be an injective resolution. Then we have exact sequences

$$0 \to \operatorname{coker} \varepsilon \to I^0 \to \operatorname{coker} d^0 \to 0$$

and

$$0 \to \operatorname{coker} d^n \to I^{n+1} \to \operatorname{coker} d^{n+1} \to 0$$

for $n > 0$. Taking the long exact sequences in Theorem 125.9 of these yields the result in view of Lemma 125.16 and Corollary 125.17. $\square$

Looking at lengths of injective resolutions leads to the study of a cohomological dimension of a ring. Schanual's Theorem allows us to inductively compute the shortest length of an injective R-resolution of an R-module. We indicate this.

Definition 125.19. If M is an R-module, define the *left injective dimension* of M by

$$\operatorname{lid}_R(M) = \min\{n \mid \mathcal{I}^n(M) = 0\}$$

(or infinity if no minimum exists) and the *left global injective dimension* of R to be

$$\operatorname{lgl\,inj\,dim}(R) = \max\{\operatorname{lid}_R(M) \mid M \text{ an } R\text{-module}\}$$

(or infinity if no maximum exists).

Of course, we also have right injective dimension of right R-modules and right global injective dimension $\operatorname{rid}(R)$. For non commutative rings $\operatorname{lgl\,inj\,dim} R)$ and rgl inj $\dim(R)$ may be different.

Corollary 125.20. *Let N an R-module. Then the following are equivalent:*

(1) $\operatorname{lid}_R(N) \leq n$.
(2) $\mathcal{I}^n(N)$ *is an injective R-module.*
(3) $\operatorname{Ext}^1\big(M,\mathcal{I}(N)\big) = 0$ *for all R-modules M.*
(4) $\operatorname{Ext}_R^{n+1}(M,N) = 0$ *for all $i > 0$ and all R-modules M.*

Corollary 125.21. *The following are equivalent:*

(1) $\operatorname{lgl\,inj\,dim}(R) \leq n$.
(2) $\operatorname{Ext}_R^{n+1}(M,N) = 0$ *for all R-modules M and N.*
(3) $\operatorname{Ext}_R^{n+i}(M,N) = 0$ *for all $i > 0$ and R-modules M and N.*
(4) $\operatorname{lgl\,inj\,dim}(R) = \sup\{\operatorname{id}_R(N) \mid N$ *a cyclic R-module*$\}$.

Exercises 125.22. 1. Prove the properties in Properties 125.5.

2. Prove Theorem 125.7

3. Prove Theorem 128.9.
4. Prove that if
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
is a split exact sequence of R-modules, then
$$0 \to \operatorname{Ext}^n_R(M, A) \xrightarrow{\overline{f}^n} \operatorname{Ext}^n_R(M, B) \xrightarrow{\overline{g}^n} \operatorname{Ext}^n_R(M, C) \to 0$$
is exact for all n.
5. Prove that if
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
is a split exact sequence of R-modules, then
$$0 \to \operatorname{Ext}^n_R(C, N) \xrightarrow{\overline{g}^n} \operatorname{Ext}^n_R(B, N) \xrightarrow{\overline{f}^n} \operatorname{Ext}^n_R(A, N) \to 0$$
is exact for all n.
6. Prove that the pushout in Definition 125.12 exists.
7. Establish equation (125.14).
8. Let R be a commutative Noetherian ring and M an R-module. For every prime ideal $\mathfrak{p}$ in R, suppose that $\operatorname{Ext}^1_R(R/\mathfrak{p}, M) = 0$. Prove that M is an injective R-module.

126. Projective Modules

We now generalize the notion of free modules. This will result in in many statements that are dual to those that we did for injective R-modules, i.e., by reversing the arrows. Therefore, we shall not prove results where that is all that is necessary.

Definition 126.1. Let P be an R-module. We call P a *projective R-module* or *R-projective* if given any R-epimorphism $f : B \to C$ and an R-homomorphism $g : P \to C$, then there exists an R-homomorphism $h : P \to B$ such that the diagram

$$\begin{tikzcd} & P \arrow[dl, "h"'] \arrow[d, "g"] \\ B \arrow[r, "f"'] & C \end{tikzcd}$$

commutes.

Example 126.2. Every free R-module P is projective:
Let $\mathcal{B}$ be a basis for P and

$$\begin{tikzcd} & P \arrow[dl, dashed, "h"'] \arrow[d, "g"] \\ B \arrow[r, "f"'] & C \end{tikzcd}$$

a diagram with f surjective. Let $a_i \in B$ satisfy $g(a_i) = f(x_i)$ for each $x_i \in \mathcal{B}$. By the Universal Property of Free Modules 39.3, there exists an $h : P \to B$ such that $h(x_i) = a_i$ for all $x_i \in \mathcal{B}$. So P is R-projective.

Lemma 126.3. *A direct summand of an R-free module is projective and a direct sum of R-modules is projective if and only if each direct summand of it is R-projective.*

Example 126.4. The $\mathbb{Z}/6\mathbb{Z}$-module $\mathbb{Z}/2\mathbb{Z}$ is a $\mathbb{Z}/6\mathbb{Z}$ projective but not is not $\mathbb{Z}/6\mathbb{Z}$-free.

Lemma 126.5. *Let P be an R-module. Then P is a projective R-module if and only if, whenever*

$$(*) \qquad 0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \operatorname{Hom}_R(P, A) \xrightarrow{f_*} \operatorname{Hom}_R(P, B) \xrightarrow{g_*} \operatorname{Hom}_R(P, C) \to 0$$

is exact. In particular, if C is R-projective, then $(*)$ is split exact.

Corollary 126.6. *Let C be an R-module. Then C is R-projective if and only if any short exact sequence of R-modules of the form*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

splits.

Proof. The proof of the converse of splitting in the projective case is easier than the injective case as any direct summand of a free module is projective. The rest is as before. □

Example 126.7. Let K be field and $R = K[t]/(t^2)$, the ring of dual numbers over K. The ring R is the image of the ring epimorphism $\overline{} : K[t] \to R$ by $t^2 \mapsto 0$. The ideal $\mathfrak{m} = R\bar{t}$ is the unique prime ideal in R, so R is a Noetherian local ring. K is a field, but as an R-module it is not even R-projective, since the short exact sequence $0 \to R\bar{t} \to R \to K \to 0$ of R-modules does not split. It also follows that the submodule $R\bar{t}$ of R is not R-projective, as the exact sequence $0 \to K \to R \to R\bar{t} \to 0$ does not split. In particular, a submodule of a projective module need not be projective. A special case of this is $K = \mathbb{Z}/2\mathbb{Z}$.

Although projective modules are not necessarily free, they do satisfy a weaker version of the existence of a basis. In particular, they have a generating set that gives rise to coordinate functions just as in the free case.

Proposition 126.8. (Projective Basis) *Let M be an R-module. Then M is R-projective if and only if there exist sets*

(1) $\{m_i \mid m_i \in M,\ i \in I\}$
(2) $\{f_i : M \to R \mid f_i \text{ an } R\text{-homomorphism},\ i \in I\}$

satisfying for each $m \in M$, $f_i(m) = 0$ for almost all $i \in I$ and $m = \sum_I f_i(m) m_i$. In particular, if this is the case, then $\{m_i \mid i \in I\}$ generates M.

Proof. ($\Rightarrow$): Let F be a free R-module with basis $\mathcal{B} = \{x_i\}_I$ such that $g : F \to P$ is an R-epimorphism and $f : P \to F$ a splitting of g given by Corollary 126.6. We have $f(m) = \sum_I f_i(m) x_i$ for unique $f_i(m)$ for all $i \in I$. As $\mathcal{B}$ is a basis for F, we must have

$f_i(m) = 0$ for almost all $i \in I$ and each $f_i : P \to R$ is R-linear. Since $gf = 1_P$, we also have $m = \sum_I f_i(m)g(x_i)$ for all i. Define $m_i = g(x_i)$ for each $i \in I$. Then this works.

($\Leftarrow$): Suppose that we are given the f_i and m_i, $i \in I$. Then $P = \sum_I Rm_i$. Let $F = \coprod_I Rx_i$ be a free R-module with basis $\mathcal{B} = \{x_i\}_I$ and set $g : F \to P$ to be the unique R-epimorphism defined by $x_i \mapsto m_i$ for all $i \in I$. Let $f : P \to F$ be defined by $f(m) = \sum_I f_i(m)x_i$. This makes sense as almost all $f_i(m) = 0$. The map f is an R-homomorphism, since all the f_i are. Moreover, $fg(m) = \sum_I f_i(m)g(x_i) = \sum_I f_i(m)m_i$ for all $m \in P$. Thus f splits and P is isomorphic to a direct summand of the R-free module F, hence is R-projective. $\square$

If P is a projective R-module, the $\{m_i \mid i \in I\}$ with $\{f_i \mid i \in I\}$ given by the proposition is called a *projective basis* for P.

We also have a test for the projectivity of R-modules that uses injective R-modules.

Proposition 126.9. *Let P be an R-module. Then P is a projective R-module if and only whenever we have a diagram*

$$\begin{tikzcd} & P \arrow[d, "g"] \arrow[dl, dashed, "f"'] \\ Q \arrow[r, "h"'] & C \end{tikzcd}$$

with Q an injective R-module and h an R-epimorphism, there exists an R-homomorphism $f : P \to Q$, such the diagram commutes.

Proof. Suppose that we have a diagram

$$\begin{tikzcd} & P \arrow[d, "g"] & \\ B \arrow[r, "h"'] & C \arrow[r] & 0. \end{tikzcd}$$

with h an R-epimorphism. Let $K = \ker h$ and assume that $C = B/K$. We know that there exists an injective R-module Q and an R-monomorphism $B \to Q$ by Proposition 124.13, which we may view as an inclusion as well as $C \subset Q/K$. So we have a commutative diagram

$$\begin{tikzcd} & P \arrow[d, "g"] \arrow[dl, dashed] & \\ B \arrow[r] \arrow[d, "inc"'] & B/K \arrow[r] \arrow[d, "inc"] & 0 \\ Q \arrow[r, "\pi"'] & Q/K \arrow[r] & 0 \end{tikzcd}$$

with π the canonical R-epimorphism. By assumption, there exists an R-homomorphism $f : P \to Q$ making the diagram commute. Let $x \in P$. By commutativity of the resulting diagram, $f(x) + K \in \operatorname{im}(inc \circ g)$. Since $g(x) \in C$, we have $\operatorname{im} f \subset \operatorname{im}(inc : B \to Q)$ by the commutativity of this diagram. Therefore, $f : P \to B$ and the diagram commutes. $\square$

We turn to those rings that satisfy the strong condition that any submodule of a projective module is projective. This special case is interesting for it includes the collection of Dedekind domain studied in Chapter XV. Indeed, we shall see here the beginnings of a generalization of the Fundamental Theorem of Finitely Generated Modules Over a PID and which we shall prove in the next section.

Definition 126.10. Let R be a ring. Then R is called *left hereditary* if every left ideal in R is a projective R-module.

A key result is the following theorem, which in the case of Dedekind domains helps generalize the decomposition of finitely generated modules over a Dedekind domain.

Theorem 126.11. (Kaplansky) *Let R be a left hereditary ring. If P is a submodule of a free R-module F, then P is isomorphic to a direct sum of left ideals. In particular, P is projective.*

Proof. Let $\mathcal{B} = \{x_i \mid i \in I\}$ be a basis for F. Using the Well-Ordering Principle (which is equivalent to Zorn's Lemma by Appendix A), we may assume that I is well-ordered. Let $F_0 = 0$ and for each $i \in I$, set

$$F_i := \bigoplus_{j<i} Rx_j \quad \text{and} \quad \overline{F}_i := \bigoplus_{j\leq i} Rx_j = Fx_j \oplus Rx_i.$$

Fix $i \in I$. Suppose that $m \in P \cap \overline{F}_i$. Then $m = m' + rx_i$ for some $m' \in P \cap F_i$ and $r \in R$. Since the restriction $f_i : P \cap \overline{F}_i \to R$ by $m \mapsto r$ of the coordinate map is an R-homomorphism, $r = f_i(m)$. Let $\mathfrak{A}_i = \operatorname{im} f_i$, a left ideal of R. By assumption $\mathfrak{A}_i$ is R-projective. In particular, the exact sequence

$$0 \to P \cap F_i \to P \cap \overline{F}_i \to \mathfrak{A}_i \to 0$$

splits with $P \cap \overline{F}_i = (P \cap F_i) \oplus N_i$ and $N_i \cong \mathfrak{A}_i$. In particular, $F_i \cap N_i = 0$.

Claim. $P = \bigoplus_I N_i$.

Clearly, $\sum_I N_i \subset P$. We first show that $\sum_I N_i = \bigoplus_I N_i$. Suppose that we have an equation $y_{i_1} + \cdots + y_{i_n} = 0$ with $y_{i_j} \in N_{i_j}$ and $i_1 < \cdots < i_n$. Since $F_{i_n} \cap N_{i_n} = 0$, and $y_{i_n} \in N_{i_n}$, we must have $y_{n_i} = 0$. By induction $y_{i_j} = 0$ for all $j = 1, \ldots, n$. Therefore, $\sum_I N_i = \bigoplus_I N_i$. So we need only show that $P = \sum_I N_i$. Assume this is not true. By well-ordering, there exists a least such i say j such that $(P \cap \overline{F}_j) \setminus \sum_I N_i$ is nonempty. Let $0 \neq y \in (P \cap \overline{F}_j) \setminus \sum_I N_i$. As $P \cap \overline{F}_j = (P \cap F_j) \bigoplus N_j$, there exist $y = y' + y''$ with $y' \in P \cap F_j$ and $y'' \in N_j$. Since $y' \in F_j$, it must have coordinate zero on all but finitely many $x_k \in \mathcal{B}$ with $k < j$. Let k_0 be the maximum of these. Then $k_0 < j$. Therefore, $y' \in P \cap \overline{F}_{k_0}$. Consequently, we have $y' \in \sum_I N_i$. Hence $y \in \sum_I N_i$, a contradiction. □

Corollary 126.12. *Let R be a PID. Then any submodule of a free R-module is free.*

Proof. Ideals in a PID are free of rank one. □

Corollary 126.13. *Let R be a PID. If M is an R-module generated by n elements then any submodule Of M can be generated by n elements.*

Using Proposition 126.9, we establish a nice characterization of hereditary rings.

Theorem 126.14. *Let R be a ring. Then the following are equivalent:*

(1) R is left hereditary
(2) Every quotient of an injective R-module is injective.
(3) Every submodule of a projective R-module is projective.

Proof. Statements (1) and (3) are equivalent by Kaplansky's Theorem 126.11 as every projective R-module is a direct summand of a free R-module.

(2) $\Rightarrow$ (3): Suppose that we have the following diagram with exact rows (with the dashed arrows to be defined) of R-modules and R-homomorphisms:

$$\begin{tikzcd}
P \arrow[d, dashed, "h"'] \arrow[dr, dashed, "\alpha"] & P' \arrow[l, "f"'] \arrow[d] & 0 \arrow[l] \\
I \arrow[r, "g"'] & I' \arrow[r] & 0
\end{tikzcd}$$

We must fill in the dashed arrows and obtain a commutative diagram. By Proposition 126.9, we may assume that I is an injective R-module. Therefore, by assumption I' is also an injective R-module. In particular, $\alpha : P \to I'$ exists. Since P is R-projective, $h : P \to I'$ exists. Therefore, the composition $hf : P' \to I$ shows that P' is R-projective.

(3) $\Rightarrow$ (2): Using the Baer Criterion 124.7, we use the similar proof as above except reversing arrows, i.e., we fill in the diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathfrak{A} \arrow[r] \arrow[d] & R \arrow[dl, dashed] \arrow[d, dashed] \\
0 & I' \arrow[l] & I \arrow[l]
\end{tikzcd}$$

As $\mathfrak{A}$ is R-projective, we have an R-homomorphism $\mathfrak{A} \to I$ such that the diagram commutes. As I is an injective R-module, this defines $R \to I$. The result now easily follows by the the Baer Criterion 124.7. $\square$

Corollary 126.15. *If R is a left hereditary ring, an R-module is projective if and only if it can be embedded into a free R-module.*

Exercises 126.16. 1. Show if $e \in R$ is an idempotent, i.e., $e^2 = 1$, then Re is a projective R-module.

2. Show that $\mathbb{Q}$ is not $\mathbb{Z}$-projective.
3. Show Example 126.4 does produce a non-free projective module.
4. Show that the sequence in Example 126.7 does not split.
5. Let K be a field and $R = K[t_1, t_2]$. Show that the ideal $Rt_1 + Rt_2 \subset R$ is not R-projective.

6. Let F be an R-free module on basis $\mathcal{B}$ and K a submodule of F generated by $X \subset K$. We say that an R-module M is *generated by* $\mathcal{B}$ with *relations* X if $M \cong F/K$. This is called a *presentation of* M (cf. with the case of groups). An R-module M is called *finitely presented* if there exists an exact sequence
$$R^m \to R^n \to M \to 0$$
of R-modules and R-homomorphisms. Show that an R-module M is finitely presented if and only if M has a presentation with a finite generating set with and finite generated set of relations.
7. Show the following:
   (i) If R is left Noetherian, then any finitely generated R-module is finitely presented.
   (ii) If P is a finitely generated projective R-module, then P is finitely presented.
8. Let M be an R-module and $M^* = \mathrm{Hom}_R(M, R)$ the *dual R-module* of M. Show that M^* is a right R-module and the map $M \to M^{**}$ by $x \mapsto e_x$ with $e_x(f) = f(x)$, the evaluation map, is a R-homomorphism.
9. if P is a nonzero projective R-module show that P^* is nonzero.
10. Let P be projective R-module. Show that the canonical map $P \to P^{**}$ is an R-monomorphism and an isomorphism if P is finitely generated.
11. (Eilenberg Swindle) Let P be a projective R-module. Then there exists a free R-module F such that $F \coprod P$ is R-free.

127. Projective Modules over Commutative Rings

We turn to the case that R is a commutative. We shall show that if R is a local ring, then any finitely generated projective R-module is in fact, free in Lemma 127.2 below. (The result is in fact true without the finitely generated hypothesis.) This is very useful in commutative algebra as we can use localization techniques since the localization of exact sequences of modules over a commutative ring is exact and localization takes split exact sequences to split exact sequence (cf. Exercise 92.31(6)). In particular, localization takes projective module to projective modules. In particular, if $\mathfrak{p}$ is a prime ideal in commutative R and M is an R-module, we can look at the $R_\mathfrak{p}$ module $M_\mathfrak{p} = \{rm \mid r \in R_\mathfrak{p},\, m \in M\}$, the *localization of M at* $\mathfrak{p}$. It will follow that if M is R-projective, then $M_\mathfrak{p}$ is $R_\mathfrak{p}$-projective, hence $R_\mathfrak{p}$-free. Moreover, we shall show in Theorem 127.4 that if R is a commutative Noetherian ring and M a finitely generated R-module, then M is a projective R-module if and only if $M_\mathfrak{m}$ is a free $R_\mathfrak{m}$-module for all maximal ideals $\mathfrak{m}$ in R (using Exercise 92.31(7)). In differential geometry, one studies vector bundles. This means that finitely generated projective modules are the algebraic analog of vector bundles on the euclidean pieces that glue to define the differential manifold with the prime ideals corresponding in the points on the euclidean pieces.

To prove this result, we need a few lemmas. The first (and its proof) was mentioned above.

Lemma 127.1. *Let R be a commutative ring and $S \subset R$ a multiplicative set with $0 \notin S$. Suppose that P is a projective R-module. Then $S^{-1}P$ is a projective $S^{-1}R$-module.*

Lemma 127.2. *Let $(R, \mathfrak{m})$ be a local ring. Then every finitely generated R-projective module is R-free.*

Proof. Let P be a finitely generated free R-module. By Corollary 93.12 of Nakayama's Lemma, we may assume P is generated by n elements with n the minimal number of generators for P. Let $f : R^n \to P$ be the R-epimorphism taking a basis of R^n to these n generators. As P is projective, f is a split R-epimorphism, say with splitting g. In particular, $R^n = \ker f \oplus g(P)$ with $g(P) \cong P$. The induced surjection $R^n/\mathfrak{m}R^n \to P/\mathfrak{m}P$ of $R/\mathfrak{m}$-vector spaces must be an isomorphism as they have the same dimension. It follows that $R^n = g(P) + \mathfrak{m}R^n$. By Corollary 93.11 of Nakayama's Lemma, $R^n = g(P)) \cong P$.

This induces a commutative diagram of $R/\mathfrak{m}$-vector spaces

$$\begin{tikzcd} g(P)/\mathfrak{m}g(P) \arrow[dr, "\overline{f}|_{g(P)}"] \arrow[dd, "\overline{inc}"'] & \\ & P/\mathfrak{m}P \\ R^n/\mathfrak{m}R^n. \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

We have $g(P)/\mathfrak{m}g(P) \to P/\mathfrak{m}(P)$ an isomorphism, as both are $R/\mathfrak{m}R$-vector spaces of the same rank n. Therefore, $R^n/\mathfrak{m}R^n \to P/\mathfrak{m}(P)$ is an isomorphism. It follows that $R^n = g(P) + \mathfrak{m}R^n$. By Nakayama's Lemma 93.10, $R^n = g(P)) \cong P$. □

In fact, it can be shown that the lemma holds without the condition that the module be finitely generated.

Lemma 127.3. *Let R be a commutative Noetherian ring and $S \subset R$ a multiplicative set. Suppose that M a finitely generated R-module and N is an arbitrary R-module such that there exists an $S^{-1}R$-homomorphism $\varphi : S^{-1}M \to S^{-1}N$. Then there exists an R-homomorphism $f : M \to N$ and an element $s \in S$ satisfying*

$$\varphi\big(\frac{x}{1}\big) = \frac{f(x)}{s}$$

for all $x \in M$.

Proof. Let $M = \sum_{j=1}^m Rx_j$, so $S^{-1}M = \sum_{j=1}^m S^{-1}R\big(\frac{x_j}{1}\big)$. Choose $y_j \in N$ such that

$$\varphi\big(\frac{x_j}{1}\big) = \frac{y_j}{s'} \text{ for } j = 1, \dots, m \text{ and } s' \in S.$$

(We can take a uniform s' as S is a multiplicative set.) As R is Noetherian and M finitely generated, there exists an R-epimorphism $h : R^m \to M$ induced by $e_j \mapsto x_j$ with $\{e_1, \dots, e_m\}$ the standard basis for R^m and satisfying $\ker h$ is finitely generated. In particular, there exists $a_{ij} \in R$ for $1 \leq i \leq n$, some $n \geq 1$, and $1 \leq j \leq m$, such that $\ker h$ is generated by $\{\sum_{j=1}^n a_{ij}e_j \mid i = 1, \dots, n\}$. As h induces an isomorphism $R^m/\ker h \to M$ and localization is exact, we have

$$\sum_{j=1}^m \big(\frac{a_{ij}}{1}\big)\big(\frac{x_j}{1}\big) = \Big(\frac{\sum_{j=1}^m a_{ij}x_j}{1}\Big) = 0$$

in $S^{-1}M$ for $i = 1, \dots, n$. Taking φ of this equation yields

$$\sum_{j=1}^{m} \big(\frac{a_{ij}}{1}\big)\big(\frac{y_j}{s'}\big) = 0$$

in $S^{-1}N$ for $i = 1, \dots, n$. Hence there exists $s'' \in S$ with

$$s'' \sum_{j_1}^{m} a_{ij} y_j = 0 \ \text{ in } \ N \ \text{ for all } \ i = 1, \dots, n.$$

Suppose that we have an equation $\sum_{j=1}^{m} b_j m_j = 0$ in M with $b_i \in R$, $j = 1, \dots, m$. Then there exist $c_1, \dots, c_m$ in R satisfying $b_j = \sum_{i=i}^{n} c_i a_{ij}$. It follows that $f : M \to N$ by $m_j \mapsto s'' y_j$, $j = 1, \dots, m$, is a well-defined R-homomorphism. Moreover, we have

$$\varphi(\frac{x_j}{1}) = \frac{y_j}{s'} = \frac{f(x_j)}{s's''}.$$

Setting $s = s's''$ shows that

$$\varphi(\frac{m}{1}) = \frac{f(m)}{s}$$

as needed. $\square$

Theorem 127.4. *Let R be a commutative Noetherian ring and M a finitely generated R-module. Then M is a projective R-module if and only if $M_{\mathfrak{m}}$ is a free $R_{\mathfrak{m}}$-module for all maximal ideals $\mathfrak{m} \subset R$.*

Proof. By Lemma 127.1, we know if M is a projective R-module so are all its localizations, hence $R_{\mathfrak{m}}$-free by Lemma 127.2. Therefore, we need only show the converse.

Let $\mathfrak{m}$ be a fixed maximal ideal in R. Then there exists a free R-module F and an exact sequence $F \xrightarrow{g} M \to 0$. This induces a split exact sequence $0 \to \ker g_{\mathfrak{m}} \to F_{\mathfrak{m}} \xrightarrow{g_{\mathfrak{m}}} M_{\mathfrak{m}} \to 0$ by Exercise 92.31(6). Let φ split $g_{\mathfrak{m}}$, so $g_{\mathfrak{m}}\varphi = 1_{M_{\mathfrak{m}}}$. By Lemma 127.3, there exists an R-homomorphism $f : M \to F$ (depending on $\mathfrak{m}$) and an element $s \in R \setminus \mathfrak{m}$ satisfying

$$\varphi(\frac{m}{1}) = \frac{f(m)}{s}$$

for all $m \in M$. In particular, $\dfrac{gf(m)}{1} = \dfrac{m}{1}$ for every $m \in M$. Let $M = \sum_{i=1}^{n} Rx_i$. For each x_i, $i = 1, \dots, n$, there exists a $c_i \in R \setminus \mathfrak{m}$ satisfying $c_i gf(x_i) = c_i s x_i$. Setting $c = c_1 \cdots c_n$. we have $c \in R \setminus \mathfrak{m}$ and $cgf(x_i) = csx_i$ for $i = 1, \dots, n$. Therefore, we have

$$(*) \qquad cgf(m) = csm \ \text{ for all } m \in M \text{ and } cs \notin \mathfrak{m}.$$

If $r \in R$, let $\lambda_r : M \to M$ be the R-homomorphism $m \mapsto rm$ and $\mathfrak{A}$ the set of elements $a \in R$ satisfying

(+)

$$\begin{tikzcd} & M \arrow[dl, "h"'] \arrow[d, "\lambda_a"] \\ F \arrow[r, "g"'] & M \end{tikzcd}$$

commutes for some $h : M \to F$. In particular, such an r exists with $r = cs$ for the above f, so $\mathfrak{A} \neq \emptyset$. Clearly, $\mathfrak{A}$ is an ideal in R and by (*), $\mathfrak{A} \cap (R \setminus \mathfrak{m}) \neq \emptyset$. Since this holds for each maximal ideal $\mathfrak{m}$ in R, we conclude that $1 \in \mathfrak{A}$. Therefore there exists a splitting $\sigma : M \to F$, i.e., $g\sigma = 1_M$. In particular, $\sigma(M)$ is a direct summand of F, hence R-projective. As $M \cong \sigma(M)$, M is R-projective as needed. □

Corollary 127.5. *Every Dedekind domain is a hereditary ring.*

Proof. If R is a Dedekind domain and $\mathfrak{p}$ a nonzero prime ideal, then $R_{\mathfrak{p}}$ is a local Dedekind domain. Therefore, $R_{\mathfrak{p}}$ is a PID by Lemma 87.6 so ideals in $R_{\mathfrak{p}}$ are $R_{\mathfrak{p}}$-free. The result follows by Theorem 127.4. □

We have generalized this corollary in Proposition 87.8. For convenience, we repeat it here with a different proof.

Theorem 127.6. *Let R be a domain with quotient field K. Then R is a hereditary ring if and only if every fractional ideal in R is invertible. In particular, if this is the case, then R is a Noetherian ring.*

Proof. ($\Leftarrow$): If $\mathfrak{A}$ is invertible, i.e., $\mathfrak{A}\mathfrak{A}^{-1} = R$, then there exists finitely many $a_i \in \mathfrak{A}$ and $q_i \in \mathfrak{A}^{-1}$, $i = 1, \dots, n$, some n, satisfying $1 = \sum_{i=1}^n q_i a_i$. with $q_i a_i \in R$ for $i = 1, \dots, m$. In particular, if $a \in \mathfrak{A}$, then $a = \sum_{i=1}^n (q_i a_i)a = \sum_{i=1}^r a(q_i a_i)$. Consequently, $\mathfrak{A}$ is a finitely generated R-module, so R is a Noetherian ring. Let $f_i = \lambda_{q_i} : \mathfrak{A} \to R$ be the R-homomorphism given by $a \mapsto q_i a$. Then $a = \sum_{i=1}^n f_i(a)a_i$ for all $a \in \mathfrak{A}$. In particular, $\{f_1, \dots, f_n\}$, $\{a_1, \dots, a_n\}$ form a projective basis for $\mathfrak{A}$ by Proposition 126.8, i.e., $\mathfrak{A}$ is R-projective.
($\Rightarrow$): Suppose that $\mathfrak{A}$ is a projective R-module. By Proposition 126.8, there exists a projective basis $\{f_i\}_I$, $\{a_i\}_I$ for $\mathfrak{A}$. In particular, the R-homomorphisms $f_i : \mathfrak{A} \to R$ satisfy $a = \sum_I f_i(a)a_i$ for all $a \in \mathfrak{A}$ with $f_i(a) = 0$ for almost all $i \in I$. For each $i \in I$, let $q_i = f_i(a)/a$ in K for $0 \neq a \in \mathfrak{A}$. The q_i's are independent of a. Indeed if $0 \neq a' \in \mathfrak{A}$, then
$$a' f_i(a) = f_i(a'a) = a f_i(a').$$
Therefore, $f_i = \lambda_{q_i}$ for $i \in I$. In particular, $q_i\mathfrak{A} \subset R$, hence $q_i \in \mathfrak{A}^{-1}$ for all $i \in I$. Since $a = \sum_I (q_i a)a_i$ in R for all $a \in \mathfrak{A}$, we have $1 = \sum_I q_i a_i$ in K hence in R. Therefore, $\mathfrak{A}\mathfrak{A}^{-1} = R$ and $\mathfrak{A}$ is invertible. □

Corollary 127.7. *Let R be a domain, not a field. Then R is a hereditary ring if and only if R is a Dedekind domain.*

Proof. We have seen that Dedekind domains are hereditary, so we may assume that R is hereditary. By Theorem 127.6 this is equivalent to every fractional ideal in R is invertible and, in particular, that R is Noetherian. Since R is a domain, $R = \bigcap_{\mathrm{Max}(R)} R_{\mathfrak{m}}$ by Exercise 29.4(8). Let $\mathfrak{m} \in \mathrm{Max}(R)$. Then every fractional ideal in $R_{\mathfrak{m}}$ is invertible. As the intersection of integrally closed domains is integrally closed (if $A = \bigcap_i B_i, x \in B_i$ integral for domains A, B_i, then $x \in x(\cap B_i) = \cap B_i = A$) and all fractional ideals in $R_{\mathfrak{m}}$ are invertible, for every $\mathfrak{m} \in \mathrm{Max}(R)$, it suffices to show that $R_{\mathfrak{m}}$ is a local Dedekind domain, i.e., a discrete valuation ring. In particular, we may assume that $R = (R, \mathfrak{m})$ is a local ring. Since R is hereditary, all ideals in R are R-projective, hence R-free by Lemma

127.2 as R is local. But a free submodule of R must be of rank 0 or 1 (as $xy - yx = 0$ in R for all $x, y \in R$). It follows that R must be a local PID (i.e., a discrete valuation ring) hence integrally closed. The result follows. $\square$

We turn to finitely generated modules over a Dedekind domain. We want to generalize the Fundamental Theorem of Finitely Generated Modules over a PID to the Dedekind case. In the PID case, we decomposed a finitely generated module into a direct sum of a free module and torsion module. In the case of a PID, a finitely generated free module was the same as a torsion-free module. This is not the true in general in the Dedekind case. We first shall show that it suffices to show that finitely generated R module over a Dedekind domain is a direct sum of torsion R-module and torsion-free R-modules. This allows us to study the torsion and torsion-free cases separately.

Lemma 127.8. *Let R be a Dedekind domain and $\mathfrak{A} \subset R$ an ideal. Then every ideal in $R/\mathfrak{A}$ is principal.*

PROOF. If $\mathfrak{A} = \mathfrak{p}_1^{e_1} \dots \mathfrak{p}_r^{e_r}$, then $S = R \setminus \bigcup_I \mathfrak{p}_i$ is a multiplicative set. This domain $S^{-1}R$ is a Dedekind domain that is semi-local with maximal ideals $\{S^{-1}\mathfrak{p}_1, \dots, S^{-1}\mathfrak{p}_r\}$. In particular, $S^{-1}R$ is a PID by Lemma 87.6. Since the ring homomorphism $R/\mathfrak{A} \to S^{-1}(R/\mathfrak{A})$ has $S \cap R = \emptyset$, the result follows. $\square$

Corollary 127.9. *Let R be a Dedekind domain with quotient field K and $\mathfrak{A}$ a fractional ideal of R. If $\mathfrak{B} \subset R$ is an ideal, then there exists $0 \neq y \in K$ that satisfies $R = y\,\mathfrak{A} + \mathfrak{B}$.*

PROOF. Let $0 \neq c \in K$ satisfy $c\,\mathfrak{A} \subset R$. Then $c\,\mathfrak{A}/c\,\mathfrak{A}\mathfrak{B}$ is an ideal in $R/c\,\mathfrak{A}\mathfrak{B}$. By the lemma, it is a principal ideal, say $c\,\mathfrak{A} = Rx + c\,\mathfrak{A}\mathfrak{B}$, $x \in R$. It follows that

$$R = (c\,\mathfrak{A})^{-1}(c\,\mathfrak{A}) = (c\,\mathfrak{A})^{-1}x + \mathfrak{B}.$$

Set $y = c^{-1}x$. Then replace $\mathfrak{A}$ by $\mathfrak{A}^{-1}$ in the above. The result follows. $\square$

Corollary 127.10. *Every ideal in a Dedekind domain can be generated by two elements and one of the generators can be chosen to be any nonzero element in $\mathfrak{A}$. Moreover, this generator a may be chosen such that (a) is relatively prime to any given principal ideal relatively prime to $\mathfrak{A}$.*

PROOF. Let $a \in \mathfrak{A}$ Then $\mathfrak{A}/(a)$ is a principal ideal. If $\mathfrak{A}/(a) = b\mathfrak{A}/(a)$, and $\mathfrak{A} = (a, b)$, Moreover, if (c) is relatively prime to $\mathfrak{A}$, then there exists an equation $cx + a = 1$, for some $x \in R$, $a \in \mathfrak{A}$ and we can start with this a. $\square$

This result does not generalize to Prüfer domain. This generalization of Dedekind domains in which every localization at a finitely generated prime ideal is a valuation ring is equivalent to domains that are semi-hereditary, i.e., every finitely generated ideal is projective. But there exist Prüfer domains having ideals minimally generated by n elements for any $n \geq 1$, although it has been shown that every finitely generated maximal ideal in a Prüfer domain can be generated by two elements.

Every ideal in a PID is free. This is not true in general. We can only say that every ideal in a domain R is R-torsion-free. The key result to the decomposition of finitely generated modules over a Dedekind domain is the following result (that is not true in general) is that the converse of Kaplansky's Theorem 126.11 is also true for Dedekind domains.

Proposition 127.11. *Let R be a Dedekind domain and M a finitely generated R-module. Then M is R-torsion-free if and only if M is a projective R-module.*

Proof. Let K be the quotient field of R. If M is projective, then it is a submodule of a free R-module. By Kaplansky's Theorem 126.11 and Corollary 127.7, M is R-torsion-free. So we need only prove the converse.

Suppose that M is a finitely generated R-torsion-free R-module. Let $\varphi : M \to K\otimes_R M$ be defined by $x \mapsto 1\otimes x$. As $K\otimes_R M$ is a finite dimensional K-vector space, it is isomorphic to K^r, some r. Let $\mathcal{B} = \{e_1, \dots, e_r\}$ be a basis for $K \otimes_R M$ and $M = \sum_{i=1}^n Ru_i$. Then

$$\varphi(u_i) = \sum_{i,j} \frac{a_{ij}}{b_{ij}} e_i \text{ with } a_{ij}, b_{ij} \in R, b_{ij} \neq 0 \text{ for all } i, j.$$

Set $0 \neq b = \prod_{i,j} b_{ij}$ in R. Then $bM \subset R^r$ is a submodule, so R-projective as R is hereditary by Corollary 127.7. It follows that M is also R-projective. □

Corollary 127.12. *Let R be a Dedekind domain and M a finitely generated R-module. Then $M \cong M_t \coprod M/M_t$.*

Proof. M/M_t is R-projective as it is R-torsion-free. Therefore, the canonical map $^- : M \to M/M_t$ splits. □

This shows that we can study the R-torsion-free and the R-torsion cases separately. We look at the R-torsion case.

Theorem 127.13. *Let R be a Dedekind domain and M a finitely generated torsion R-module. Then there exist ideals $\mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n$ unique up to isomorphism such that*

$$M \cong R/\mathfrak{A}_1 \coprod \cdots \coprod R/\mathfrak{A}_n.$$

Proof. Let $M = \sum_{i=1}^n Ru_i$ and $\mathfrak{A}_i = \operatorname{ann}_R u_i$ for $i = 1, \dots n$. Set $\mathfrak{A} = \mathfrak{A}_1 \cdots \mathfrak{A}_n$. We have $\mathfrak{A} = \operatorname{ann}_R(M)$. Let $\mathfrak{A} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ be a factorization of $\mathfrak{A}$ into prime ideals and $S = R \setminus \cup_{i=1}^r \mathfrak{p}_i$, a multiplicative set in R. It follows if $x \in S$, then $(x) + \mathfrak{A} = R$ by the definition of the greatest common divisor of ideals in a Dedekind domain. In particular, there exists $r \in R$ and $a \in \mathfrak{A}$ satisfying $rx + a = 1$. Therefore, for all $m \in M$, we have $m = rxm + am = rxm$, i.e., the R-homomorphism $\lambda_x : M \to M$ by $m \mapsto xm$ is an R-automorphism with inverse λ_r. Since $S^{-1}R$ is a semi-local with maximal ideals $S^{-1}\mathfrak{p}_1, \dots, S^{-1}\mathfrak{p}_r$, it is a PID by Lemma 87.6. Therefore, by the Fundamental Theorem of Finitely Generated Modules over a PID, $S^{-1}M \cong \coprod_{i=1}^r S^{-1}/(a_i)$ for some $a_1 \mid \cdots \mid a_r$ unique up to isomorphism. This lifts to a decomposition for M as desired. □

If R is a Dedekind domain and $\mathfrak{p}$ a prime ideal, let

$$M_{(\mathfrak{p})} := \{x \in M \mid \mathfrak{p}^n x = 0 \text{ for some } n > 0\}$$

called the $\mathfrak{p}$-*primary part* of M. An analogous proof of the Primary Decomposition Theorem (essentially the Chinese Remainder Theorem in the finitely generated case) yields:

Proposition 127.14. *If R is a Dedekind domain and M a torsion R-module, then $M = \bigoplus_{\mathfrak{p} \text{ a prime}} M_{(\mathfrak{p})}$. If M is finitely generated, then $M_{(\mathfrak{p})} = 0$ for almost all prime ideals $\mathfrak{p}$.*

Therefore, we can also decompose a R-torsion module over a commutative ring in this way also. In particular, in the Dedekind case for finitely generated R-torsion modules. We turn to the R-torsion-free case. We first need a general result.

Lemma 127.15. *Let R be a domain with quotient field K and $\mathfrak{A}$, $\mathfrak{B}$ fractional ideals of R. If $f : \mathfrak{A} \to \mathfrak{B}$ is an R-homomorphism, then there exists an element $c \in K$ such that $f(x) = cx$ for all $x \in \mathfrak{A}$. In particular, f is either the zero map or is a monomorphism.*

PROOF. Let $x \in \mathfrak{A}$. Then for all all $a \in \mathfrak{A}$, we have the equation

$$(xf)(a) = f(xa) = af(x).$$

Fix $a \neq 0$ in $\mathfrak{A}$ and set $c = a^{-1}f(a)$ in K. Then $f(x) = cx$ for all $x \in \mathfrak{A}$. □

Corollary 127.16. *Let R be a domain with quotient field K and $\mathfrak{A}$, $\mathfrak{B}$ fractional ideals of R. Then $\mathfrak{A} \cong \mathfrak{B}$ if and only if there exists an element $c \in K$ satisfying $\mathfrak{A} = c\,\mathfrak{B}$.*

Theorem 127.17. *Let R be a Dedekind domain and M a finitely generated R-torsion-free module. Then $M \cong R^r \coprod \mathfrak{A}$ for some ideal $\mathfrak{A} \subset R$ and $r \geq 0$ with*

$$R^r \coprod \mathfrak{A} \cong R^s \coprod \mathfrak{B} \text{ if and only if } r = s \text{ and } \mathfrak{A} \cong \mathfrak{B}.$$

In particular, if $\mathfrak{A}_1, \dots, \mathfrak{A}_m$ and $\mathfrak{B}_1, \dots, \mathfrak{B}_n$ are fractional ideals of R, then

$$\begin{gathered}\mathfrak{A}_1 \coprod \cdots \coprod \mathfrak{A}_m \cong \mathfrak{B}_1 \coprod \cdots \coprod \mathfrak{B}_n \text{ if and only if} \\ r = s \text{ and } \mathfrak{A}_1 \cdot \dots \cdot \mathfrak{A}_m \cong \mathfrak{B}_1 \cdots \mathfrak{B}_n.\end{gathered} \tag{127.18}$$

PROOF. We know that M is a direct sum of ideals in R by Kaplansky's Theorem 126.11 and Corollary 127.7. To finish, we need only show statement 127.18 holds. Suppose that we have an K-isomorphism $f : \mathfrak{A}_1 \coprod \cdots \coprod \mathfrak{A}_m \to \mathfrak{B}_1 \coprod \cdots \coprod \mathfrak{B}_n$. This isomorphism f can be viewed as a matrix $C = \big(c_{ij}\big) \in R^{n \times m}(K)$ such that $\mathfrak{B}_i = c_{ij}\mathfrak{A}_j$. Then f an isomorphism if and only if C is invertible. In particular, this shows that $m = n$. We next show that

$$\mathfrak{B}_1 \cdots \mathfrak{B}_m = \det(C)\mathfrak{A}_1 \cdots \mathfrak{A}_m. \tag{*}$$

Let $a_j \in \mathfrak{A}_j$. Then $c_{ij}a_j \in \mathfrak{B}_j$, hence $\det(C)\mathfrak{A}_1 \cdots \mathfrak{A}_m \subset \mathfrak{B}_1 \cdots \mathfrak{B}_m$. By symmetry we have the reverse inclusion. This shows (*). [Note that this argument only used that R was a domain.]

For the converse, we shall show that

$$\mathfrak{A}_1 \coprod \cdots \coprod \mathfrak{A}_m \cong R^{m-1} \coprod \mathfrak{A}_1 \cdots \mathfrak{A}_m.$$

Suppose that $m = 2$. We need to show if $\mathfrak{A} \coprod \mathfrak{B} \cong R \coprod \mathfrak{A}\mathfrak{B}$. Multiplying by a suitable nonzero element in K, we may assume the fractional ideal $\mathfrak{B}$ is an ideal in R. Consequently, by Corollary 127.9, we have $y\,\mathfrak{A} + \mathfrak{B} = R$ for some nonzero $y \in K$. Hence the fractional ideal $y\,\mathfrak{A}$ is also an ideal in R and the ideals $y\,\mathfrak{A}$ and $\mathfrak{B}$ are comaximal. By the Chinese Remainder Theorem, $y\,\mathfrak{A} \cap \mathfrak{B} = y\,\mathfrak{A}\mathfrak{B}$. Let $g : y\,\mathfrak{A} \coprod \mathfrak{B} \to R$ be the R-homomorphism given by $(ya, b) \mapsto ya - b$. We have $\ker g = y\,\mathfrak{A} \cap \mathfrak{B} = y\,\mathfrak{A}\mathfrak{B}$. Since the exact sequence $0 \to kerg \to y\,\mathfrak{A} \coprod \mathfrak{B} \xrightarrow{g} R \to 0$ splits and $y\,\mathfrak{A} \cong \mathfrak{A}$, equation (127.18) is established for the $n = 2$ case. By induction,

$$\mathfrak{A}_1 \coprod \cdots \coprod \mathfrak{A}_m \cong R \coprod \mathfrak{A}_1\mathfrak{A}_2 \coprod \mathfrak{A}_3 \coprod \cdots \coprod \mathfrak{A}_m \cong R^{m-1}\mathfrak{A}_1 \cdots \mathfrak{A}_m.$$

□

Corollary 127.19. *Let R be a Dedekind domain and M a nonzero finitely generated R-projective module. Then there exist ideals $\mathfrak{A}_1, \cdots, \mathfrak{A}_m$ in R such that $M \cong \mathfrak{A}_1 \coprod \cdots \coprod \mathfrak{A}_m$ with $m \geq 1$ unique and the class of $\mathfrak{A}_1 \cdots \mathfrak{A}_m$ in the ideal class group of R unique.*

Exercises 127.20. 1. Let $(R, \mathfrak{m})$ be a local ring and M a finitely presented R-module. Show the following are equivalent:
   (*i*) M is R-free.
   (*ii*) There exists a projective R-module P and an exact sequence of R-modules $0 \to N \xrightarrow{f} P \xrightarrow{g} M \to 0$ such that the induced map $\overline{f} : N/\mathfrak{m}N \to P/\mathfrak{m}P$ is a monomorphism.
2. Let R be a commutative ring and M a finitely generated R-module. Show that M is R-projective if and only if M is finitely presented and $M_{\mathfrak{p}}$ is R-free for all prime ideals $\mathfrak{p}$ in R.
3. Let R be a commutative ring and P a finitely generated projective R-module. If $\mathfrak{p}$ is a prime ideal in R, define the *rank of* P at $\mathfrak{p}$ to be $\dim_{R/\mathfrak{p}} P_{\mathfrak{p}}$. Show if the rank of $P_{\mathfrak{p}}$ is r, then there exists an element $f \in R$, such that the rank of $P_{\mathfrak{P}}$ is r for all $\mathfrak{P}$ in the open set $D(f) := \{\mathfrak{P} \mid f \notin \mathfrak{P}\}$ in $\operatorname{Spec} R$.
4. Prove that Lemma 127.3 holds only assuming that R a commutative ring. and M is a finitely presented R-module. (Cf. Exercise 126.16(6).)
5. Prove that Theorem 127.4 holds only assuming that R a commutative ring and M is a finitely presented R-module. (Cf. Exercise 126.16(6.)
6. Let R be a domain with quotient field K and $\mathfrak{A}$ an invertible fractional ideal of R. Show all of the following:
   (*i*) Let $\mathfrak{B}$ be a fractional ideal of R. Then the canonical map $\mathfrak{A} \otimes_R \mathfrak{B} \to \mathfrak{A}\mathfrak{B}$ is an R-isomorphism.
   (*ii*) $\mathfrak{A} \cong \operatorname{Hom}_R(\mathfrak{A}, R)$.
   (*iii*) $\mathfrak{A}$ is R-free if and only if $\mathfrak{A}$ is principal.
7. Let R be a UFD. Then an R-projective ideal in R is R-free if and only if it is principal.
8. Prove that a domain is a Dedekind domain if and only if every divisible R-module (cf. Exercise 124.16(4)) is an injective R-module.
9. A ring is called *left semi-hereditary* if every finitely generated left ideal is projective. Prove that any Prüfer domain is left hereditary.

128. Ext II

Let R be a ring and M an R-module and (A_*, d_*) be a positive chain complex of R-modules. Let $\varepsilon : A_0 \to M$ be an R-epimomorphism such that

$$\cdots \xrightarrow{d_1} A_1 \xrightarrow{d_0} A_0 \xrightarrow{\varepsilon} M \to 0$$

is exact. We call this an *acyclic resolution of* of M with *augmentation* ε. We write this as $A_* \xrightarrow{\varepsilon} M \to 0$ is an acyclic resolution of M. If (A_*, d_*) consists of projective R-modules, then we call such an acyclic resolution a *projective resolution* of M.

Lemma 128.1. *Let M be an R-module. Then a projective resolution of M exists.*

Proof. We know that there exists a free module P and an R-epimorphism $P \to M$. The result now follows as in the injective resolution case. $\square$

The theorems about injective resolutions and cohomology have analogues for projective resolutions. The proofs that are the same if amounting to reversing arrows and substituting projective modules for injective modules will be omitted.

Theorem 128.2. (Comparison Theorem) *Let $f : M \to M'$ be an R-homomorphism and*

$$\begin{array}{ccccc} P_* & \xrightarrow{\varepsilon} & M & \longrightarrow & 0 \\ & & \Big\downarrow{\scriptstyle f} & & \\ X_* & \xrightarrow{\varphi} & M' & \longrightarrow & 0 \end{array}$$

be a diagram of positive chain complexes with the bottom complex exact and each P_i projective in the top complex. Then there exists a chain homomorphism $f_* : (P_*, d_{P*}) \to (X_*, d_{X*})$ such that

The diagram

$$\begin{array}{ccccc} P_* & \xrightarrow{\varepsilon} & M & \longrightarrow & 0 \\ {\scriptstyle f_*}\Big\downarrow & & \Big\downarrow{\scriptstyle f} & & \\ X_* & \xrightarrow{\varphi} & M' & \longrightarrow & 0 \end{array}$$

commutes. Moreover, the chain map f_* is unique up to chain homotopy.

Theorem 128.3. *Let M be an R-module and $P_{M*} \xrightarrow{\varepsilon} M \to 0$ an R-projective resolution. If N is an R-module, then the cohomology of the chain complex $\big(\mathrm{Hom}_R(P_{M*}, N), d_{\mathrm{Hom}_R(P_{M*},N)}\big)$ is independent of the projective resolution of M and is denoted $\underline{\mathrm{Ext}}^*_R(M, N)$.*

We continue to use Notation 125.4 when talking about cohomology groups.

Properties 128.4. of $\underline{\mathrm{Ext}}^n_R(\quad,\quad)$. Let M and N be R-modules. Then we have

1. $\underline{\mathrm{Ext}}^n_R(M, N)$ is independent of a projective resolution $P_{M*} \xrightarrow{\varepsilon} M \to 0$.
2. $\underline{\mathrm{Ext}}^0_R(M, N) = \mathrm{Hom}_R(M, N)$.
3. Suppose that $f : N \to N'$ is an R-homomorphism. Then f induces an abelian group homomorphism (R-homomorphisms if R is commutative)

$$f^n : \underline{\mathrm{Ext}}^n_R(M, N) \to \underline{\mathrm{Ext}}^n_R(M, N')$$

depending only on f.

4. Let A_i, $i \in I$, and B_j, $j \in J$ be R-modules. Suppose for each $i \in I$, there exist an R-homomorphism $f_{ij} : A_i \to B_j$. Then for all R-modules M, we have a commutative diagram of abelian groups (R-homomorphisms if R is commutative)

$$\begin{array}{ccc}
\underline{\mathrm{Ext}}_R^n(M, \prod_I A_i) & \xrightarrow{f^*} & \underline{\mathrm{Ext}}_R^n(M, \prod_J B_j) \\
\varphi_I \downarrow & & \downarrow \varphi_J \\
\prod_I \underline{\mathrm{Ext}}_R^n(M, A_i) & \xrightarrow[\widetilde{f}]{} & \prod_J \underline{\mathrm{Ext}}_R^n(M, B_j).
\end{array}$$

where the maps are induced by from the maps in Proposition 123.5.

5. Suppose that $g : M \to M'$ is an R-homomorphism. Then g induces an abelian group homomorphism (R-homomorphism if R is commutative)

$$g^n : \underline{\mathrm{Ext}}_R^n(M', N) \to \underline{\mathrm{Ext}}_R^n(M, N)$$

depending only on g.

6. Let A_i, $i \in I$, and B_j, $j \in J$, be R-modules. Suppose for each $j \in J$, there exist an R-homomorphism $f_{ji} : B_j \to A_i$. Then for all R-modules N, we have a commutative diagram of abelian groups (R-homomorphisms if R is commutative)

$$\begin{array}{ccc}
\underline{\mathrm{Ext}}_R^n(\coprod_J B_j, N) & \xleftarrow{g^*} & \underline{\mathrm{Ext}}_R^n(\coprod_I A_i, N) \\
\psi_J \downarrow & & \downarrow \psi_I \\
\prod_J \underline{\mathrm{Ext}}_R^n(B_j, N) & \xleftarrow[\widetilde{g}]{} & \prod_I \underline{\mathrm{Ext}}_R^n(A_i, N).
\end{array}$$

where the maps are induced by from the maps in Proposition 123.9.

Lemma 128.5. (Horseshoe Lemma) *Let*

$$0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$$

be a short exact sequence of R-modules and

$$P_*{}' \xrightarrow{\varepsilon'} M' \longrightarrow 0 \quad \textit{and} \quad P_*{}'' \xrightarrow{\varepsilon'} M'' \longrightarrow 0$$

projective resolutions. Then there exist an R-projective resolution

$$P_* \xrightarrow{\varepsilon} M \longrightarrow 0$$

such that

$$0 \longrightarrow P_*{}' \xrightarrow{f_*} P_* \xrightarrow{g_*} P_*{}'' \longrightarrow 0$$

is a split exact sequence of projective chain complexes, i.e.,

$$\begin{tikzcd}
 & \vdots \arrow[d, phantom] & \vdots & \vdots & \\
0 \arrow[r] & {P_1}' \arrow[r, "f_1"] \arrow[d, "(d_{P'})_1"'] & P_1 \arrow[r, "g_1"] \arrow[d, "(d_P)_1"'] & {P_1}'' \arrow[r] \arrow[d, "(d_{P''})_1"'] & 0 \\
0 \arrow[r] & {P_0}' \arrow[r, "f_0"] \arrow[d, "\varepsilon'"'] & P_0 \arrow[r, "g_0"] \arrow[d, "\varepsilon"'] & {P_0}'' \arrow[r] \arrow[d, "\varepsilon''"'] & 0 \\
0 \arrow[r] & M' \arrow[r, "f"] \arrow[d] & M \arrow[r, "g"] \arrow[d] & M'' \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd}$$

commutes and has exact columns and split exact rows (except for the bottom row).

Theorem 128.6. *Let N be an R-module and*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

a short exact sequence of R-modules and R-homomorphisms. Then the exact sequence of abelian groups (respectively, R-homomorphisms if R is commutative)

$$0 \to \operatorname{Hom}_R(C, N) \xrightarrow{g^*} \operatorname{Hom}_R(B, N) \xrightarrow{f^*} \operatorname{Hom}_R(A, N)$$

extends to a long exact sequence of abelian groups (respectively R-modules if R is commutative) in cohomology

$$\cdots \to \underline{\operatorname{Ext}}_R^{n-1}(A, N) \xrightarrow{\partial^{n-1}} \underline{\operatorname{Ext}}_R^{n}(C, N) \xrightarrow{\overline{g}^n} \underline{\operatorname{Ext}}_R^{n}(B, N)$$
$$\xrightarrow{\overline{f}^n} \underline{\operatorname{Ext}}_R^{n}(A, N) \xrightarrow{\partial^{n}} \underline{\operatorname{Ext}}_R^{n+1}(C, N) \to \cdots$$

Theorem 128.7. *Let N be an R-module and*

$$\begin{tikzcd}
0 \arrow[r] & (A_*, d_A) \arrow[r, "f"] \arrow[d, "\alpha_*"'] & (B_*, d_B) \arrow[r, "g"] \arrow[d, "\beta_*"'] & (C_*, d_C) \arrow[r] \arrow[d, "\gamma_*"'] & 0 \\
0 \arrow[r] & (A'_*, d_{A'}) \arrow[r, "f'"] & (B'_*, d_{B'}) \arrow[r, "g'"] & (C'_*, d_{C'}) \arrow[r] & 0
\end{tikzcd}$$

an exact sequence of chain complexes. Then there exists a commutative diagram

$$\begin{tikzcd}
\underline{\operatorname{Ext}}_R^{n}(A, N) \arrow[r, "\partial^n"] \arrow[d, "\alpha_n"'] & \underline{\operatorname{Ext}}_R^{n+1}(C, N) \arrow[d, "\gamma_{n+1}"] \\
\underline{\operatorname{Ext}}_R^{n}(A', N) \arrow[r, "\partial'^n"] & \underline{\operatorname{Ext}}_R^{n+1}(C', N).
\end{tikzcd}$$

Theorem 128.8. *Let M be an R-module. If*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms. Then the exact sequence

$$0 \to \operatorname{Hom}_R(M, A) \xrightarrow{f_*} \operatorname{Hom}_R(M, B) \xrightarrow{g_*} \operatorname{Hom}_R(M, C)$$

extends to a long exact sequence in cohomology

$$\begin{aligned} \cdots \to \underline{\operatorname{Ext}}_R^{n-1}(M, C) \xrightarrow{\partial^{n-1}} \underline{\operatorname{Ext}}_R^{n}(M, A) \xrightarrow{\overline{g}^n} \underline{\operatorname{Ext}}_R^{n}(M, B) \\ \xrightarrow{\overline{f}^n} \underline{\operatorname{Ext}}_R^{n}(M, C) \xrightarrow{\partial^n} \underline{\operatorname{Ext}}_R^{n+1}(M, A) \to \cdots \end{aligned}$$

Theorem 128.9. *Let M be an R-module and*

$$\begin{tikzcd} 0 \arrow[r] & (A_*, d_A) \arrow[r, "f"] \arrow[d, "\alpha_*"'] & (B_*, d_B) \arrow[r, "g"] \arrow[d, "\beta_*"'] & (C_*, d_C) \arrow[r] \arrow[d, "\gamma_*"'] & 0 \\ 0 \arrow[r] & (A'_*, d_{A'}) \arrow[r, "f'"] & (B'_*, d_{B'}) \arrow[r, "g'"] & (C'_*, d_{C'}) \arrow[r] & 0 \end{tikzcd}$$

an exact sequence of chain complexes. Then there exists a commutative diagram

$$\begin{tikzcd} \underline{\operatorname{Ext}}_R^{n}(M, C) \arrow[r, "\partial^n"] \arrow[d, "\gamma_n"'] & \underline{\operatorname{Ext}}_R^{n+1}(M, A) \arrow[d, "\alpha_{n+1}"] \\ \underline{\operatorname{Ext}}_R^{n}(M, C') \arrow[r, "\partial'^n"] & \underline{\operatorname{Ext}}_R^{n+1}(M, A'). \end{tikzcd}$$

Corollary 128.10. *Let M be a projective R-module. Then $\operatorname{Ext}_R^n(M, N) = 0$ for all R-modules N and all $n > 0$.*

The "dual" of the pushout is the following:

Definition 128.11. Suppose that we are given a diagram of R-modules and R-homomorphisms

$$(*) \qquad \begin{tikzcd} & N_1 \arrow[d, "f_1"] \\ N_2 \arrow[r, "f_2"'] & M. \end{tikzcd}$$

Then (X, g_1, g_2), with X an R-module and g_i an R-homomorphism for $i = 1, 2$, is called the *fiber product* or *pullback* of (*) if we have a commutative diagram

$$(\dagger) \qquad \begin{tikzcd} X \arrow[r, "g_1"] \arrow[d, "g_2"'] & N_1 \arrow[d, "f_1"] \\ N_2 \arrow[r, "f_2"'] & M. \end{tikzcd}$$

and if (Y, h_1, h_2) is another such triple satisfying (†), then there exists an R-homomorphism $\alpha : Y \to X$, unique up to isomorphism, satisfying the following commutative diagram

$$
\begin{tikzcd}
Y \arrow[drr, "h_1"] \arrow[dr, "\alpha"] \arrow[ddr, "h_2"'] & & \\
& X \arrow[r, "g_1"] \arrow[d, "g_2"] & N_1 \arrow[d, "f_1"] \\
& N_2 \arrow[r, "f_2"] & M.
\end{tikzcd}
$$

Lemma 128.12. (Schanuel's Lemma) *Let A and M be R-modules. Suppose that*

$$0 \to N \xrightarrow{f_1} P \xrightarrow{g_1} M \to 0 \quad \textit{and} \quad 0 \to N' \xrightarrow{f_2} P' \xrightarrow{g_2} M \to 0$$

are short exact sequences of R-modules with P and P' projective R-modules. Then $M \coprod P' \cong M' \coprod P$.

In the analogous proof to Schanuel's Lemma for injectives, if $P = P_1 \coprod P_2$ and

$$X = \{(x_1, x_2) \mid x_i \in P_1 \coprod P_2 \mid x_i \in \ P_i \ i = 1, 2, \ \text{ with } \ f_1(x_1) = f_2(x_2)\},$$

then (X, π_1, π_2), with π_i, $i = 1, 2$, the restrictions of the projections into P, is the pullback of

$$
\begin{tikzcd}
& P_1 \arrow[d, "f_1"] \\
P_2 \arrow[r, "f_2"'] & M.
\end{tikzcd}
$$

(Cf. equation (125.14).)

Definition 128.13. Define the *projective shift operator* $\mathcal{P}$ on the collection of R-modules by $\mathcal{P}(M) := [N]$ if there exists an exact sequence

$$0 \to N \xrightarrow{f} P \xrightarrow{g} M \to 0$$

with P a projective R-module. This is well-defined by Schanuel's Lemma and $\mathcal{P}(M)$ only depends on $[M]$. In addition, $\mathcal{P}(M_1 \coprod M_2) = \mathcal{P}(M_1) + \mathcal{P}(M_2)$. Let $\mathcal{P}^0(M) = [0]$ and $\mathcal{P}^n = \mathcal{P}(\mathcal{P}^{n-1})$. In particular, if $P_* \to M \to 0$ is an projective resolution, then $\mathcal{P}^n(M) = [\ker d_n)]$. A representative of $\mathcal{P}^n(M)$ is called an *nth zyzygy* of M.

Lemma 128.14. *Let M be an R-module. Then the following are equivalent:*

(1) M is a projective R-module.
(2) $\underline{\mathrm{Ext}}_R^n(M, N) = 0$ for all $n \geq 1$ and all R-modules N.
(3) $\underline{\mathrm{Ext}}_R^1(M, N) = 0$ for all R-modules N.

Corollary 128.15. *Let M and N be R-modules. Then $\underline{\mathrm{Ext}}_R^n(M, N)$ depends only on $[M]$.*

Abusing notation, we write $\underline{\mathrm{Ext}}_R^n(\mathcal{P}(M), N)$ for M' if $\mathcal{P}(M) = [M']$ and have

Theorem 128.16. (Dimension Shift) *Suppose that M and N are R-modules. Then* $\underline{\mathrm{Ext}}_R^{n+1}(M,N) = \underline{\mathrm{Ext}}_R^n\big(\mathcal{P}(M),N\big)$ *for all $n \geq 1$. In particular, we have* $\underline{\mathrm{Ext}}_R^{n+1}(M,N) = \underline{\mathrm{Ext}}_R^1\big(\mathcal{P}^n(M),N\big)$ *for all $n \geq 1$.*

Lemma 128.17. *Let M_i, N be R-modules with $i \in I$. Then*

$$\underline{\mathrm{Ext}}_R^*(\coprod_I M_i, N) \cong \prod_I \underline{\mathrm{Ext}}_R^*(M_i, N).$$

Note if I is finite, then $\underline{\mathrm{Ext}}_R^*(\coprod_I M_i, N) \cong \coprod_I \underline{\mathrm{Ext}}_R^*(M_i, N)$.

Definition 128.18. If M is an R-module, define the *projective dimension* of M by

$$\mathrm{lpd}_R(M) = \min\{n \mid \mathcal{P}^n(M) = 0\}$$

(or infinity if no minimum exists) and the *left global projective dimension* of R to be

$$\mathrm{lgl\,proj\,dim}(R) = \max\{\mathrm{lpd}_R(M) \mid \mathcal{P}^n(M) = 0\}$$

(or infinity if no maximum exists).

Of course, we also have right projective dimension of right R modules and right global projective dimension rgl proj dim(R). For non-commutative rings $\mathrm{lgl\,proj\,dim}(R)$ and rgl proj dim(R) may be different.

Corollary 128.19. *Let N be an R-module. Then the following are equivalent:*

(1) $\mathrm{lpd}_R(N) \leq n$.
(2) $\mathcal{P}^n(N)$ is an projective R-module.
(3) $\underline{\mathrm{Ext}}^1\big(\mathcal{P}(M),N\big) = 0$ *for all R-modules N.*
(4) $\underline{\mathrm{Ext}}_R^{n+1}(M,N) = 0$ *for all $n > 0$ and all R-modules N.*

We can also tie injective modules into our computations of $\underline{\mathrm{End}}_R$ as Lemma 125.16 is also valid for $\underline{\mathrm{End}}_R$.

Proposition 128.20. *Let N be an R-module. Then the following are equivalent:*

(1) N is an injective R-module.
(2) $\underline{\mathrm{Ext}}_R^n(M,N) = 0$ *for all $n \geq 1$ and all R-modules M.*
(3) $\underline{\mathrm{Ext}}_R^1(M,N) = 0$ *for all R-modules M.*
(4) $\underline{\mathrm{Ext}}_R^1(M,N) = 0$ *for all cyclic R-modules M.*

PROOF. (1) $\Rightarrow$ (2): Let $P_* \to M \to 0$ be a projective resolution. In particular, P_* is exact (i.e., $P_0 \to P_1 \to P_2 \to \cdots$ is exact.) As Q is an injective R-module, $\mathrm{Hom}\big(\quad, N\big)$ is exact, by Exercise 124.16(1), i.e.,

$$\mathrm{Hom}_R(P_0,Q) \to \mathrm{Hom}_R(P_1,Q) \to \mathrm{Hom}_R(P_2,Q) \to \cdots .$$

is exact. Therefore, its cohomology $\mathrm{End}_R^n(M,Q) = 0$ for all $n \geq 1$
(2) $\Rightarrow$ (3) $\Rightarrow$ (4) are immediate.
(4) $\Rightarrow$ (1): We show N is an injective R-module using the Baer Criterion 124.7. Let $0 \to \mathfrak{A} \xrightarrow{\mathrm{inc}} R \xrightarrow{f} M \to 0$ be an exact sequence of R-modules. Then M is a cyclic R-module. Taking the long exact sequence in the first variable Theorem 125.9 yields

$$0 \to \mathrm{Hom}_R(M,N) \to \mathrm{Hom}_R(R,N) \xrightarrow{f^*} \mathrm{Hom}_R(\mathfrak{A},N) \to \underline{\mathrm{Ext}}_R^1(M,N).$$

As $\underline{\mathrm{Ext}}^1_R(M,N)=0$ by hypothesis, $\mathrm{Hom}_R(R,N)\xrightarrow{f^*}\mathrm{Hom}_R(\mathfrak{A},N)$ is surjective as needed. $\square$

By Lemma 125.16 and Proposition 128.20, we have

Corollary 128.21. *Let N be an R-modules. Then N is an injective R-module if and only if $\mathrm{Ext}^n_R(M,N)=\underline{\mathrm{Ext}}^n_R(M,N)=0$ for all $n\geq 1$ and all R-modules M.*

Analogously to our previous argument, Proposition 128.20 implies that we have

Corollary 128.22. *Let M and N be R-modules. Then $\underline{\mathrm{End}}_R(M,N)$ depends only on the injective equivalence class of N.*

We can now show the following:

Theorem 128.23. *Let M and N be R-modules. Then $\underline{\mathrm{Ext}}^n_R(M,A)\cong\mathrm{Ext}^n_R(M,A)$ for all $n\geq 1$. We write them both as $\mathrm{Ext}^n_R(A,B)$. In particular, $\mathrm{Ext}_R(\ ,\ )$ commutes with finite direct sums in both variables.*

Proof. Let $0\to N\xrightarrow{\varepsilon}I_*$ be an injective resolution of N. We know that $\mathrm{End}^0_R(M,N)=\mathrm{Hom}_R(M,N)=\underline{\mathrm{End}}^0_R(M,N)$. Let $K_1=\operatorname{coker}\varepsilon$ and $K_n=\operatorname{coker}(d^{n-1}:I_n\to I_{n+1})$ for all $n>1$. taking the long exact sequence on the short exact sequence $0\to N\to I_0\to K_1$ for both Ext_R and $\underline{\mathrm{Ext}}_R$ gives a commutative diagram with exact rows

$$\begin{array}{ccccccc}
\mathrm{Hom}_R(M,I_0) & \longrightarrow & \mathrm{Hom}_R(M,K_1) & \xrightarrow{\delta} & \mathrm{Ext}^1_R(M,N) & \longrightarrow & \mathrm{Ext}^1_R(M,I_0)\\
\| & & \| & & \downarrow & & \\
\mathrm{Hom}_R(M,I_0) & \longrightarrow & \mathrm{Hom}_R(M,K_1) & \xrightarrow{\delta} & \underline{\mathrm{Ext}}^1_R(M,N) & \longrightarrow & \underline{\mathrm{Ext}}^1_R(M,I_0).
\end{array}$$

Since $\mathrm{Ext}^n_R(M,I)=0=\underline{\mathrm{Ext}}^n_R((M,I)$ for any injective R-module I by Lemma 125.16 and Proposition 128.20 for any $n\geq 1$, we see that see that the result holds for $n=1$. As

$$0\to K_{n-1}\to I_{n-1}\to I_n\cdots$$

is an injective resolution of K_{n-1}. By induction, we may assume that

$$\underline{\mathrm{End}}^n_R(M,N)\cong\underline{\mathrm{End}}^n_R\big(M,\mathcal{I}^{n-1}(N)\big)\cong\mathrm{End}^1_R(M,K_{n-1}).$$

It follows that $\underline{\mathrm{Ext}}^1_R(M,K_{n-1})$ is the cohomology of

$$0\to\mathrm{Hom}_R(M,I_{n-1})\to\mathrm{Hom}_R(M,I_n)\to\mathrm{Hom}_R(M,I_{n+1})\to\cdots$$

at $\mathrm{Hom}_R(M,I_n)$ which is precisely $\mathrm{End}^n_R(M,N)$. $\square$

Identifying Ext_R and $\underline{\mathrm{Ext}}_R$, we see that we still have dimension shifting for projective R-modules.

Corollary 128.24. *Let M and N be R-modules. Then*

$$\mathrm{Ext}^n_R(\mathcal{P}(M),N)=\mathrm{Ext}^{n+1}_R(M,N)=\mathrm{Ext}^R_n(M,\mathcal{I}(N))$$

for all $n>1$.

Hence by Corollary 125.21, we have

Corollary 128.25. *The following are equivalent:*

(1) $\operatorname{lgl\,proj\,dim}(R) \leq n$.
(2) $\operatorname{Ext}_R^{n+1}(M, N) = 0$ *for all R-modules M and N.*
(3) $\operatorname{Ext}_R^{n+i}(M, N) = 0$ *for all* $i > 0$ *and R-modules M and N.*
(4) $\operatorname{lid}(R) = \sup\{N \mid N \text{ a cyclic } R\text{-module}\}$.

In particular,

Corollary 128.26. *Let N be an R-module. Then the following are equivalent:*

(1) M is a projective R-module.
(2) $\operatorname{Ext}_R^n(M, N) = 0$ *for all* $n \geq 1$ *and all R-modules N.*
(3) $\operatorname{Ext}_R^1(M, N) = 0$ *for all R-modules N.*
(4) $\operatorname{Ext}_R^1(M, N) = 0$ *for all cyclic R-modules N.*

Putting this together gives us the following theorem:

Theorem 128.27. (Auslander) *Let R be a ring. Then*

$$\operatorname{lgl\,inj\,dim}(R) = \operatorname{lgl\,proj\,dim}(R).$$

We write $\operatorname{lgl\,dim}(R)$ *for it. Moreover,*

$$\operatorname{lgl\,dim}(R) = \sup\{\operatorname{lpd}(M) \mid M \textit{ a cyclic } R\textit{-module}\}.$$

Proof. We may assume that $\operatorname{lgl\,dim}(R)$ is finite. Let

$$n = \sup\{\operatorname{lpd}(M) \mid M \text{ a cyclic } R\text{-module}\}.$$

It suffices to show that $\operatorname{lgl\,inj\,dim}(R) = \operatorname{lgl\,proj\,dim}(R) < n$. This follows from $\operatorname{Ext}_R^1(M, N) = \operatorname{Ext}_R^1(\mathcal{P}^n(M), N) = 0$, since $\mathcal{P}^n(M)$ is projective by Corollary 128.26. □

Corollary 128.28. *Let R be a ring. Then the following are equivalent:*

(1) $\operatorname{lgl\,dim}(R) = 0$
(2) Every R-module is projective.
(3) Every R-module is injective.
(4) Every short exact sequence of R-modules and R-homomorphisms splits.

Corollary 128.29. *A ring R is left hereditary if and only* $\operatorname{lgl\,dim}(R) \leq 1$.

Proof. This follows from Theorem 126.14 □

Exercises 128.30. 1. Let $0 \to A \to B \to C \to 0$ be an exact sequence of R-modules. Show if two of $\operatorname{lpd} A, \operatorname{lpd}(B), \operatorname{lpd}(C)$ are finite so is the third.

2. Let $0 \to A \to B \to C \to 0$ be an exact sequence of R-modules. Show all of the following:
   (i) If $\operatorname{lpd}(A) < \operatorname{lpd}(B)$, then $\operatorname{lpd}(C) = \operatorname{lpd}(A)$.
   (ii) If $\operatorname{lpd}(A) > \operatorname{lpd}(B)$, then $\operatorname{lpd}(C) = \operatorname{lpd}(A) + 1$.
   (iii) If $\operatorname{lpd}(A) = \operatorname{lpd}(B)$, then $\operatorname{lpd}(C) \leq \operatorname{lpd}(A) + 1$.
3. Let $0 \to A \to B \to C \to 0$ be an exact sequence of R-modules. Show all of the following:
   (i) If $\operatorname{lpd}(B) < \operatorname{lpd}(A)$, then $\operatorname{lpd}(C) = \operatorname{lpd}(A)$.
   (ii) If $\operatorname{lpd}(B) > \operatorname{lpd}(A)$, then $\operatorname{lpd}(C) = \operatorname{lpd}(B) + 1$.

(*iii*) If $\mathrm{lpd}(B) = \mathrm{lpd}(A)$, then $\mathrm{lpd}(C) \leq \mathrm{lpd}(B) + 1$.

4. Let $0 \to A \to B \to C \to 0$ be an exact sequence of R-modules. Show all of the following:
   (*i*) If $\mathrm{lpd}(A) = n$ and $\mathrm{lpd}(C) \leq n$, then $\mathrm{lpd}(B) \leq n$.
   (*ii*) $\mathrm{lpd}(B) \leq \max(\mathrm{lpd}(A), \mathrm{lpd}(C))$.
   (*iii*) In (*ii*), we have inequality if the sequence does not split.
5. Let $0 \to A \to B \to C \to 0$ be an exact sequence of R-modules. Show all of the following:
   (*i*) $\mathrm{lpd}(C) \leq \max(\mathrm{lpd}(A), \mathrm{lpd}(B))$.
   (*ii*) If B is projective, then either all three are projective or $\mathrm{lpd}(C) = 1 + \mathrm{lpd}(A)$.
6. Let R be the ring $\begin{pmatrix} \mathbb{Z} & \mathbb{Q} \\ 0 & \mathbb{Q} \end{pmatrix}$. Show $\mathrm{rgd}(R) = 1$ and $\mathrm{lgd}(R) > 1$.
7. Let $(R, \mathfrak{m})$ be a Noetherian local ring and M a nonzero finitely generated R-module. Suppose that a maximal R-sequence on M has length m. Then all maximal R-sequences on M have length m. Moreover, m is the smallest integer n satisfying $\mathrm{Ext}^n_R(R/\mathfrak{m}, M) \neq 0$.

129. Tensor Product Revisited

We generalize the tensor product of modules over a commutative ring to the tensor product of modules over an arbitrary ring, i.e., not necessarily commutative. We indicate this, leaving the proofs as exercises. We first must replace the definition of a bilinear map.

Definition 129.1. Let R be a ring with M a right R-module, N a left R-module, and A an abelian group. A map $j : M \times N \to A$ is called an *R-balanced biadditive map*, it it is additive in each variable and is *balanced*, i.e., $j(mr, n) = j(m, rn)$ for all $r \in R$, $m \in M$, and $n \in N$.

An abelian group T is then called a *tensor product* of M and N if there exists an R-balanced biadditive group homomorphism $\iota : M \times N \to T$ satisfying the following universal property: If A is an abelian group and $j : M \times N \to A$ an R-balanced biadditive map, then there exists a unique group homomorphism $f : T \to A$ such that

$$\begin{tikzcd} M \times N \arrow[r, "\iota"] \arrow[dr, "j"'] & T \arrow[d, "f"] \\ & A \end{tikzcd}$$

commutes.

Theorem 129.2. *Let R be a ring with M a right R-module and N a left R-module. Then a tensor product $\iota : M \times N \to T$ exists and is unique up to a unique isomorphism. We denote it by $M \otimes_R N$.*

Remark 129.3. Let Z be the center of R and $\mathfrak{A}$ the ideal in $M \otimes_Z N$ generated by $\{r \in R \mid mr \otimes n - m \otimes rn,\ m \in M,\ n \in N,\ r \in R\}$. Then $M \otimes_R N \cong (M \otimes_Z N)/\mathfrak{A}$.

As for the case of a commutative ring, we have (with similar proofs that we omit).

Properties 129.4. Let R be a ring and M, M' be right R-modules and N, N' left R-modules. [We leave to the reader the analogous statement by switching right and left in the statement.]

1. $0 \otimes 0 = m \otimes 0 = 0 \otimes n$ is the zero of $M \otimes_R N$ for any m in M, any n in N.
2. Let $f : M \to M'$ and $f' : M' \to M$ be R-homomorphisms of right R-modules and $g : N \to N'$ and $g' : N' \to N''$ R-homomorphisms of left R-modules. Then there exists a unique group homomorphism $f \otimes g : M \otimes_R M' \to N \otimes_R N'$ sending $m \otimes m' \mapsto f(m) \otimes g(m')$. Moreover, $(f' \otimes g')(f \otimes g) = f'f \otimes g'g$ using the universal property of tensor product.
3. The canonical R-homomorphism $\iota_N : R \otimes_R N \to N$ induced by $r \otimes n \mapsto rn$ and the universal property of tensor products is a $\mathbb{Z}$-isomorphism and it is a *natural* $\mathbb{Z}$-homomorphism, , i.e., if $\varphi : N \to N'$ is an R-homomorphism, then we have commutative diagram
$$\begin{array}{ccc} R \otimes_R N & \xrightarrow{\iota_N} & N \\ {\scriptstyle 1_R \otimes \varphi}\downarrow & & \downarrow{\scriptstyle \varphi} \\ R \otimes_R N' & \xrightarrow{\iota_{N'}} & N'. \end{array}$$
4. Let M_i, $i \in I$, be right R-modules and N a left R-module. Then there exists a canonical map
$$f_N : (\coprod_I M_i) \otimes_R N \to \coprod_I (M_i \otimes_R N).$$
Moreover, the map is natural as above, i.e., if $g : N \to N'$ is an R-homomorphism of left R-modules, then we have a commutative diagram
$$\begin{array}{ccc} (\coprod_I M_i) \otimes_R N & \xrightarrow{f_N \otimes 1_N} & \coprod_I (M_i \otimes_R N) \\ {\scriptstyle 1_M \otimes g}\downarrow & & \downarrow{\scriptstyle 1_M \otimes g} \\ (\coprod_I M_i) \otimes_R N' & \xrightarrow[f_{N'} \otimes 1_{N'}]{} & \coprod_I (M_i \otimes_R N') \end{array}$$
where $M = \coprod_I M_i$. (Similarly, we have the analogous result switching sides.)
5. Let M_i be right R-modules, $i \in I$, and $N_j$be left R-modules, $j \in J$. Then $\coprod_I M_i \otimes_R \coprod_J N_j \cong \coprod_I \coprod_J (M_i \otimes_R N_j)$. We say that $\otimes_R$ and $\coprod$ commute.

Proposition 129.5. *Let M be a left R-module (respectively, right R-module) and*
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
a short exact sequence of left R-modules (respectively right R-modules). Then
$$A \otimes_R M \xrightarrow{f \otimes 1_M} B \otimes_R M \xrightarrow{g \otimes 1_M} C \otimes_R M \to 0$$
(*respectively,*
$$M \otimes_R A \xrightarrow{1_M \otimes f} M \otimes_R B \xrightarrow{1_M \otimes g} M \otimes_R C \to 0 \quad)$$
are exact.

Proof. We show the first induced sequence is exact. It is easy to see that the sequence is a zero sequence.

We first show exactness at $B \otimes_R M$. Let $D = \operatorname{im}(1_M \otimes f)$ and $\pi = \overline{\ \ } : M \otimes_R B \to (M \otimes_R B)/D$ be the canonical map. As $D \subset \ker(1_M \otimes g)$, the composition of the canonical map and $1_M \otimes g$ induce $\overline{g} : (M \otimes_R B)/D \to M \otimes_R C$. We claim that $\overline{g}$ is an isomorphism. If we do so, then we are done, as $\ker(1_M \otimes g) = \ker \overline{g}\pi = \ker \pi = D = \operatorname{im}(1_M \otimes g)$ as needed.

Define $h : M \times C \to (M \otimes_R B)/D$ by $h(m, c) = \overline{m \otimes b}$ where $g(b) = c$. We show h is well-defined. Since g is surjective, such a b exists. Suppose that $g(b') = c = g(b)$ for some $b' \in B$, then $b' - b$ lies in $\ker g = A$. So there exists $a \in A$ satisfying $b - b' = f(a)$. Therefore,

$$m \otimes b' - m \otimes b = m \otimes (b' - b) = (1 \otimes f)(m \otimes a)$$

which lies in $\operatorname{im}(1_M \otimes f)$. Therefore, h is well-defined. It is clearly R-biadditive as well as R-balanced, hence induces a $\mathbb{Z}$-homomorphism $\widetilde{f} : M \otimes_R C \to (M \otimes_R B)/D$ sending $m \otimes c \mapsto \overline{m \otimes b}$ with $g(b) = c$. Since $\widetilde{f}$ is the inverse of $\overline{g}$, we have $\overline{g}$ is an isomorphism.

To show $1_M \otimes g$ is surjective, let $\sum m_i \otimes c_i$ lie in $M \otimes_R C$. As g is surjective, there exist b_i in B such that $g(b_i) = c_i$ for all i. Then $(1 \otimes g)(\sum m_i \otimes c_i) = \sum m_i \otimes c_i$ lies in $\operatorname{im}(1_M \otimes g)$. $\square$

We say that $\otimes_R$ is *right exact* in each variable. The proposition is still true if we do not assume the first map f in the proposition is a monomorphism.

In general, $\otimes_r$ is not exact. For example, if M is a torsion abelian group, then $M \otimes_Z$ does not preserve the injectivity of the exact sequence $0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$.

Definition 129.6. Let R and S be rings. An abelian group M is called an $(R\text{-}S)$-*bimodule* if M is a left R-module and a right S-module satisfying $r(ms) = (rm)s$ for all $s \in S$, $r \in R$, and $m \in M$. We sometimes write $M = {}_RM_S$ if M is an $(R\text{-}S)$-bimodule. It is also sometimes convenient to write ${}_SN$ if N is a left S-module and N_S if N_S is a right S-module.

For example, if S is an R-algebra (so R is commutative), then, by definition, S is an $(R\text{-}R)$-bimodule (as well as an $(S\text{-}S)$-bimodule).

Remarks 129.7. Let R, S be rings. (We leave variations to the reader.)

1. Let M be a right R-module, N an $(R\text{-}S)$-bimodule, and P a left S-module. Then $(M \otimes_R N) \otimes_S P \cong M \otimes_R (N \otimes_S P)$ (and the isomorphism is natural).
2. R is an $(R\text{-}R)$-bimodule. If N is a left R-module, then so is $R \otimes_R N$ with action given by $r(1 \otimes n) = r \otimes n$. By the universal property of tensor products, we have an R-homomorphism $N \to R \otimes_R N$ determined by $n \mapsto 1 \otimes n$ and it is a (natural) isomorphism and often identified as an equality.
3. If M is an $(R\text{-}S)$-bimodule and N a left S-module, then $M \otimes_S N$ is a left R-module.
4. Suppose that M_i are $(R\text{-}R)$-bimodules for $i \in I$ and N a left R-module, then $(\coprod_I M_i) \otimes_R N \cong \coprod_I (M_i \otimes_R N)$ as left R-modules.
5. If N is a free left R-module on basis $\mathcal{B}$, then $R \otimes_R N$ is free on basis $\{1 \otimes_R x\}_{\{x \in \mathcal{B}\}}$ and isomorphic to $R^{|\mathfrak{B}|}$. (Of course, $R^m \cong R^n$ is possible with $m \neq n$ general).

We also need the structure on Hom-sets.

Remark 129.8. Let R and S be rings. If M and N are left R-modules then $\mathrm{Hom}_R(M,N)$ is an abelian group. If, in addition, $M = {}_RM_S$ is an $(R$-$S)$-bimodule, then we can make $\mathrm{Hom}_R(M,N)$ into a left S-module as follows: For each $s \in S$ and $\varphi \in \mathrm{Hom}_R(M,N)$, define $s\varphi(m) = \varphi(ms)$ for all $m \in M$, $s \in S$. Then φ is a left R-homomorphism, hence induces a left R-action on $\mathrm{Hom}_R(M,N)$. This S-action makes $\mathrm{Hom}_R(M,N)$ into a left S-module as desired.

Theorem 129.9. (Adjoint Associativity Theorem) *Let R and S be rings, $A = A_R$ a right R-module, $B = {}_RB_S$ an $(R$-$S)$-bimodule, and $C = C_S$ a right S-module. Then there is a natural isomorphism of abelian groups,*

$$\mathrm{Hom}_S(A \otimes_R B, C) \cong \mathrm{Hom}_R\big(A, \mathrm{Hom}_S(B,C)\big).$$

We say that $-\otimes_R B$ and $\mathrm{Hom}_S(B,-)$ are adjoints.

Proof. We have $A \otimes_R S$ is a right S-module (by the variant of Remark 129.7(3)). Let $\varphi : A \otimes_R B \to C$ be an S-homomorphism. For each $a \in A$, define a map $\Phi(a) : B \to C$ by $\Phi(a)(b) = \varphi(a \otimes b)$. Check that $\Phi(a)$ is S-homomorphism of right S-modules and the map $\Phi : A \to \mathrm{Hom}_S(B,C)$ given by $a \mapsto \varphi(a)$ is an R-homomorphism of right R-modules. Define $f : \mathrm{Hom}_S(A \otimes_R B, C) \to \mathrm{Hom}_R\big(A, \mathrm{Hom}_S(B,C)\big)$ by $f(\varphi) = \Phi$. Now let $\Psi : A \to \mathrm{Hom}_S(B,C)$ be an R- homomorphism. The map $\psi : A \times B \to C$ defined by $(a,c) \mapsto \Psi(a)(c)$ is an R-bidadditive, R-balanced map, so induces a group homomorphism $\psi : A \otimes_R B \to C$. Then $g : \mathrm{Hom}_R\big(A, \mathrm{Hom}_S(B,C)\big) \to \mathrm{Hom}_S(A \otimes_R B, C)$ defined by $g(\Psi) = \psi$ is a group homomorphism and the inverse of f. We leave the proof that the map is natural as an exercise. $\square$

In a similar manner, we see the following:

Theorem 129.10. *Let R and S be rings, $A = {}_RA$ a left R-module, $B = {}_SB_R$ an $(R$-$S)$-bimodule, and $C = {}_SC$ a left S-module. Then there is a natural isomorphism of abelian groups,*

$$\mathrm{Hom}_S(B \otimes_R A, C) \cong \mathrm{Hom}_R\big(A, \mathrm{Hom}_S(B,C)\big).$$

Exercises 129.11. 1. Let R be a commutative ring. Then the tensor product of two projective R modules is projective.

2. Let $A \xrightarrow{f} B \xrightarrow{g} C \to 0$ be a sequence of right R-modules. Suppose that for every left R-module M, the sequence

$$) \to \mathrm{Hom}_R(C,M) \xrightarrow{g^*} \mathrm{Hom}_R(B,M) \xrightarrow{f^*} \mathrm{Hom}_R(B,M)$$

is exact. Show the first sequence is exact. Also prove the "dual" statement.

3. Let M be a right R-module. Then $B\otimes_R$ takes a sequence exact sequence $A \xrightarrow{f} B \xrightarrow{g} C \to 0$ of abelian groups to an exact sequence of abelian groups.

4. Let A be a left R-module, B an $(R$-$S)$-bimodule, and C a right S-module. Show that there is a natural isomorphism of abelian groups

$$\mathrm{Hom}_R(A, \mathrm{Hom}_S(C,B)) \cong \mathrm{Hom}_S\big(C, \mathrm{Hom}_R(A,B)\big).$$

5. Let M be a left R-module and A an abelian group. Show that $\operatorname{Hom}_{\mathbb{Z}}(M, A) \cong \operatorname{Hom}_R(M, \operatorname{Hom}_{\mathbb{Z}}(R, A))$.
6. Use the previous exercise to help give another proof that every R-module embeds into an injective R-module.

130. Limits

We need a generalization of an infinite direct sum (and also give a generalization of direct products). In this section we shall continue to label modules as either right or left if R is an arbitrary ring.

Definition 130.1. Let I be a partially ordered ordered set under $\leq$ and $\{M_i\}_I$ be a collection of left (respectively, right) R-modules. Suppose for all $i \leq j$ in I, there exist R-homomorphisms $\theta_{i,j} : M_i \to M_j$ with $\theta_{i,i} = 1_{M_i}$ and

$$\begin{tikzcd} M_i \arrow[d, "\theta_{ij}"'] \arrow[dr, "\theta_{ik}"] & \\ M_j \arrow[r, "\theta_{jk}"'] & M_k \end{tikzcd}$$

commutes whenever $i \leq j \leq k$. We call $(M_i, \{\theta_{ij} \mid i \leq j\}_J)_I$ a *directed system* and write it as $(M_i, \theta_{ij})_I$ (or even simply $(M_i)_I$ if the θ_{ij} are understood). Let $(M_i, \theta_{ij})_I$ be a directed system of left (respectively right) R-modules. Let M be a left (respectively, right) R-module together with R-homomorphisms $\psi_i : M_i \to M$ for all i in I satisfying

$$\begin{tikzcd} M_i \arrow[d, "\theta_{ij}"'] \arrow[dr, "\psi_i"] & \\ M_j \arrow[r, "\psi_j"'] & M \end{tikzcd}$$

commutes for all $i \leq j$ in I. Then $(M, \psi_i)_I$ is called a *direct limit* of the $(M_i, \theta_{ij})_I$ if it satisfies the following universal property: If M' is a left (respectively, right) R-module together with commutative diagrams of R-homomorphisms

$$\begin{tikzcd} M_i \arrow[d, "\theta_{ij}"'] \arrow[dr, "\varphi_i"] & \\ M_j \arrow[r, "\varphi_j"'] & M' \end{tikzcd}$$

for all $i \leq j$ in I, then there exists a unique R-homomorphism $\mu : M_i \to M'$ satisfying

$$\begin{tikzcd}
M_i \arrow[dd, "\theta_{ij}"'] \arrow[dr, "\psi_i"] \arrow[drr, "\varphi_i"] & & \\
& M \arrow[r, "\mu"] & M' \\
M_j \arrow[ur, "\psi_j"] \arrow[urr, "\varphi_j"'] & &
\end{tikzcd}$$

commutes for all $i \leq j$ in I. We denoted this direct limit by $\varinjlim_{\theta} M_i$ or $\varinjlim M_i$ if the θ_{ij} are clear. [We shall see that the ψ_i are essentially unique.]

Proposition 130.2. *Let I be a partially ordered ordered set under $\leq$ and $(M_i, \theta_{ij})_I$ a directed system left (respectively, right) R-modules. Then there exists a direct limit $(M_i, \psi_i)_I$ of $(M_i, \theta_{ij})_I$ unique up to an R-isomorphism.*

PROOF. Let $\iota_j : M_i \to \coprod_I M_i$ be the injection map into the jth coordinate and $S_\theta := S_{(M_i,\theta_{ij})_I}$ be the left R-module generated by

$$\{\iota_j\theta_{ij}(m_i) - \iota_j(m_i) \mid m_i \in M_i,\, i \leq j \text{ in } I\}$$

and $^{-} : \coprod M_i \to \coprod_I M_i/S_\theta$ be the canonical R-homomorphism. Set

$$\psi_j : M_j \to (\coprod_I M_i)/S_\theta \text{ by } m_j \mapsto \overline{\iota_j(m_j)}.$$

Then it is easily checked that $\varinjlim M_i = (\coprod_I M_i)/S_\theta$. $\square$

Examples 130.3. 1. Let M_i, $i \in I$, be a left (respectively, right) R-module for all $i \in I$ and $M = \coprod_I M_i$. Then $M = \varinjlim M_i$ with no θ_{ij} if $i \neq j$ in I, by the Universal Property of Direct Sums. In the language of the proof of the Proposition 130.2, a better description is given by letting $S_\theta = 0$ by setting $\theta_{ii} = 1_{M_i}$ for all $i \in I$.

2. The pushout of a diagram of left (respectively, right) R-modules

$$\begin{tikzcd}
M_1 \arrow[r, "f_2"] \arrow[d, "f_3"'] & N_2 \\
N_3 &
\end{tikzcd}$$

is the direct limit with $I = \{1, 2, 3\}$ partially ordered by $1 < 2$ and $1 < 3$.

3. If M is a left (respectively, right) R-module, then it is the direct limit of its finitely generated submodules.

Definition 130.4. A partially ordered set I is called *directed* under $\leq$ if for all $i, j \in I$, there exists a $k \in I$ satisfying $i \leq k$ and $j \leq k$.

Example 130.5. Let M_i, $i \in I$, be a left (respectively, right) R-module for all $i \in I$ and $M = \coprod_I M_i$. If $J = \{j_1, \dots, j_n\}$ is a finite subset of I, set $M_J = M_{j_1} \coprod \cdots \coprod M_{j_n}$ and let $inc_J : M_J \to M$ be the inclusion map for all finite $J \subset I$. Let $\mathcal{I} := \{J \mid J \subset I \text{ finite}\}$ a partially ordered by inclusion. If $I, J \in \mathcal{I}$, then $I \subset I \cup J$ and $J \subset I \cup J$. So $\mathcal{I}$

is a directed partially ordered set. Define $inc_{IJ} : M_I \to M_J$ if $I \subset J$ in $\mathcal{I}$ to be the inclusion map. Then $(M_I, inc_{IJ})_{\mathcal{I}}$ is a directed system by a direct partially ordered set and $M = (M_J, inc_J)_{\mathcal{I}} = \varinjlim M_J$ is the direct limit via the directed partially ordered set $\mathcal{I}$. We leave the verification as an exercise.

Definition 130.6. Let I be a directed partially order set and $M = (M_i, \theta_{ij})_I$ and $N = (N_i, \rho_{ij})_I$ be directed systems of left (respectively, right) R-modules. An R-homomorphism of directed systems $M \to N$ is a map $f = (f_i)_I$ where $f_i : M_i \to N_i$ is an R-homomorphism for all $i \in I$ and satisfies $f_j\theta_{ij} = \rho_{ij} f_j$ whenever $i \le j$ in I. We call a sequence of directed systems

$$0 \to (M_i', \theta_{ij})_I \xrightarrow{(f_i)_I} (M_i, \theta_{ij}')_I \xrightarrow{(g_i)_I} (M_i'', \theta''{}_ij)_I \to 0$$

an *exact sequence* of directed systems if $0 \to M_i' \xrightarrow{f_i} M_i \xrightarrow{g_i} M_i'' \to 0$ is exact for all $i \in I$.

For the rest of our discussion about direct limits, we shall need our partially ordered sets to always be directed.

Construction 130.7. Let I be a directed partially ordered set and M_i, $i \in I$, be right (respectively, left) R-modules. Suppose that $M = (M_i, \theta_{ij})_I$ is a directed system with $\varinjlim_{\theta}$ its direct limit. As in the proof of Proposition 130.2, let $S_\theta = S_{(M_i,\theta_{ij})_I}$. Let $(M_i, \rho_{ij})_I$ be the directed system with $\rho_{ii} = 1_{M_i}$ for all $i \in I$. So $\coprod_I M_I$ is the direct limit of the directed system $(M_i, \rho_{ij})_I$ with corresponding $S_{(M_i,\theta_{ij})_I} = 0$. The map $(M_i, \theta_{ij})_I \to (M_i, \rho_{ij})_I$ by $\theta_{ij} \mapsto \rho_{ij}$ is a R-homomorphism of directed systems and by definition gives an exact sequence

$$0 \to S_{(M_i,\theta_{ij})_I} \to \coprod_I M_i \to \varinjlim_{\theta} M_i \to 0$$

of left (respectively, right) R-modules. If $N = (N_i, \tau_{ij})_I$ is another directed system of left (respectively, right) R-modules and $f_i : M_i \to N_i$ is an R-homomorphism for all $i \in I$ satisfying $f = (f_i)_I : M \to N$ is an R-homomorphism of directed systems, then f induces an R-homomorphism $f_0 : S_\theta \to S_\tau$ such that the diagram

$$\begin{tikzcd}
0 \arrow[d] & 0 \arrow[d] \\
S_\theta \arrow[r, "f_0"] \arrow[d] & S_\tau \arrow[d] \\
\coprod_I M_i \arrow[r, "f"] \arrow[d] & \coprod_I N_i \arrow[d] \\
\varinjlim_{\theta} M_i \arrow[r, "\widetilde{f}"] & \varinjlim_{\tau} N_i
\end{tikzcd}$$

commutes with $\widetilde{f}$ the induced map and the columns exact. We leave the details of the verification to the reader.

In the notation of this construction, we have

Lemma 130.8. *Let*

$$0 \to (M_i', \theta_{ij}')_I \xrightarrow{(f_i)_I} (M_i, \theta_{ij})_I \xrightarrow{(g_i)_I} (M_i'', \theta_{ij}'')_I \to 0$$

be an exact sequence of directed systems of left (respectively, right) R-modules by a directed partially ordering I. Then

$$0 \to S_{\theta'} \xrightarrow{(f_i)_I} S_\theta \xrightarrow{(g_i)_I} S_{\theta''} \to 0$$

is exact.

We leave the proof as an exercise.

Proposition 130.9. *Let*

$$0 \to (M_i', \theta_{ij}')_I \xrightarrow{(f_i)_I} (M_i, \theta_{ij})_I \xrightarrow{(g_i)_I} (M_i'', \theta_{ij}'')_I \to 0$$

be an exact sequence of left (respectively, right) R-modules of directed systems by a directed partially ordering I. Then

$$0 \to \varinjlim_{\theta'} M_i' \xrightarrow{\widetilde{f}} \varinjlim_{\theta} M \xrightarrow{\widetilde{g}} \varinjlim_{\theta''} M_i'' \to 0$$

is exact.

Proof. By Construction 130.7 and Lemma 130.8, we have a commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & S_{\theta'} \arrow[r, "(f_i)_I"] \arrow[d] & S_\theta \arrow[r, "(g_i)_I"] \arrow[d] & S_{\theta''} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \coprod_I M_i' \arrow[r, "(f_i)_I"] \arrow[d] & \coprod_I M_i \arrow[r, "(g_i)_I"] \arrow[d] & \coprod_I M_i'' \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \varinjlim M_i' \arrow[r, "\widetilde{f}"] & \varinjlim M_i \arrow[r, "\widetilde{g}"] & \varinjlim M_i'' \arrow[r] & 0
\end{tikzcd}$$

with exact columns and the top two rows are exact. The result follows by the Snake Lemma 122.1. □

The proposition says that $\varinjlim$ is *exact*, i.e., it takes short exact sequences of directed systems of left (respectively, right) R-modules by a directed partially ordered set to short exact sequences.

Lemma 130.10. *Let R be a ring and $(M_i, \theta_{ij})_I$ a directed system of right R-modules over a directed partially ordered set I. Then for any left R-module B, the R-homomorphism*

$$f_B : (\varinjlim M_i) \otimes_R B \to \varinjlim (M_i \otimes_R B)$$

induced by $(m_i)_I \otimes b \mapsto (m_i \otimes b)_I$ *is an isomorphism of abelian groups. Moreover, if* $g : A \to B$ *is an* R*-homomorphism of left* R*-modules and* $M = \varinjlim M_i$*, then we have a commutative diagram*

$$\begin{tikzcd}
(\varinjlim M_i) \otimes_R A \arrow[r, "f_A \otimes 1_A"] \arrow[d, "1_M \otimes g"'] & \varinjlim (M_i \otimes_R A) \arrow[d, "\widetilde{1_M \otimes g}"] \\
(\varinjlim M_i) \otimes_R B \arrow[r, "f_B \otimes 1_B"'] & \varinjlim (M_i \otimes_R B)
\end{tikzcd}$$

of abelian groups where the right vertical is the homomorphism to the direct limit. An analogous result holds for direct limits of left R*-modules and right* R*-modules.*

Proof. Exercise. □

Proposition 130.11. *Let* R *be a ring and* $(M_i, \theta_{ij})_I$ *a directed system of right* R*-modules over a directed partially ordered set* I*. If*

$$0 \to A \to B \to C \to 0$$

is an exact sequence of left R*-modules, then*

$$0 \to \varinjlim M_i \otimes_R A \to \varinjlim M_i \otimes_R B \to \varinjlim M_i \otimes_R C \to 0$$

is an exact sequence of abelian groups.
An analogous result holds for left and right R*-modules.*

Proof. This follows by Proposition 130.9, Lemma **??**, and Lemma 130.10 □

Remark 130.12. If the M_i is an (S-R) bimodule, in the Lemma **??** and Lemma 130.10, the isomorphisms are S-isomorphisms and in Proposition 130.11, the sequence is an exact sequence of S-modules. (An analogous result for M_i are (R-S) on the other side.)

We define the dual of the direct limit. We leave the analogous proofs as exercises.

Definition 130.13. Let I be a partially ordered ordered set under $\leq$ and $\{M_i\}_I$ be a collection of left (respectively, right) R-modules. Suppose for all $i \leq j$ in I, there exist R-homomorphisms $\theta_{ij} : M_j \to M_i$ with $\theta_{i,i} = 1_{M_i}$ and

$$\begin{tikzcd}
M_i & \\
& M_k \arrow[ul, "\theta_{ik}"'] \arrow[dl, "\theta_{jk}"] \\
M_j \arrow[uu, "\theta_{ij}"] &
\end{tikzcd}$$

commutes whenever $i \leq j \leq k$. We call $(M_I, \theta_{ij})_I = (M_i, \{\theta_{ij}\}_J)_I$ an *inverse system.* Let $(M_i, \theta_{ij})_I$ be such an inverse system and M be an left (respectively, right) R-module together with R-homomorphisms $\psi_i : M_i \to M$ for all i in I and commutative diagrams

$$\begin{tikzcd}
M_i & \\
& M \arrow[ul, "\psi_i"'] \arrow[dl, "\psi_j"] \\
M_j \arrow[uu, "\theta_{ij}"] &
\end{tikzcd}$$

for all $i \leq j$ in I. Then $(M, \psi_i)_I$ is called an *inverse limit* or *projective limit* of the $(M_i, \{\theta_{ij}\}_I)_I$ and denoted by $\varprojlim M_i$ if it satisfies the following universal property: If M' is a left R-module together with commutative diagrams of R-homomorphisms

$$\begin{tikzcd}
M_i & \\
& M' \arrow[ul, "\varphi_i"'] \arrow[dl, "\varphi_j"] \\
M_j \arrow[uu, "\theta_{ij}"] &
\end{tikzcd}$$

for all $i \leq j$ in I, then there exists a unique R-homomorphism $\mu : M' \to M$ satisfying

$$\begin{tikzcd}
M_i & & \\
& M \arrow[ul, "\psi_i"'] \arrow[dl, "\psi_j"] & M' \arrow[l, "\mu"'] \arrow[ull, "\varphi_i"'] \arrow[dll, "\varphi_j"] \\
M_j \arrow[uu, "\theta_{ij}"] & &
\end{tikzcd}$$

commutes for all $i \leq j$ in I. Of course, we have an analogous definition for right R-modules.

Proposition 130.14. *Let I be a partially ordered ordered set under $\leq$ and $(M_i, \theta_{ij})_I$ an inverse system of left (respectively right) R-modules Then an inverse limit M of $(M_i, \psi_i)_I$ exists and is unique up to an R-isomorphism.*

Here we have $\varprojlim M_i$, is the submodule S_θ of $\prod_I M_i$ generated by the set $\{\alpha = (a_i)_I \in \prod_I M_i \mid a_i = \theta_{ij} a_j \text{ if } i \leq j\}$ with $\psi_j : \varprojlim M_i \to M_j$ given by the restriction of the projection map $\pi_j : \prod_I M_i \to M_j$.

Examples 130.15. 1. $M = \prod_I M_i$, the direct product of left (respectively, right) R-modules, is the inverse limit when $S_\theta = 0$ in the above, e.g., where $\theta_{ii} = 1_{M_i}$ and $\theta_{ij} = 0$ if $i < j$.

2. The pushout of a diagram of R-modules

$$\begin{array}{ccc} M_1 & \xrightarrow{f_2} & N_2 \\ {\scriptstyle f_3}\downarrow & & \\ N_3 & & \end{array}$$

When we try to reverse arrows in our proofs of direct limits, we run into a problem about exactness. In general the inverse limit does not take exact sequences to exact sequences of inverse systems under a directed partially ordered set. It is only left exact. You will see this if you go through the proofs for direct limits when we invoked the Snake Lemma. [In general, subobjects are harder to work with then quotient objects.] Inverse limits are very important. In commutative algebra they give rise to completions, the analog of completions in topology. For example, if R is a commutative ring and $\mathfrak{A}$ an ideal, then $\varprojlim(R/\mathfrak{A}^i)$ is the completion of R in the $\mathfrak{A}$adic topology, e.g., the completion of $R[t_1, \dots, t_n]$ is the ring of formal power series $R[[t_1, \dots, t_n]]$. One can also define the inverse product for groups. For example, if K/F is a field extension with K the separable closure of F, then the Galois group of K/F is the inverse limit of finite Galois extensions of F.

Exercises 130.16. 1. Let (M_i, θ_{ij}) be a directed system of left (resp, right) R-modules over a directed partially order I. Let $\psi_i : M_i \to \varinjlim M_i$. Show that every $x \in \varinjlim M_i$, there exists an $i \in I$ and an $x_i \in M_i$ satisfying $x = \psi_i(x_i)$. Show also that $\psi_i(x_i) = 0$, then there exists $i \leq j$ satisfying $\theta_{ij}(x_i) = 0$.

2. Verify the assertions in Construction 130.7.
3. Prove Lemma 130.8
4. Prove Lemma **??**.
5. Prove Lemma 130.10.

131. Flat Modules

Definition 131.1. Let R be a ring and N a left R-module. We say that N is a *flat* R-module if whenever

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$

is exact sequence of right R-modules, so is

$$0 \to M' \otimes_R N \xrightarrow{f \otimes 1_N} M \otimes_R N \xrightarrow{g \otimes 1_N} M'' \otimes_R N \to 0.$$

An analogous result holds for right R-modules.

Example 131.2. Any free right (respectively, left) R-module is flat as tensor products and direct sums commute (cf. Remark 129.4(5) and $R \otimes_R \cong M$ for any left R-module M.

Remark 131.3. Suppose that $f_i : N_i' \to N$ is an R-homomorphism of left R-modules for all $i \in I$. Then $\coprod_I f_i : \coprod_I N_i' \to \coprod_I N_i$ an R-monomorphism if and only if f_i, $i \in I$ is an R-monomorphism for all $i \in I$.

Proposition 131.4. *Let M_i, $i \in I$, be right (respectively, left) R-modules and $M = \coprod_I M_i$. Then M is flat if and only if every M_i, $i \in I$, is flat.*

Proof. We must show that $M\otimes_R$ preserves R-monomorphisms if and only if $M_i\otimes_R$ does for all $i \in I$. Consider the following commutative diagram:

$$\begin{array}{ccc} (\coprod_I M_i) \otimes_R N' & \xrightarrow{1_M\otimes f} & (\coprod_I M_i) \otimes_R N \\ \downarrow & & \downarrow \\ (\coprod_I M_i \otimes_R N') & \xrightarrow[\coprod(1_{M_i}\otimes f)]{} & (\coprod_I M_i \otimes_R N) \end{array}$$

with the first vertical map is the isomorphism $(\sum_I m_i)\otimes n \mapsto (\sum_I m_i \otimes n)$ and the second analogous. Then $1_M \otimes f$ is $\mathbb{Z}$- monic if and only if $1_{M_i} \otimes f$ is $\mathbb{Z}$-monic for all $i \in I$. $\square$

Corollary 131.5. *Every projective left (respectively, right) R-module is R-flat.*

Let M be a right (respectively, left) R-module. An acyclic positive chain complex $F_* \to M \to 0$ is called a *flat resolution* of M if every F_i, $i \geq 0$, is R-flat.

Corollary 131.6. *Every right (respectively, left R-module) has a flat resolution.*

Proposition 131.7. *Let R be a ring and N a left (respectively, right) R-module. Then N is R-flat if every finitely generated submodule of N is flat.*

Proof. N is the direct limit of its finitely generated flat submodules. $\square$

Definition 131.8. Let M be a right (respectively, left) R-module. We called $M^* := \operatorname{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$ the *dual* or *character* module of R. Recall that as $\mathbb{Q}$ is a divisible abelian group, hence so is $\mathbb{Q}/\mathbb{Z}$ by Observation 15.11(3). In particular, $\mathbb{Q}/\mathbb{Z}$ is an injective abelian group by Lemma 124.10. Moreover, the abelian group M^* is a left (respectively, right) R-module with R-action given by $(rf)(m) = f(mr)$ (respectively, $(fr)(m) = f(rm)$ for all $r \in R$ and $m \in M$.

Lemma 131.9. *A sequence $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ of right R-modules is exact if and only if $0 \to C^* \xrightarrow{g^*} B^* \xrightarrow{f^8} A^* \to 0$ is an exact sequence of left R-modules.*

Proof. $(\Rightarrow)$: As $\mathbb{Q}/\mathbb{Z}$ is $\mathbb{Z}$-injective, $\operatorname{Hom}_{\mathbb{Z}}(\ ,\mathbb{Q}/\mathbb{Z})$ takes exact R-sequences of to exact sequences.

$(\Leftarrow)$: It suffices to prove if $C^* \xrightarrow{g^*} B^* \xrightarrow{f^*} A^*$, so is $A \xrightarrow{f} B \xrightarrow{g} C$ is exact. We first show that it is a zero sequence. Suppose that $a \in A$. To show that $gf(a) = 0$. As $gf(a) \in C$, by Exercise 15.18(15), it suffices to show that $\big(\sigma(gf)\big)(a) = 0$ for all $\sigma \in C^*$. But $\sigma(gf(a)) = (g^*\sigma)\big(f((a)\big) = f^*\big(g^*(\sigma)\big)(a) = 0$. To show exactness, suppose that $b \in B$ satisfies $g(b) = 0$. We must show that $b \in \operatorname{im} f$. To show this, it suffices to show if $\sigma \in B^*$ vanishes on $\operatorname{im} f$, then it vanishes on b. But if $0 = \sigma(f(a)) = 0$ for all $a \in A$, then $f^*(\sigma)(a) = 0$ for all $a \in A$, i.e., $f^*(\sigma) = 0$ in A^*. Therefore, $\sigma = g^*(\tau)$ for some $\tau \in C^*$ and $\sigma(b) = (g^*\tau)(b) = \tau(g(b) = g(0) = 0$. $\square$

Theorem 131.10. *(*Lambek's Theorem) *Let R be a ring and M a right R-module. Then M is R-flat if and only if M^* is R-injective.*

Proof. (⇒): Suppose that M is R-flat. To show that M^* is left injective. Suppose that $0 \to A \to B$ is an R-monomorphism of right R-modules. Then we have a commutative diagram

$$(*)\qquad \begin{array}{ccc} \mathrm{Hom}_R(B, M^*) & \longrightarrow & \mathrm{Hom}_R(A, M^*) \\ \downarrow & & \downarrow \\ \mathrm{Hom}_R(M \otimes_R B, \mathbb{Q}/\mathbb{Z}) & \longrightarrow & \mathrm{Hom}_R(M \otimes_R A, \mathbb{Q}/\mathbb{Z}) \end{array}$$

with the verticals the natural isomorphisms given by the Adjunct Associativity Theorem 129.9. Since M is right flat, $0 \to M \otimes_R B \to M \otimes_R A$ is a monomorphism. Therefore, the bottom row of (*) is surjective as $\mathbb{Q}/\mathbb{Z}$ is $\mathbb{Z}$-injective. as needed

(⇐): Suppose that M^* is a left injective R-module and $0 \to A \to B$ is an R-monomorphism of left R-modules. We must show that $0 \to M \otimes_R A \to M \otimes_R B$ is a monomorphism. To do so, it suffices to show $(M \otimes_R B)^* \to (M \otimes_R A)^* \to 0$ is exact by Lemma 131.9. The commutative diagram

$$\begin{array}{ccc} (M \otimes_R B)^* & \longrightarrow & (M \otimes_R A)^* \\ \| & & \| \\ \mathrm{Hom}_R(M \otimes_R B, \mathbb{Q}/\mathbb{Z}) & \longrightarrow & \mathrm{Hom}_R(M \otimes_R A, \mathbb{Q}/\mathbb{Z}) \\ \downarrow & & \downarrow \\ \mathrm{Hom}_R(B, M^*) & \longrightarrow & \mathrm{Hom}_R(A, M^*) \end{array}$$

with the vertical arrows isomorphisms as before has bottom row an epimorphism as M^* is an injective R-module. □

The analogue of the Baer Criterion for injective modules is the following:

Corollary 131.11. *A right R-module M is R-flat if and only if given any finitely generated left ideal $\mathfrak{A}$ of R, we have $M \otimes_R \mathfrak{A} \xrightarrow{1_M \otimes inc} M \otimes_R R$ is an R-monomorphism.*

Proof. We show that M is R-flat if it satisfies the criterion. The converse is immediate. By Lambek's Theorem, it suffices to show that M^* is an injective left R-module. By the Baer Criterion 124.7, it suffices to show the induced map $\mathrm{Hom}_R(R, M^*) \to \mathrm{Hom}_R(\mathfrak{A}, M^*)$ is surjective whenever $\mathfrak{A}$ is a left ideal in R. The Adjoint Associativity Theorem 129.9 shows that we need only show $\mathrm{Hom}_R(M \otimes_R R, \mathbb{Q}/\mathbb{Z}) \to \mathrm{Hom}_R(M \otimes_R \mathfrak{A}, \mathbb{Q}/\mathbb{Z})$ is surjective. But this is true since $M \otimes_R \mathfrak{A} \to M \otimes_R R$ is a monomorphisms, since $\mathbb{Q}/\mathbb{Z}$ is $\mathbb{Z}$-injective.

Since taking $\varinjlim$ takes exact sequences to exact sequences by Proposition 130.11 and every ideal is the direct limit of its finitely generated subideals, we may assume that $\mathfrak{A}$ is finitely generated by Lemma 130.10.

□

Corollary 131.12. *A right R-module M is flat if and only if $M \otimes_R \mathfrak{A} \to M \otimes_R R$ is an R-monomorphism for all finitely generated left ideals of R, i.e., identifying $M \otimes_R R$ and M, we have $M \otimes_R \mathfrak{A} \to M\mathfrak{A}$ is a (natural) isomorphism.*

Proof. This follows from the commutative diagram

$$\begin{array}{ccc} M\otimes_R \mathfrak{A} & \longrightarrow & M\otimes_R R \\ \downarrow & & \| \\ M\mathfrak{A} & \longrightarrow & M \end{array}$$

as the bottom map is a monomorphism. □

Proposition 131.13. *Let F be a flat right R-module and $0\to K\to F\xrightarrow{f} B\to 0$ an exact sequence of right R-modules. Then the following are equivalent:*

1. *B is flat.*
2. *$K\cap F\mathfrak{A} = K\mathfrak{A}$ for every left ideal $\mathfrak{A}$ in R.*
3. *$K\cap F\mathfrak{A} = K\mathfrak{A}$ for every finitely generated left ideal $\mathfrak{A}$ in R.*

Proof. We have an exact sequence $K\otimes_R\mathfrak{A}\to F\otimes_R\mathfrak{A}\to B\otimes\mathfrak{A}\to 0$. As F is flat, we can identify $F\mathfrak{A}$ and $F\otimes_R\mathfrak{A}$. It follows that $F\mathfrak{A}/K\mathfrak{A}\cong B\otimes_R\mathfrak{A}\to B\mathfrak{A}\cong F\mathfrak{A}/K\cap F\mathfrak{A}$.
(1) ⇒ (2): If B is flat, then $B\otimes_F\mathfrak{A}\to B\mathfrak{A}$ is an isomorphism, hence (1) implies (2).
(2) ⇒ (3) is immediate.
(3) ⇒ (1): If $K\mathfrak{A}\cong (K\cap F\mathfrak{A})$, then $B\otimes_R\mathfrak{A}\to B\mathfrak{A}$ is an isomorphism, hence (3) implies (1) by Corollary 131.11. □

Corollary 131.14. *Let $r\in R$ not be a right zero divisor and M a flat right R-module. Then $r\notin \operatorname{ann}_R(M)$, i.e., $\lambda_r : M\to M$ by $m\mapsto rm$ is an R-monomorphism. In particular, if R is a domain and M is R-flat, then M is R-torsion-free.*

Proof. Let $0\to K\to F\xrightarrow{f} M\to 0$ be an exact sequence of right R-modules with F a free R-module, hence flat. Suppose that $mr=0$ for $m\in M$. Choose $x\in F$ such that $g(x)=m$. Therefore, $xr\in K\cap F\mathfrak{A}$ with $\mathfrak{A}=Rr$. By Proposition 131.13, $K\cap F\mathfrak{A}=K\mathfrak{A}=KRr=Kr$. So there exists $k\in K$ satisfying $xr=kr$. Since r is not a zero divisor in R, we have Kr is free of rank one. Writing kr and xr in a basis for F, we see that $x=k$. Consequently, $m=g(x)=g(k)=0$ □

Theorem 131.15. (Villamayor) *Let F be a right free R-module and $0\to K\to F\xrightarrow{f} B\to 0$ be an exact sequence of right R-modules. Then the following are equivalent:*

1. *B is flat.*
2. *For all $x\in K$, there exists an R-homomorphism $\theta : F\to K$ such that $\theta|_K = 1_K$.*
3. *For all $v_1,\dots,v_n\in K$, let $K_0=\sum_{i=1}^n Rv_i$. Then there exists an R-homomorphism $\theta : F\to K$ such that $\theta|_{K_0}=1_K$*

Proof. Let $\mathcal{B}$ be a basis for F.
(1) ⇒ (2): Assume that B is R-flat and $v\in K$. Then $v=\sum_{i=1}^s x_{n_i}a_i$ for some $x_{n_i}\in\mathcal{B}$ and $a_i\in R$ for some s. Set $\mathfrak{A}=\sum_{i=1}^s Ra_i$ a left ideal in R. Since B is R-flat, $v\in K\cap F\mathfrak{A}=K\mathfrak{A}$. There exist finitely many k_j, $j=1,\dots,m$, in K such that $v=\sum_{j=1}^m k_jc_j$ with $c_j\in\mathfrak{A}$. Write each $c_j=\sum_{i=1}^s r_{ji}a_i$ with $r_{ji}\in R$ for all i,j. Then $v=\sum_{i=1}^s(\sum_{j=1}^m k_jr_{ji})a_i=\sum_{i=1}^r k_i'a_i$ where $k_i'=\sum_{j=1}^m k_jr_{ji}\in K$. As F is R-free, there exists an R-homomorphism $\theta : F\to K$ satisfying $x_{n_j}\mapsto k_j'$, $j=1,\dots,m$, and $x\mapsto 0$ for all $x\in\mathcal{B}\setminus\{x_{n_1},\dots,x_{n_m}\}$. By construction, $\theta(v)=v$.

(2) $\Rightarrow$ (1): Let $\mathfrak{A}$ be a left ideal in R. We want to show that $K \cap F\mathfrak{A} \subset K\mathfrak{A}$ (as $K\mathfrak{A} \subset K \cap F\mathfrak{A} \subset K\mathfrak{A}$ is always true). Let $v \in K \cap F\mathfrak{A}$. We can write $v = \sum_{i=1}^{s} x_{n_i} a_i$ for some $x_{n_i} \in \mathcal{B}$ and $a_i \in \mathfrak{A}$ for some s. By (2), there exists an R-homomorphism $\theta : F \to K$ satisfying $v = \theta(v) = \sum_{i=1}^{s} \theta(x_{n_i}) a_i$ lies in $K\mathfrak{A}$, since $\theta(x_{n_i}) \in K$ for all n_i.

(3) $\Rightarrow$ (2) is immediate.

(2) $\Rightarrow$ (3): We induct on n. Assume the case $n-1$ is true. Let $v_i' = v_i - \theta_n$ in K, $i = 1, \dots, n$, (so $v_n = 0$) where θ_n satisfies (2). By induction, there exists $\theta' : F \to K$ that fixes all the v_i'. Define $\theta = 1_F - (1_F - \theta')(1_F - \theta_n)$, an endomorphism of F satisfying $\theta(v_n) = v_n$ and $\theta(v_i) = v_i - (1_F - \theta')(v_i') = 0$ for $i = 1, \dots, n-1$. To finish, we need only show that $\theta(F) \subset K$. Since $\theta = -\theta'\theta_n + \theta' + \theta_n$ and $\theta'(K) \subset K$ and $\theta_n(F) \subset K$, this follows. □

Corollary 131.16. *Let R be a right Noetherian ring and M be a finitely generated right R-module. Then M is flat if and only if M is projective.*

PROOF. If M is a finitely generated right R-module, we have an exact sequence

$$0 \to K \to F \to M \to 0$$

with F a finitely generated free. Then K is finitely generated as R is right Noetherian. We may assume that $K \to F$ is the inclusion. If $K = Rv_1 + \cdots + Rv_n$, then and M is flat, there exists $\theta : F \to K$ such that $\theta(v_i) = v_i$ for all i. Therefore, the inclusion is a split monomorphism so a direct summand of F. Therefore, M is also isomorphic to a direct summand of F, hence projective. □

Exercises 131.17. 1. If $(F_i, \theta_{ij})_I$ is a directed system of flat right (respectively left) R-modules over a directed partially ordered set I, show that $\varinjlim F_i$ is flat.

2. Let B be an (R-S)-bimodule that is S-flat and C an injective left R-module. Show that $\mathrm{Hom}_R(B, C)$ is an injective left R-module.
3. Let M be a finitely presented right (respectively, left) R-module. Then M is flat if and only if M is projective.

132. Tor

Just as we developed a cohomology theory Ext_R^* for Hom_R, we can define a homology $\otimes_R$. Since the statements are similar as well as the proofs, we briefly discuss the common results and let the reader fill in not only all the proofs but even most of the statements.

Let A be a right R-module and B a left R-module. Let $P_A \to A \to 0$ and $P_B \to B \to 0$ be projective resolutions. Define $\mathrm{Tor}_*^R(A, B) := H_*(P_A \otimes_R B)$ and $\underline{\mathrm{Tor}}_*^R(A, B) := H_*(A \otimes_R P_B)$.

Just as for Ext_R^*, we see that these abelian groups are independent of projective resolutions, Ext_R^0 is (essentially) Hom_R, and lead to long exact sequences in homology together with naturality between homomorphisms of short exact chain complexes and their homology. We also know that $\mathrm{Tor}_n^R(A, B) = 0$ for $n > 0$ if A is projective and $\underline{\mathrm{Tor}}_n^R(A, B) = 0$ for $n > 0$ if B is projective, since if P is projective, then $P \xrightarrow{1_P} P \to 0$ is a projective resolution of P. Moreover, the coproduct of right (respectively, left) R-modules commutes with Tor_n^* in the left (respectively, right) variable. What is missing is the determination of when all positive homology vanishes and to connect it to flatness. Schanuel's Lemma does not hold for flat modules, but the key to equality of Ext and $\underline{\mathrm{Ext}}$ was dimension shifting. Of course, if R is a commutative ring, the homologly groups constructed are R-modules.

We need to show the following.

Theorem 132.1. *Let F be a right R-module. Then the following are equivalent:*

(1) F is R-flat.
(2) $\underline{\mathrm{Tor}}_n^R(F,B)=0$ for all $n\geq 1$.
(3) $\underline{\mathrm{Tor}}_1^R(F,B)=0$.
(4) $\underline{\mathrm{Tor}}_1^R(F,B)=0$ for all cyclic left modules B.

We have an analogous equivalence for left R-modules.

PROOF. (1) $\Rightarrow$ (2): Let $P_* \xrightarrow{\varepsilon} B \to 0$ be a projective resolution of B. As F is flat, $F\otimes_R$ take exact sequences to exact sequences. In particular, $\cdots \to F\otimes_R P_1 \to F\otimes_R P_0 \xrightarrow{1_F\otimes\varepsilon} F\otimes_R B\to 0$ is exact. So we have $H_n(F\otimes_R P_i)=0$ for all $i\geq 0$ and $n\geq 1$. Therefore, $\underline{\mathrm{Tor}}_n^R(F,B)=0$ for all $n\geq 1$.
(2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (4) are immediate
(4) $\Rightarrow$ (1): We use the criterion given by Corollary 131.11 to establish flatness. Let $0\to\mathfrak{A}\xrightarrow{inc} R\to B\to 0$ be an exact sequence of left R-modules with $\mathfrak{A}$ a finitely generated left ideal. In particular,

$$\underline{\mathrm{Tor}}_1^R(F,B)\to F\otimes_R\mathfrak{A}\xrightarrow{1_F\otimes inc} F\otimes_R R.$$

is exact. By (4), $1_F\otimes inc$ is monic. Therefore, F is flat as tensoring is right exact. □

The theorem says that we can dimension shift using the long exact sequence in homology So from this we can deduce, as before the following:

Theorem 132.2. *Let A be a right R-module and B a left R-module. Then $\mathrm{Tor}_n^R(A,B)=\underline{\mathrm{Tor}}_n^R(A,B)$ for all $n\geq 0$ with $\mathrm{Tor}_0^R(A,B)=A\otimes_R B$.*

Of course, we write $\mathrm{Tor}_n^R(A,B)$ for both of these.

Corollary 132.3. *Let $F_*\to A\to 0$ be a flat R resolution of the right R-module A. Then $\mathrm{Tor}_n^R(A,B)=H_n(F_*\otimes_R B)$ for all R-modules B and $n\geq 0$. We also have if $F'_*\to B\to 0$ is a flat R resolution of the left R-module B. Then $\mathrm{Tor}_n^R(A,B)=H_n(A\otimes_R F'_*)$ for all R-modules B and $n\geq 0$.*

PROOF. Showing dimension shifting only depended on projectivity was used only to show the vanishing of $\mathrm{Tor}_1^R(P,B)$ when P was a projective right R-module. □

Dimension shifting also shows implies

Corollary 132.4. *Let $(F_A)_*\to A\to 0$ be a flat R-resolution of the right R-module A and $F_{B*}\to B\to 0$ a flat R-resolution of the left R-module B. Then the following are true:*

(1) $H_*(F_A\otimes_R B)\cong\mathrm{Tor}_*^R(A,B)\cong H_*(A\otimes_R F_B)$.
(2) $\mathrm{Tor}_*^R(A,B)$ is independent of flat resolutions of either A or B.

In particular, $\operatorname{Tor}(A, B)$ *can be computed using flat resolutions.*

The therefore have a new homological dimension.

Definition 132.5. Define the right (respectively, left) *flat dimension* of a right R-module M to be

$$\operatorname{rfd}_R M := \max\{n \mid \operatorname{Tor}_n^R(M, N) \neq 0 \text{ for all left } R\text{-modules } N\}$$

(or infinity if no minimum exists) and the *right global weak dimension* of R to be

$$\operatorname{rweak\,dim}(R) = \max\{\operatorname{rfd}_R(M) \mid M \text{ a right } R\text{-module}\}$$

(or infinity if no maximum exists). Of course, we have the analogue on for left modules and we denote these by lfd and lweak dim

As projective modules are flat, we have $\operatorname{rweak\,dim}(R) \leq \operatorname{rgl\,proj\,dim}(R)$ and $\operatorname{lweak\,dim}(R) \leq \operatorname{lgl\,proj\,dim}(R) = \operatorname{gl\,dim}(R)$. Since Tor_*^R is determined by flat modules, we can have inequality. However, we do have, in view of Theorem 132.1, which holds for right and left R-modules:

Lemma 132.6. *For any ring* R*, we have* $\operatorname{lweak\,dim}(R) = \operatorname{rweak\,dim}(R)$.

We let $\operatorname{weak\,dim}(R) = \operatorname{lweak\,dim}(R) = \operatorname{rweak\,dim}(R)$ and call it the *weak dimension* of R.

Theorem 132.7. *Let* R *be a commutative Noetherian ring. Then* $\operatorname{weak\,dim}(R) = \operatorname{gl\,dim}(R)$.

Proof. If M is a finitely generated right R-module, then there exists a projective resolution $P_* \to M \to 0$ with all P_i finitely generated, since submodules of a finitely generated submodule is finitely generated (i.e., all finitely generated modules are finitely presented). As projective modules are flat, by Theorem 132.1 we have $\operatorname{lfd}(M) \leq \operatorname{lpd}(M)$. Suppose that $\operatorname{lfd}(M) = n$. Let $N = \ker(P_{n-1} \to P_{n-2})$. Then N must be flat by assumption. As it is finitely generated, it is a finitely generated flat R-module so projective by Corollary 131.16. Therefore, $\operatorname{lpd}(M) \leq \operatorname{lfd}(M)$ also. $\square$

It can be shown if K is field, then $\operatorname{lgl\,dim}(K[t_1, \ldots, t_n]) = n$. This is called the Hilbert Syzygy Theorem. [If the nth kernel of $P_{n-1} \to P_{n-2})$ where $P_* \to M \to 0$ is a projective resolution of M is called an nth syzygy of M.]

Exercises 132.8. 1. Let R be a PID and M an R-module. Show M is flat if and only if M is R-torsion-free..

2. Let M be a finitely presented right R-module. Show M is flat if and only if M is projective.
3. Prove Corollary 132.4.
4. Let $(A_i, \theta_{ij})_I$ be a directed system with I a directed partially order set of right R-modules. Show that $\operatorname{Tor}_n^R(\varinjlim A_i, B) \cong \varinjlim \operatorname{Tor}_n^R(A_i, B)$ for every left R-module B. We have an analogous result for left R-modules A_{iI}
5. Let M be a right R-module. Show M is flat if and only if $\operatorname{Tor}_1^R((M, R/\mathfrak{A}) = 0$ for all left ideals $\mathfrak{A}$ in R.
6. Let R be a ring. Show that the following are equivalent:

(i) All left ideals of R are flat.
(ii) All right ideals of R are flat.
(iii) $\operatorname{weak}\dim(R) \leq 1$.
(iv) $\operatorname{Tor}^2_R(M, N) = 0$ for all right R-modules M and left R-modules N.

133. Regular Local Rings II

We wish to give an application of global dimension to ring theory. Our main motivation is to prove that a regular local ring is a UFD without the characteristic assumption that we made before. The proof of this showed the power of using homological algebra in commutative algebra. Much of our development is also applicable to all rings, not just commutative ones.

We know if $\varphi : R \to S$ be a ring homomorphism and N an S-module, then the pull back via $rm := \varphi(r)n$ for $r \in R$, $n \in N$, makes M into an R-module. We leave it as an exercise to show that the pullback takes exact sequences of S-modules to exact sequences of R-modules.

Theorem 133.1. (General Change of Rings Theorem) *Let $\varphi : R \to S$ be a ring homomorphism and M an S-module. Then*

$$\operatorname{lpd}_R M \leq \operatorname{lpd}_R S + \operatorname{lpd}_S M.$$

Proof. We may assume that the right hand side is finite. Let $n = \operatorname{lpd}_R S$ and $i = \operatorname{lpd}_S M$. We induct on i.

$i = 0$: In this case we have M is S-projective, so there exists an S-module M' satisfying $M' \coprod M$ is S-free. The projective dimension of a direct sum of modules is equal to the supremum of the projective dimension of its direct summands (which we leave as any exercise). In particular,

$$\operatorname{lpd}_R M \leq \operatorname{lpd}_R(M' \coprod M) = \operatorname{lpd}_R S = \operatorname{lpd}_R S + \operatorname{lpd}_S M.$$

$i > 0$: In this case we know that M is not S-projective. Choose a free S-module P satisfying $0 \to N \to P \to M \to 0$ is an exact sequence of S-modules. In particular, the pullback of this sequence is an exact sequence of R-modules. Since P is S-free, and M not S-projective, we know that $\operatorname{lpd}_S M = \operatorname{lpd}_S N + 1$ by Exercises 128.30(128.30) and 128.30(3). By induction $\operatorname{lpd}_R N \leq \operatorname{lpd}_R S + \operatorname{lpd}_S N = n+i-1$. Since P is a free S-module, it follows that $\operatorname{lpd}_R P = \operatorname{lpd}_R S = n$. We apply Exercises 128.30(128.30) and 128.30(3).

$$\begin{aligned}
&\text{If } \operatorname{lpd}_R N > \operatorname{lpd}_R P, \text{ then } \operatorname{lpd}_R M = \operatorname{lpd}_R N + 1 \leq n + i\\
&\text{If } \operatorname{lpd}_R N = \operatorname{lpd}_R P, \text{ then } \operatorname{lpd}_R M \leq \operatorname{lpd}_R P + 1 \leq n + 1 \leq n + i\\
&\text{If } \operatorname{lpd}_R N < \operatorname{lpd}_R P, \text{ then } \operatorname{lpd}_R M = \operatorname{lpd}_R P + 1 = n + 1 \leq n + i.
\end{aligned}$$

The result follows. □

Examples 133.2. Let R be a ring.

1. $R[t]$ is a R-module with t central. Therefore, any $R[t]$-module satisfies $\operatorname{lpd}_R M \leq \operatorname{lpd}_{R[t]} M$ be the theorem. But if M is an R-module, then as an $R[t]$-module, $tm = 0$ for all $m \in M$. It follows that $\operatorname{lpd}_R M \leq \operatorname{lpd}_{R[t]} M$. In particular $\operatorname{lgl}\dim R \leq \operatorname{lgl}\dim R[t]$.

2. Let $x \in R$ be a central and not a zero divisor in R and $^{-} : R \to R/(x)$ be the canonical ring epimorphism where $R/(x) = R/RxR$. Then $\operatorname{lpd}_R M \leq \operatorname{lpd}_{\overline{R}} +1$.

   PROOF. We may assume that $\operatorname{lpd}_{\overline{R}} M = n < \infty$. By the General Change of Rings Theorem 133.1, it suffices to show $\operatorname{lpd}_R \overline{R} \leq 1$. Since x is not a zero divisor, the sequence $0 \to R \xrightarrow{x} (x) \to 0$ of R-modules is exact, hence $(x) \cong R$ is R-free. Therefore, the exact sequence $0 \to (x) \to R \to \overline{R} \to 0$ of $\overline{R}$-modules shows that $\operatorname{lpd}_R \overline{R} \leq 1$. $\square$

3. Let $(R, \mathfrak{m})$ be a commutative local ring with $0 \neq \mathfrak{m} = (x)$ satisfying $x^n = 0$ but $x^{n-1} \neq 0$ for some $n > 1$. Then $\operatorname{pd}_R R/\mathfrak{m} = \infty$. In particular, $\operatorname{gl\,dim} R = \infty$.

   PROOF. As $\mathfrak{m} = (x)$, our hypothesis implies that $\operatorname{ann}_R(\mathfrak{m}) = (x^{n-1})$ and $\operatorname{ann}_R(x^{n-1}) = (x) = \mathfrak{m}$. Therefore, we have exact sequences of R-modules:

$$0 \to \mathfrak{m} \to R \xrightarrow{-} R/\mathfrak{m} \to 0$$
$$0 \to (x^{n-1}) \to R \xrightarrow{x} R/\mathfrak{m} \to 0$$
$$0 \to \mathfrak{m} \to R \xrightarrow{x^{n-1}} R/(x^{n-1}) \to 0.$$

   Hence, we have $\mathcal{P}(R/\mathfrak{m}) = [\mathfrak{m}]$, $\mathcal{P}^2(R/\mathfrak{m}) = [(x^{n-1})]$, $\mathcal{P}^3(R/\mathfrak{m}) = [\mathfrak{m}]$, and $\mathcal{P}^{n+2}(R/\mathfrak{m}) = \mathcal{P}^n(R)/\mathfrak{m})$ for all n. Since R is local, any finitely generated projective R-module is R-free by Lemma 127.2. But $(y) < R$ cannot be R-free if $\operatorname{ann}_R y > 0$. So $\mathfrak{m}$ and (x^{n-1}) are not free. It follows that the sequence $R \to R/\mathfrak{m} \to 0$ cannot be extended to a finite projective resolution. $\square$

   Note: We shall see below that if x is also a non-unit in the above, that $\operatorname{pd}_R \overline{R} \leq 1$.

4. Let R be a discrete valuation ring that is not a field. Then $\mathfrak{m} = (x) > 0$. As $\mathfrak{m}$ is R-free, $\operatorname{pd}_R R/\mathfrak{m} = 1$. Let $^{-} : R \to R/(x^2)$. Then by the previous example, $\operatorname{lgl\,dim} \overline{R}/\overline{\mathfrak{m}} = \infty$. So we can have inequality in Example 2.

Theorem 133.3. (First Change of Rings Theorem) *Let R be a ring and $x \in R$ a central element that is not a zero-divisor. Let $^{-} : R \to R/(x)$ be the canonical ring epimorphism and $0 \neq M$ an $\overline{R}$-module (so $x \notin R^{\times}$) satisfying $\operatorname{lpd}_{\overline{R}} M$ is finite. Then $\operatorname{lpd}_R M = 1 + \operatorname{lpd}_{\overline{R}} M$.*

PROOF. By Example 133.2(2), we have $\operatorname{lpd}_R \leq 1 + \operatorname{lpd}_{\overline{R}} M$, so we need only show the converse inequality. Let $n = \operatorname{lpd}_{\overline{R}} M$. We prove the result by induction on n.

$n = 0$: We show that M cannot be R-projective, i.e., $\operatorname{lpd}_R M = 1$. In particular, this establishes the Note in 133.2(2):

If $x \notin R^{\times}$, then left multiplication $\lambda_x : R \to R$ by $y \mapsto xy$ in a monomorphism of R-modules. It follows that $\lambda_x : P \to P$ by $y \mapsto xy$ is a monomorphism for all free R-modules P and hence all projective R-modules P. We are given that the $\overline{R}$-module M is not zero. Since $xM = 0$, M cannot be R-projective.

$n = 1$: Suppose that $\operatorname{lpd}_R M \leq 1$. Then we have an exact sequence of R-modules $0 \to P \to F \xrightarrow{\varphi} M \to 0$ with F a free R-module and with $P \subset F$ and P a projective R-module. Since $xM = 0$, we have $xP \subset F$. In particular, φ induces a sequence

$$(*) \qquad 0 \to P/xF \to F/xF \xrightarrow{\overline{\varphi}} M \to 0$$

of R-modules, hence of $\overline{R}$-modules, as $xF \subset xP$. Moreover, the sequence (*) is exact, since $\varphi(z) \in xM = 0$ if and only if $z \in P$. We know that $\overline{R}\otimes_R$ takes R-free (respectively, R-projective) modules to $\overline{R}$-free (respectively, $\overline{R}$-projective) modules. Therefore, $\overline{R}\otimes P = P/xP$ is $\overline{R}$-projective and $F = F/xF$ is $\overline{R}$-free. Since $\operatorname{lpd}_{\overline{R}} M = 1$, P/xP must be $\overline{R}$-projective. As $P \subset F$ implies $xP \subset xF$, the exact sequence

$$0 \to F/xP \to P/xP \to P/xF \to 0$$

must be a split exact sequence of $\overline{R}$-modules. In particular, xF/xP must also be $\overline{R}$-projective. But we also have $\lambda_x : F \to F$ by $y \mapsto xy$ is a monomorphism of R-modules. It follows that $M \cong F/P \cong xF/xP$ is projective. This contradicts our hypothesis and completes the $n = 1$ case.

$n > 1$: Choose a free $\overline{R}$-module F and an exact sequence $0 \to N \to F \to M \to 0$ of $\overline{R}$-modules. Then $\operatorname{lpd}_{\overline{R}} N = n-1$. By the $n = 0$ case, $\operatorname{lpd}_R = 1$. Hence $\operatorname{lpd}_R N = n-1$ and $\operatorname{lpd}_R M = \operatorname{lpd}_R N + 1 = n+1$ by induction. □

Corollary 133.4. *Let R be a ring. Then* $\operatorname{lgl\,dim} R[t] \geq \operatorname{lgl\,dim} R + 1$.

PROOF. Apply the theorem to $R[t] \to R[t]/(t) = R$. □

Notation 133.5. Let M be an R-module. Set $M[t] := R[t] \otimes_R M$, an $R[t]$-module.

Lemma 133.6. *Let M be an R-module. The* $\operatorname{lpd}_{R[t]} M[t] = \operatorname{lpd}_R M$.

PROOF. $R[t]$ is R-free so R-flat. In particular, $R[t] \otimes_R$ takes R-free modules to $R[t]$-free modules, hence R-projective modules to $R[t]$-projective modules. Therefore, $R[t] \otimes_R$ takes projective resolutions of an R-module M to projective resolutions of $R[t]$-modules $M[t]$. In particular, $\operatorname{lpd}_R M \geq \operatorname{lpd}_{R[t]} M[t]$. Suppose that $n < \operatorname{lpd}_R M$. We must show if $\mathcal{P}$ is the projection operator, then $\mathcal{P}^n M[t] \neq 0$ in $R[t]$. This is equivalent to showing that M is a projective R-module if $M[t]$ is a projective $R[t]$-module. So assume that $M[t]$ is $R[t]$-projective. Since $R[t]$ is R-free. $R[t] \cong R^J$ for some J. Therefore, $M[t] \cong M^J$ as R-modules. Since $M[t]$ is a direct summand of $R[t]^I$, some I, M is a direct some of R^L, for some L. Therefore, M is projective. □

Theorem 133.7. $\operatorname{lgl\,dim} R[t] = 1 + \operatorname{lgl\,dim} R$.

PROOF. Let M be an $R[t]$-module. By Corollary 133.4, we need only show that $\operatorname{lpd}_{R[t]} M \leq \operatorname{lgl\,dim} R + 1$. To do this we construct the generalization of the characteristic sequence done in **??**, i.e., we construct an exact sequence

$$0 \to M[t] \xrightarrow{\psi} M[t] \xrightarrow{\varphi} M \to 0. \tag{*}$$

If we show that (*) is exact then the lemma implies

$$\operatorname{lpd}_{R[t]} M \leq 1 + \operatorname{lpd}_{R[t]} M[t] = 1 + \operatorname{lpd}_R M \leq 1 + \operatorname{lgl\,dim} R$$

and we would be done.

Define

$$\psi : M[t] \to M[t] \ \text{ by } \ t^i \otimes m \mapsto t^i(1 \otimes t - t \otimes 1)m$$

$$\varphi : M[t] \to M \ \text{ by } \ t^i \otimes m \mapsto t^i m.$$

In particular,

$$\psi(\sum_{i=0}^{n} t^i \otimes m_i) = 1 \otimes tm_0 + \sum_{i=1}^{n} t^i \otimes (tm_i - m_{i-1}) - t^{n+1} \otimes m_n.$$

Clearly, $\operatorname{im}\psi \subset \ker\varphi$. So suppose that $\psi(\sum_{i=0}^{n} t^i \otimes m_i) = 0$. Since $M[t] = \coprod Rt^i \otimes_R M$ as R-modules, we have

$$0 = -m_n = tm_n - m_{n-1} = \cdots = tm_1 - m_0.$$

It follows that $m_i = 0$ for all i and hence ψ is a monomorphism. To finish, we must show that $\operatorname{im}\psi = \ker\varphi$. Let $x \in \ker\varphi$, say $x = \sum t^i \otimes v_i$, so $\sum t^i v_i = \psi(x) = 0$. Recursively define $m_{n-1}, \dots, m_0$ in M by

$$\begin{aligned} v_0 &:= -tm_0 \\ v_1 &:= tm_1 - m_0 \\ &\vdots \\ v_{n_1} &:= tm_{n_1} - m_{n-2} \\ v_n &:= -m^{n-1}. \end{aligned}$$

Then $\psi(\sum t^i \otimes m_i) = x$. $\square$

Corollary 133.8. *Let R be an Artinian semi-simple ring. Then* $\operatorname{lgl\,dim} R[t_1, \dots, t_n] = n$

Corollary 133.9. (Hilbert Syzygy Theorem (Weak Form)) *Let K be a field. Then* $\operatorname{gl\,dim} K[t_1, \dots, t_n] = n$.

The full Hilbert Syzygy Theorem says that any $K[t_1, \dots, t_n]$-module has an acyclic resolution of length at most n by free R-modules.

Corollary 133.10. *Let R be a PID that is not a field. Then* $\operatorname{gl\,dim} R[t_1, \dots t_n] = n + 1$.

Proof. Every submodule of an R-free if free. $\square$

Theorem 133.11. (Second Change of Rings Theorem) *Let R be a ring and $x \in R$ be a central non-unit and not a zero-divisor. Let $^- : R \to R/(x)$ be the canonical epimorphism and M an R-module. If x is not a zero-divisor of M, then* $\operatorname{lpd}_{\overline{R}} M/xM \leq \operatorname{lpd}_R M$.

Proof. We may assume that $n = \operatorname{lpd}_R M < \infty$. We induct on n.

$n = 0$: We have $M/xM = \overline{R} \otimes_R M$. Therefore, if M is R-free (respectively R-projective), then so is M/xM.

$n > 0$: Choose an exact sequence of R-modules

$$0 \to N \hookrightarrow F \xrightarrow{\varphi} M \to 0 \tag{1}$$

with F R-free. Since $\overline{R} \otimes_R$ is right exact, we have a commutative exact diagram

$$\begin{tikzcd} \overline{R} \otimes_R N \arrow[r] \arrow[d, "\text{nat}"', "\cong"] & \overline{R} \otimes_R F \arrow[r, "1\otimes\varphi"] \arrow[d, "\text{nat}"', "\cong"] & \overline{R} \otimes_R M \arrow[r] \arrow[d, "\text{nat}"', "\cong"] & 0 \\ N/xN \arrow[r] & F/xF \arrow[r, "\overline{\varphi}"'] & M/xM \arrow[r] & 0. \end{tikzcd}$$

As x is not a zero divisor in R, x is not a zero divisor in F. Hence x not a zero divisor in $N \subset F$.

Claim. The sequence of R-modules

$$(2) \qquad 0 \to N/xN \to F/xF \xrightarrow{\overline{\varphi}} M/xM \to 0$$

is exact:

Since $\overline{R} \otimes_R \;$ is right exact, we need only show that $N/xN \to F/xF$ is monic. To do this it suffices to show $N \cap xF = xN$. It is clear that $N \cap xF \supset xN$. We show the opposite inclusion. Let $y \in N \cap xF$. Then $y = xf$ for some $f \in F$ and $0 = \varphi(y) = x\varphi(f)$. Since x is not a zero divisor on M, $\varphi(f) = 0$, i.e., $y \in xN$. This proves the Claim.

If $\operatorname{lpd}_R M = 1$ equation (1) is a projective resolution of M. Therefore, by the $n = 0$ case, equation (2) is a projective resolution for M/xM. So we may assume that $n > 1$. By equation (1), $\operatorname{lpd}_R N = n - 1$. By induction, $\operatorname{lpd}_{\overline{R}} N/xN \le \operatorname{lpd}_R N = n - 1$. Therefore, by equation (2) we have $\operatorname{lpd}_{\overline{R}} M/xM \le \operatorname{lpd}_{\overline{R}} N/xN + 1 \le n$. □

Remark 133.12. In the above, $\operatorname{lpd}_R \overline{R} = 1$. It follows that $\operatorname{Tor}_i^R(\overline{R},\ M) = 0$ for all R-modules M and for all $i > 1$. This means that x is a zero divisor on an R-module M if and only if $\operatorname{Tor}_1^R(\overline{R}, M) = 0$:

Proof. ($\Rightarrow$): This follows from the exactness of equation (1) implies the exactness of equation (2).
($\Leftarrow$): Applying the long exact sequence of Tor on the short exact sequence $0 \to (x) \to R \to \overline{R} \to 0$ yields this. □

We need the analog of the Jacobson radical and Nakayama's Lemma in the non-commutative case. We leave the proofs as exercises.

Theorem 133.13. *Let R be a ring and $a \in R$. Then the following are equivalent.*

(1) If M is a simple R-module, then $aM = 0$.
(2) a belongs to every maximal left ideal of R.
(3) $1 - ya$ has a left inverse for all $y \in R$.
(4) $1 - xay$ has an inverse for all $x, y \in R$.

The right analogs of the above hold.

We let $J(R)$ be the set of elements satisfying the theorem. It is called the *Jacobson radical* of R.

Lemma 133.14. *Let R be ring. Then $J(R)$ is a two-sided ideal. Moreover, it is the intersection of all left maximal ideals in R as well as the intersection of all right maximal ideals in R.*

The analogue of Nakayama's Lemma also holds.

Lemma 133.15. (Nakayama's Lemma) *Let M be a nonzero finitely generated R-module. Then $J(R)M \neq M$ and if $N \subset M$ is a submodule satisfying $M = N + J(R)M$, then $M = N$.*

Theorem 133.16. (Third Change of Rings Theorem) *Let R be a left Noetherian ring and $J(R) = \cap\mathfrak{m}$, the intersection of all left maximal ideals in R. Suppose that $x \in J(R)$ is central (so not a unit) and not a zero divisor in R and $^{-} : R \to R/(x)$, the canonical ring epimorphism. Let M be a nonzero finitely generated R module such that x is not a zero divisor of M. Then $\operatorname{lpd}_R M = \operatorname{lpd}_{\overline{R}} M/xM$.*

Proof. By the Second Change of Rings Theorem 133.11, $\operatorname{lpd}_R M \geq \operatorname{lpd}_{\overline{R}} M/xM$, so we need only show that $\operatorname{lpd}_R M \leq \operatorname{lpd}_{\overline{R}} M/xM$. We may assume that $n = \operatorname{lpd}_{\overline{R}} M/xM < \infty$. Since $x \in J(R)$, we know that $xM < M$ by Nakayama's Lemma 133.15. We induct on n.

$n = 0$: We first show if $\overline{M} = M/xM$ is $\overline{R}$-free, then M is R-free. Let $^{-} : M \to M/xM$. Choose $v_1, \dots, v_m$ in M so that $\mathcal{B} = \{\overline{v}_1, \dots, \overline{v}_m\}$ is a basis for $\overline{M}$. We show that $\{v_1, \dots, v_m\}$ is a basis for M. It spans by Nakayama's Lemma 133.15. Suppose that $\sum_{i=0}^m a_{i1}v_i = 0$ with $a_{i1}, \dots, a_{m1} \in R$. Since $\mathcal{B}$ is a basis for $\overline{M}$, we have $a_{i1} \in (x)$, i.e., $a_{i1} = xa_{i2}$ for some $a_{i2} \in R$, as x is central in R, for $i = 1, \dots, m$. So $x(\sum_{i=0}^m a_{i2}v_i) = 0$. As x is not a zero divisor on M, we must have $\sum_{i=1}^n a_{i2}v_i = 0$. Iterating this process, we find $a_{i3}, a_{i4}, \dots$ satisfying the ascending chain $Ra_{i1} \subset Ra_{i2} \subset Ra_{i3} \subset \cdots$ for $i = 1, \dots, m$. As R is left Noetherian, this chain stabilizes. So here exists a j such that $Ra_{ij} = Ra_{i,j+1}$ and by construction $a_{ij} = xa_{i,j+1}$ for $i = 1, \dots, m$. Let $a_{i,j+1} = y_i a_{ij}$ for some $y_i \in R$ for $i = 1, \dots, m$. It follows that $(1 - xy_i)a_{ij} = 0$ for $i = 1, \dots, m$. Since $x \in J(R)$, we have $1 - xy_i \in R^{\times}$ by Theorem 133.13. It follows that $a_{ij} = 0$ for $i = 1, \dots, m$. This shows that if $\overline{M}$ is $\overline{R}$-free, then M is R-free.
Next assume that $\overline{M}$ is $\overline{R}$-projective. As in the proof of the Second Change of Rings Theorem 133.11, we have exact sequences

$$0 \to N \hookrightarrow F \xrightarrow{\varphi} M \to 0 \tag{1}$$

$$0 \to N/xN \to F/xF \xrightarrow{\overline{\varphi}} M/xM \to 0 \tag{2}$$

with F a finitely generated free R-module (as M is finitely generated) and x not a zero divisor on N. Since R is left Noetherian, N is also finitely generated. Since $\overline{M}$ is $\overline{R}$-projective, equation (2) splits. Therefore, $F/xF \cong N/xN \coprod M/xM$ is $\overline{R}$-free. Set $A = M \coprod N$. Then A is a finitely generated R-module satisfying $A/xA \cong F/xF$ is $\overline{R}$-free. Therefore, by the first part of the proof, A is R-free. Hence P is R-projective as needed.
Suppose that $n > 0$. As above we get exact sequences (1) and (2). If $n = 1$, then N/xN is $\overline{R}$-projective, so N is R-projective by the $n = 0$ case, hence $\operatorname{lpd}_R M \leq 1$. It $n > 1$, then by induction $n - 1 = \operatorname{lpd}_R N = \operatorname{lpd}_{\overline{R}} N/xN$. Hence $\operatorname{lpd}_R M \leq M$, as needed. $\square$

Theorem 133.17. *Let R be a left Noetherian ring, $x \in R$ is central and not a zero-divisor, and $^{-} : R \to R/(x)$, the canonical ring epimorphism. Suppose that $x \in J(R)$ and $\operatorname{lgl\,dim} \overline{R} = n < \infty$. Then $\operatorname{lgl\,dim} R = n + 1$.*

Proof. Let $\mathfrak{A} \subset R$ be a left ideal. It is finitely generated as R is left Noetherian. Since x is not a zero divisor on R, it is not a zero divisor on $\mathfrak{A}$, so by the Third Change of Rings Theorem 133.16, $\operatorname{lpd}_R \mathfrak{A} = \operatorname{lpd}_{\overline{R}} \mathfrak{A}/x\mathfrak{A} \leq \operatorname{lgl\,dim} \overline{R} = n$. As $0 \to \mathfrak{A} \to R \to R/\mathfrak{A} \to 0$ is an exact sequence of R-modules, $\operatorname{lpd}_R R/\mathfrak{A} \leq 1 + \operatorname{lpd}_R \mathfrak{A} \leq n + 1$. Since the reverse inequality holds by the First Change of Rings Theorem 133.3, the result follows by Auslander's Theorem **??** □

We now look at the case when R is a commutative ring, in particular the case of local Noetherian rings. We use results that we proved in commutative algebra.

Definition 133.18. Let R be a commutative ring. An ordered sequence of elements $x_1, \dots, x_n$ in R is called an *R-sequence* of *length* n if $(x_1, \dots, x_n) \neq R$ and $x_{i+1} \notin \mathrm{zd}\big(R/(x_1, \dots x_r)\big)$ for all $i < n$. The *depth* of R is defined to be

$$\operatorname{depth} R := \max\{n \mid x_1, \dots, x_n \text{ is an } R\text{-sequence in } R\}.$$

A Noetherian local ring $(R, \mathfrak{m})$ is called *Cohen-Macaulay* or a *CM* ring if $\operatorname{depth} R = \dim R$. Note that if $(R, \mathfrak{m})$ is a local ring, then any R-sequence lies in $\mathfrak{m}$.

In general, the order of the elements in an R-sequence is crucial. However, it turns out that for a local Noetherian ring it does not matter. We will not prove this, nor need it.

Proposition 133.19. *Let R be a commutative ring and $\mathfrak{P}$ be a prime ideal in R. Suppose that $x_1, \dots, x_n$ is an R-sequence lying in $\mathfrak{P}$. Then $\operatorname{ht} \mathfrak{P} \geq n$. In particular, $\dim R \geq \operatorname{depth} R$.*

Proof. Since $x_1 \notin \mathrm{zd}(R)$, $x_1 \notin \mathfrak{p}$ for any $\mathfrak{p} \in \mathrm{Min}(R)$. In particular, $\operatorname{ht} \mathfrak{P} \geq 1$. Let $^{-} : R \to R/(x_1)$ be the canonical epimorphism. Then $\overline{x}_2, \dots, \overline{x}_n$ is an $\overline{R}$-sequence by the Third Isomorphism Theorem. By induction, $\operatorname{ht} \overline{\mathfrak{P}} \geq n - 1$. Hence $\operatorname{ht} \mathfrak{P} \geq n$ by the Correspondence Principle. The result follows. □

Corollary 133.20. *Let $(R, \mathfrak{m})$ be a Noetherian local ring. Then $\operatorname{depth} R \leq \dim R \leq V\text{-}\dim R$.*

Corollary 133.21. *Let $(R, \mathfrak{m})$ be a Noetherian local ring. If $\mathfrak{m}$ can be generated by an R-sequence of length d, then R is a regular local ring (and Cohen-Macaulay) of dimension d.*

Proposition 133.22. *Let $(R, \mathfrak{m})$ be a regular local ring. Then R is a Cohen-Macaulay ring and any minimal generating set for $\mathfrak{m}$ is an R-sequence.*

Corollary 133.23. *Let $(R, \mathfrak{m})$ be a regular local ring and $x_1, \dots, x_d$ a minimal generating set for $\mathfrak{m}$. Then $R/(x_1, \dots, x_i)$ is a regular local ring of dimension $d - i$.*

Theorem 133.24. *Let $(R, \mathfrak{m})$ be a regular local ring. Then $\dim R = \operatorname{gl}\dim R$*

Proof. Let $d = \dim R$. If $d = 0$, R is a field and the result follows. So suppose $d > 0$ and let $x_1, \dots, x_d$ be a minimal generating set for $\mathfrak{m}$. Then $R/(x_1)$ is regular of dimension $d - 1$ and $x_1 \notin \mathrm{zd}(R)$. Since $x_1 \in J(R) = \mathfrak{m}$, $\operatorname{gl}\dim R = (d - 1) + 1 = d$ by Theorem 133.17 and induction. □

We want to prove the converse of this theorem. This is a famous theorem of Serre. We need two lemmas.

Lemma 133.25. *Let $(R, \mathfrak{m})$ be a Noetherian local ring and M a finitely generated R-module with $M = Rm_1 + \cdots + Rm_n$ and n minimal. Let F be an R-free module on basis $\{v_1, \cdots, v_n\}$ and $\varphi : F \to M$ the R-homomorphism defined by $v_i \mapsto m_i$ for $i = 1, \ldots, n$. Then $\ker\varphi \subset \mathfrak{m}F \cong \coprod_{i=1}^n \mathfrak{m}$.*

PROOF. If the result is false then there exist $c_i \in R$, $i = 1, \ldots, n$, satisfying $\sum_{i=1}^n c_i v_i \in \ker\varphi \setminus \mathfrak{m}F$, i.e., some $c_i \notin \mathfrak{m}$. But then $c_i \in R^\times$, contradicting the minimality of n. □

Definition 133.26. Let R be a commutative ring $M = Rm_1 + \cdots + Rm_n$ with n minimal, F a free R-module of rank n. Then an exact sequence $0 \to K \to F \to M \to 0$ is called a *minimal presentation* of M.

Using the Primary Decomposition for ideals in a commutative Noetherian ring, we establish the following lemma.

Lemma 133.27. *Let $(R, \mathfrak{m})$ be a Noetherian local ring. Suppose that $\mathfrak{m} \subset \mathrm{zd}(R)$ and M is a finitely generated R-module. Then either $\mathrm{pd}_R M = 0$ or $\mathrm{pd}_R M = \infty$.*

PROOF. Suppose that there exists a finitely generated R-module M satisfying $0 < \mathrm{pd}_R M < \infty$. By dimension shifting (Corollary 128.24), we may assume that there exists a finitely generated R-module N such that $\mathrm{pd}_R N = 1$. Therefore, we may assume that $M = N$. Let $0 \to K \to R^n \to M \to 0$ be a minimal presentation of M (some n). Since $\mathrm{pd}_R M = 1$ and $0 \neq K$, K must be R-projective. As R is a local Noetherian ring, K must be R-free by Lemma 127.2. We know that $\mathfrak{m} \subset \mathrm{zd}(R) = \bigcup_{\mathrm{Ass}_R(0)} \mathfrak{p}$ and $|\mathrm{Ass}_R(0)| < \infty$ by Proposition 94.15. Therefore, by the Prime Avoidance Lemma 93.16 and the maximality of $\mathfrak{m} \in \mathrm{Ass}_R(0)$, we see that there exists a nonzero $x \in R$ satisfying $\mathfrak{m} = \mathrm{ann}_R(x)$. Consequently, $x\mathfrak{m} = 0$. By Lemma 133.25 above, $K \subset \mathfrak{m}R^n$. So $xK = 0$. This contradicts K is a nonzero free R-module. □

Theorem 133.28. (Serre) *Let $(R, \mathfrak{m})$ be a Noetherian local ring that is not a field. Suppose that $\mathrm{pd}_R \mathfrak{m} = n < \infty$. Then R is a regular local ring of $\dim R = \mathrm{gl}\dim R = n + 1$. In particular, R is regular if and only if $\mathrm{gl}\dim R \leq \infty$.*

PROOF. We induct on n. We first show that R is a regular local ring.

$n = 0$: Since $\mathfrak{m}$ is R-projective, it is R-free by by Lemma 127.2. But an ideal of R is R-free if and only if it is a principal ideal. As R is not a field, $\mathfrak{m} = (x)$ for some nonzero $x \in R$ and x is not a zero divisor in R. In particular, x is an R-sequence. It follows that R is a regular local ring of dimension one.

$n > 0$: Let $d = \dim R < \infty$ (by Corollary 97.22 of the Principal Ideal Theorem 97.19). By Lemma 133.27 above with $M = \mathfrak{m}$, $0 < \mathfrak{m}$ does not lie in $\mathrm{zd}(R)$. If $d = 0$, then $\mathfrak{m} \in \mathrm{Min}(R)$ using Proposition 94.14. But this means that $\mathfrak{m} \subset \mathrm{zd}(R)$ which is impossible. Therefore, we must have $d > 0$.

Claim. $\mathfrak{m} \setminus \mathfrak{m}^2 \not\subset \mathrm{zd}(R) = \bigcup_{\mathrm{Ass}_R(0)} \mathfrak{p}$:

Suppose the claim is false. By Nakayama's Lemma 93.10, we have $\mathfrak{m}^2 < \mathfrak{m}$. By Proposition 94.15, $|\mathrm{Ass}_R(0)| < \infty$. Consequently, there exists a prime ideal $\mathfrak{m} \subset \mathfrak{p}$ for some $\mathfrak{p} \in$

$\mathrm{Ass}_R(0)$ by the Prime Avoidance Lemma 93.16. Therefore, $\mathfrak{m} = \mathfrak{p} \in \mathrm{Ass}_R(0)$, so $\mathfrak{m} \subset \mathrm{zd}(R)$ which is a contradiction. This proves the claim.

Therefore, there exists $x \in \mathfrak{m} \setminus \mathfrak{m}^2$ with $x \notin \mathrm{zd}(R)$. Let $^- : R \to R/(x)$ be the canonical epimorphism. As every minimal prime lies in $\mathrm{zd}(R)$ by Proposition 94.14, x lies in no minimal prime. Therefore, $\dim \overline{R} = \mathrm{ht}\, \overline{\mathfrak{m}} = \mathrm{ht}\, \mathfrak{m} - 1 = \dim R - 1 = d - 1$. If $\overline{R}$ is a field, then R is a regular local ring, hence a domain, of dimension one, so $\mathfrak{m} = (x)$ is R-free generated by an R-sequence. In particular, $\mathrm{pd}_R\, \mathfrak{m} = 0$ and $\mathrm{gl\,dim}\ R = 1$, so we are done. Therefore, we may assume that $d > 1$, or if $d = 1$, then $\overline{R}$ is not a field.

Claim. As $\overline{R}$-modules, $\mathfrak{m}/(x)$ is isomorphic to a direct summand of the $\overline{R}$-module $\mathfrak{m}/x\mathfrak{m}$:

[To avoid confusion, we will not use $^-$.] Since $x \notin \mathfrak{m}^2$, there exists a minimal generating set $x, y_1, \dots, y_r$ for $\mathfrak{m}$. Set

$$S = x\mathfrak{m} + (y_1, \dots, y_r).$$

So $\mathfrak{m} = S + (x)$ and clearly $S \cap (x) \supset x\mathfrak{m}$. We shall show that $S \cap (x) = x\mathfrak{m}$. Let $z \in S \cap (x)$ and write

$$a = ax = b_1 y_1 + \cdots b_r y_r + cx$$

with $a, b_1, \dots, b_r \in R$ and $c \in \mathfrak{m}$. Going modulo $\mathfrak{m}^2$, we see that $a - c, b_1, \dots, b_r$ all lie in $\mathfrak{m}$. Therefore, $z = ax \in x\mathfrak{m}$ and $S \cap (x) = x\mathfrak{m}$ as needed. Thus we have

$$\frac{\mathfrak{m}}{x\mathfrak{m}} = \frac{S + (x)}{S \cap (x)} = \frac{S}{S \cap (x)} + \frac{(x)}{S \cap (x)}$$

and (check)

$$\frac{S}{S \cap (x)} \cap \frac{(x)}{S \cap (x)} = 0.$$

So

$$\frac{\mathfrak{m}}{x\mathfrak{m}} = \frac{S}{S \cap (x)} \oplus \frac{(x)}{S \cap (x)}.$$

As

$$\frac{(S)}{S \cap (x)} = \frac{S + (x)}{(x)} = \frac{\mathfrak{m}}{(x)},$$

the claim follows.

By the Third Change of Rings Theorem 133.16, we have $\mathrm{pd}_{\overline{R}}\, \mathfrak{m}/x\mathfrak{m} = \mathrm{pd}_R\, \mathfrak{m} = n$. By the Claim, $\mathrm{pd}_{\overline{R}}\, \mathfrak{m}/(x) \leq \mathrm{pd}_{\overline{R}}\, \mathfrak{m}/x\mathfrak{m} = n$ by Exercise **??**(2). By induction on d, if $\overline{\mathfrak{m}} = \mathfrak{m}/(x)$, then $(\overline{R}, \overline{\mathfrak{m}})$ is a regular local ring of dimension $d-1$, so $d > 1$, and $\mathrm{pd}_{\overline{R}}\, \overline{\mathfrak{m}} = d-2$. We know that x is an R-sequence and $\mathfrak{m}$ can be generated by the $\overline{R}$-sequence $\overline{y}_1, \dots \overline{y}_{d-1}$, where $y_1, \dots, y_{d-1}$ lie in $\mathfrak{m}$. It follows easily that $x, y_1, \dots, y_{d-1}$ is an R-sequence generating $\mathfrak{m}$. Therefore, $(R, \mathfrak{m})$ is a regular local ring of dimension d.

To finish, we must show $d = n + 1$. Since R is a regular local ring, it is a domain by Theorem 98.10. In particular (x) is R-free, so $\mathrm{pd}_R(x) = 0$. Since $\mathrm{pd}_R\, \mathfrak{m} > 0$, the exact sequence

$$0 \to (x) \to \mathfrak{m} \to \mathfrak{m}/(x) \to 0$$

implies that $\mathrm{pd}_R\, \mathfrak{m} = \mathrm{pd}_R\, \overline{\mathfrak{m}}$ where $\overline{\mathfrak{m}} = \mathfrak{m}/(x)$. By the First Change of Rings Theorem 133.3, $\mathrm{pd}_R\, \overline{\mathfrak{m}} = 1 + \mathrm{pd}_{\overline{R}}\, \overline{\mathfrak{m}} = d - 1$. Therefore, $\mathrm{pd}_R\, \mathfrak{m} = d - 1$ as required. $\square$

Remark 133.29. Note that Serre's Theorem says that a Noetherian local ring $(R, \mathfrak{m})$ has finite global dimension if and only if $\mathrm{pd}_R\, \mathfrak{m}$ is finite.

Corollary 133.30. ;f (Krull's conjecture) *Let* $(R, \mathfrak{m})$ *be a regular local ring and* $\mathfrak{p}$ *a prime ideal in* R*. Then* $(R_\mathfrak{p}, \mathfrak{p}_\mathfrak{p})$ *is a regular local ring.*

Proof. If $\mathrm{gl}\dim R$ is finite, then $\mathrm{gl}\dim R_\mathfrak{p}$ is finite. In fact, we have $\mathrm{gl}\dim R_\mathfrak{p} \leq \mathrm{gl}\dim R$ as $R_\mathfrak{p}$ is R-flat and any $R_\mathfrak{p}$-module can be written as $R_\mathfrak{p} \otimes_R N$ for some R-module N. The result follows. $\square$

Definition 133.31. A commutative Noetherian ring is called *regular* if $R_\mathfrak{m}$ is a regular local ring for all maximal ideals $\mathfrak{m}$ in R.

By Krull's Conjecture, we have

Corollary 133.32. R *is regular ring if and only if* $R_\mathfrak{p}$ *is a regular local ring for all* $\mathfrak{p} \in \mathrm{Spec}(R)$.

Corollary 133.33. *Let* R *be a commutative Noetherian ring. Then* R *is a regular ring if and only if every finitely generated* R*-module* M *satisfies* $\mathrm{pd}_R\, M < \infty$.

Proof. ($\Leftarrow$): If $\mathfrak{m}$ is a maximal ideal in R, then $\mathrm{pd}_{R_\mathfrak{m}}\, \mathfrak{m}_\mathfrak{m} \leq \mathrm{pd}_R\, \mathfrak{m} < \infty$ as $\mathrm{ht}_{R_\mathfrak{m}}\, \mathfrak{m}_\mathfrak{m} \leq \mathrm{ht}_R\, \mathfrak{m} < \infty$.
($\Rightarrow$): As R is Noetherian and M is finitely generated, there exists a projective resolution $P_* \to M \to 0$ with each P_i a Noetherian R-module and all (syzygies) $K_iF = \ker(P_i \to P_{i-1})$ are finitely generated. For each $\mathfrak{p} \in \mathrm{Spec}(R)$, there exists an integer $n = n_\mathfrak{p}$ such that $(K_n)_\mathfrak{p}$ is a finitely generated free $R_\mathfrak{p}$ module as $R_\mathfrak{p}$ is a regular local ring.. Let $\{\frac{1}{s_1}x_1, \dots, \frac{1}{s_n}x_n\}$ be a basis for $(K_n)_\mathfrak{p}$ with $s_i \notin R \setminus \mathfrak{p}$, $1 = 1, \dots, n$. Let $f_\mathfrak{p} = s_1 \cdots s_n \in R \setminus \mathfrak{p}$. Then $D(f_\mathfrak{p})$ is an open set in $\mathrm{Spec}(R)$. As $\mathrm{Spec}(R)$ is quasi-compact (cf. Proposition 92.7), $\mathrm{Spec}(R)$ has a finite open subcover of $\{D(f_\mathfrak{p}) \mid \mathfrak{p} \in \mathrm{Spec}(R)\}$, say $\{D(f_{\mathfrak{p}_1}), \dots, D(f_{\mathfrak{p}_m})\}$. Let $N = \max\{n_{\mathfrak{p}_i} \mid 1 \leq i \leq m\}$. Then K_N is finitely generated and locally free so projective. Therefore, $0 \to K_N \to P_N \to \cdots \to P_0 \to M \to 0$ is a projective resolution of M. $\square$

Remark 133.34. There exist regular rings of infinite global dimension. Some authors also call a commutative Noetherian ring regular if R also has finite global dimension.

Theorem 133.35. (Nagata) *Let* R *be a Noetherian domain and* Γ *a set of prime elements in* R*. Suppose that the localization* $S^{-1}R$ *is a UFD where* S *is the multiplicative group generated by* Γ*. Then* R *is a UFD.*

Proof. By Theorem 31.1, it suffices to show every prime ideal of height one is principal. Let $\mathfrak{p}$ be a prime ideal of height one in R. If $\mathfrak{p} \cap S \neq \emptyset$, then there exist an $x \in \Gamma \cap \mathfrak{p}$ as $\mathfrak{p}$ is a prime ideal. Therefore, $0 < (x) \subset \mathfrak{p}$. As (x) is a prime ideal, $\mathrm{ht}(\mathfrak{p}) > 1$, a contradiction. So we may assume that $\mathfrak{p} \cap S = \emptyset$. By assumption, there exist an $x \in \mathfrak{p}$ such that $S^{-1}\mathfrak{p} = S^{-1}(x)$. By Noetherian induction, there exists such an x in $\mathfrak{P}$ with (x) maximal. In particular, if $\pi \in \Gamma$, then $\pi \not| \, x$. Let $y \in \mathfrak{p}$. Then $sy = ax$, for some $s \in S$ and $a \in R$. Write $s = \pi_1 \cdots \pi_n$, $\pi_1, \dots \pi_n$ in Γ. Since $\pi_i \not| \, x$, we must have $\pi_i \mid a$ for $i = 1, \dots, n$. Induction on i shows that $s \mid a$ in R. Therefore, $y \in (x)$. Hence $\mathfrak{p} = (x)$. $\square$

Lemma 133.36. *Let R be a domain and $\mathfrak{A} \subset R$ an ideal. Suppose that $\mathfrak{A} \coprod R^n \cong R^{n+1}$ for some n. Then $\mathfrak{A}$ is a principal ideal.*

We give two proofs.

Proof. We may view $\mathfrak{A} \coprod R^n = \mathfrak{A} \oplus R^n \subset R \oplus R^n = R^{n+1}$. Let $\varphi : R^{n+1} \to \mathfrak{A} \oplus R^n$ be an isomorphism. Let $\{e_0, \dots, e_n\}$ be a basis for R^{n+1}. Let $\{f_0\}$ be a basis for R and $\{f_1, \dots, f_n\}$ a basis for R^n, So $\{f_0, \dots, f_n\}$ is a basis for $R \oplus R^n$. Suppose that $\varphi(e_i) = \sum_{j=1}^{n+1} a_{ji} f_j$. Set A to be the matrix (a_{ij}), $d = \det A$, and d_i the $(0, i)$th cofactor of A. Then by matrix theory, $\sum_{i=1}^{n+1} a_{ji} d_i = \delta_{0j}$. Therefore, $0 \neq d = \sum_{i=0}^{n} a_{0i} d_i$ and $0 = \sum_{i=0}^{n} a_{ji} d_i$ for $j \neq 0$. Let $e_0' = \sum_{i=0}^{n+1}$. So $\varphi(e_0') = df_0$. Since φ is surjective, there exist $e_i' \in R^{n+1}$ such that $\varphi(e_i') = f_i$ for $i = 1, \dots, n$. Define the matrix $C = (c_{ij})$ by $e_j' = \sum_{k=0}^{n} c_{kj} e_k$ for $j = 0, \dots, n$ In particular, $c_{k0} = d_k$ for $k = 0, \dots, n$ by definition. As As

$$AC = \begin{pmatrix} d & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

by definition, $\det C = 1$. Thus $\{e_0', \dots, e_n'\}$ is a basis for R^{n+1}. Hence $\mathfrak{A} f_0 = \varphi(Re_0') = Rdf_0$, as $\varphi(e_i') \subset 0 \oplus R^n$. Consequently, $fA = Rd$ (in the domain R) as needed. □

A less computational proof of the lemma which we now give uses the exterior algebra. In particular, we use 121.11(15).

Proof. (Second Proof) The ideal $\mathfrak{A}$ is projective and of rank one. Let $\mathfrak{p}$ be a prime in R. So $\mathfrak{A}_\mathfrak{p}$ in the local ring, $R_\mathfrak{p}$ is free of rank one. Hence $(\bigwedge^i \mathfrak{A})_\mathfrak{p} = \bigwedge^i (\mathfrak{A}_\mathfrak{p}) = 0$ for all $i > 0$. Since $\bigwedge^{n+1} R^n = 0$, using Exercise 121.11(15) , we have

$$\begin{aligned} R &\cong \bigwedge^{n+1} R^{n+1} \cong \bigwedge^{n+1} (\mathfrak{A} \coprod R^n) \\ &= \coprod_{i=0}^{n+1} \bigwedge^i \mathfrak{A} \otimes_R \bigwedge^{n+1-i} R^n = \mathfrak{A} \otimes \bigwedge^n R^n \cong \mathfrak{A}. \end{aligned}$$

So $\mathfrak{A}$ is R-free hence principal. □

We now prove that regular local rings are UFDs without a characteristic assumption. We need an elementary lemma.

Definition 133.37. Let R be a commutative ring and M a finitely genenerated R-module. We say that M has a *finite free resolution* or *FFR* of *length* $\leq n$ if there exists an exact sequence $0 \to F_n \to \cdots \to F_1 \to M \to 0$ with each F_i a finitely free R-module. We say that M is *stably free* if there exists a finitely generated free R-module F such that $M \coprod F$ is free.
Even though a finitely generated projective R-module is a direct summand of a free R-module, it may not be stably free.

Lemma 133.38. *Let R be a commutative Noetherian ring and M a finitely generated R-module. Then M has a finite free resolution if and only if M is stably free, i.e., there exists a free R-module F such that $F \coprod M$ is free.*

Proof. ($\Rightarrow$): Let

$$0 \to F_n \to \cdots \to F_0 \xrightarrow{\varepsilon} M \to 0$$

be a finite free resolution of M and $K_0 = \ker \varepsilon$. As M is R-projective, the exact sequence $0 \to K_0 \to F_0 \xrightarrow{f} M \to 0$ splits. Hence K_0 is also projective and has a finite free resolution $0 \to F_n \to \cdots \to F_1 \to K_0 \to 0$. By induction on n, K_0 is stably free, hence so is M.
($\Leftarrow$): There exists an exact sequence $0 \to F_1 \to F_0 \to M \to 0$ with F_0 and F_1 finitely generated free R-modules as M is stably free.
This proves the lemma. □

Theorem 133.39. (Auslander-Buchsbaum) *Let $(R, \mathfrak{m})$ be a regular local ring. Then R is a UFD.*

Proof. We induct on $d = \dim R$. If $d = 0$, then R is a field, and the result follows. If $\mathfrak{m}$ is generated by an R-sequence of length one, then $\mathfrak{m}$ is principal. In particular, $\operatorname{Spec}(R) = \{0, \mathfrak{m}\}$, and the result follows. So we may assume that $d > 0$.

As before, choose $x \in \mathfrak{m} \setminus \mathfrak{m}^2$. Then $R/(x)$ is a regular local ring by Corollary 98.11, hence a domain. Therefore, $(x) \in \operatorname{Spec}(R)$. By Nagata's Theorem 133.35 about UFDs, it suffices to show $x = R[x^{-1}]$ is a UFD.
Note that the localization R_x at $\{x^n \mid n \geq 0\}$ satisfies $R \subset R_x$ and $\dim R_x < \dim R$, since $\mathfrak{m}_x = R_x$. Note also that R_x need not be a regular local ring.
Let $\mathfrak{P} \in \operatorname{Spec}(R_x)$ be a prime of height one. We must show that $\mathfrak{P}$ is principal. Let $\mathfrak{p} = \mathfrak{P} \cap R$, so $\mathfrak{p} \in \operatorname{Spec}(R)$ and $\mathfrak{P} = \mathfrak{p}_x$. As $\operatorname{gl} R$ is finite, and R local, so finitely generated R-projective module is R-free, $\mathfrak{p}$ has a FFR (as R is Noetherian). As localization takes exact sequences to exact sequences (by Exercise 92.31(6)), $\mathfrak{P}$ has an FFR by finitely generated R_x-free modules. Let $\mathfrak{Q} \in \operatorname{Spec}(R_x)$. Then $(R_x)_{\mathfrak{Q}} = R_{\mathfrak{Q} \cap R}$ (check) is a regular local ring of dimension less than d, hence a UFD by induction. Therefore, any height one prime in $(R_x)_{\mathfrak{Q}}$, $\mathfrak{Q} \in \operatorname{Spec}(R_x)$, is principal. In particular, $\mathfrak{P}_{\mathfrak{Q}}$ is principal. Consequently, we also have $\mathfrak{P}_{\mathfrak{Q}}$ is $(R_x)_{\mathfrak{Q}}$-free for all $\mathfrak{Q} \in \operatorname{Spec}(R_x)$. Since R_x is Noetherian, finitely generated $\mathfrak{P}$ is R_x-projective by Theorem 127.4. Moreover, the R_x-projective module $\mathfrak{P}$ has an FFR. To finish, we must show that $\mathfrak{P}$ is principal. By Lemma 133.38, $\mathfrak{P} \coprod R_x^n \cong R_x^m$ for some m and n. Localizing shows that $n + 1 = m$. Hence $\mathfrak{P}$ is principal by Lemma 133.36. □

Theorem 133.40. *Let R be a commutative Noetherian domain. Suppose that every finitely generated R-module M has an FFR. Then R is a UFD*

Proof. Let $\mathfrak{p} \in \operatorname{Spec}(R)$. Then $\mathfrak{p} R_{\mathfrak{p}}$ has an FFR, so $\operatorname{gl} R_{\mathfrak{p}} < \infty$. Therefore, $R_{\mathfrak{p}}$ is a regular local ring by Serre's Theorem 133.28. If $\mathfrak{P} \in \operatorname{Spec}(R)$ is of height one, then $\mathfrak{P}_{\mathfrak{p}}$ is principal in the UFD $R_{\mathfrak{p}}$, hence $R_{\mathfrak{p}}$-free for all $\mathfrak{p} \in \operatorname{Spec}(R)$. Therefore, $\mathfrak{P}$ is R-projective and has an FFR, so is stably free by the Lemma 133.38. Therefore $\mathfrak{P}$ is principal by the argument at the end of the proof of Auslander-Buchsbaum Theorem 133.39. □

Remarks 133.41. 1. $\mathbb{Z}[\sqrt{-1}]$ is not a UFD, but every localization at a prime ideal is.
2. There exist Noetherian local rings that are UFD's but not a regular local ring.
3. Let R be a Noetherian ring. We know that $R[[t]]$ is Noetherian. However, it is false that $R[[t]]$ is a UFD when R is also regular. Samuel proved that if R is a both regular and a UFD, then so is R.

Exercises 133.42. 1. Let $\varphi : R \to S$ be a ring homomorphism. Show that the pullback of exact sequences of S-modules is an exact sequence of R-modules.

2. Let M_i be R-modules for $i \in I$. Show $\mathrm{pd}_R(\coprod M_i) \leq \sup(\mathrm{pd}_R(M_i)_{i\in I}$.
3. Show $\mathrm{pd}(\mathbb{Z}/4\mathbb{Z}) = \infty$, so $\mathrm{gl}\dim \mathbb{Z}/4\mathbb{Z} = \infty$ and $\mathrm{gl}\dim \mathbb{Z}/2\mathbb{Z} = 0$.
4. Let R be a commutative Noetherian ring and A a *faithfully flat* commutative R-algebra, i.e., A is R-flat and if $A \otimes_R M = 0$ for an R-module M, then $M = 0$. Let M be a finitely generated R-module. Show that $\mathrm{pd}_A A \otimes_R M = \mathrm{pd}_R M$.
5. Let R be a commutative ring and $x_1, \ldots, x_r$ be an R-sequence. Prove $\mathrm{pd}\, R/(x_1, \ldots, x_r) = r$.
6. Prove Theorem 133.13.
7. Prove that $J(R) = \cap\, \mathrm{ann}_R(M)$, where M ranges over all the simple left R-modules. In particular, $J(R)$ is an ideal.
8. Prove Lemma 133.14.
9. Prove Lemma 133.15.
10. Show if R is a commutative ring and $\mathfrak{p}$ a prime ideal in R and $M_{\mathfrak{p}}$ an R-module, then there exists an R-module N satisfying $M = R_{\mathfrak{p}} \otimes_R N$.

CHAPTER XXII

Categories

Many of the basic ideas studied in the previous chapters can be formulated as special cases of a general theory, called category theory which determines fundamental structure of mathematical objects that occur in mathematics. In this chapter, we give an introduction to it. We will begin with it in general form, but most of the results will be aimed at the structures that we have studied. Previous chapters have given a glimpse of this theory, and proofs that we have given still work in the proper context of category theory. Because of this, we leave many proofs as exercises and those given may well be left less detailed than in previous chapters. Not all connections between the various ideas that can be identified will be made totally explicit.

134. Categories

We shall begin by defining the general notion of a category. In particular, we shall not worry about set theory. (including axiom choice and the notion of a class) To make this accurate we would need to assume certain axioms, e.g., existence of large cardinals or a universe, which we shall not enumerate, that allows us to avoid versions of Russell's paradox. So instead of calling collections sets, classes, etc., we just use the word collection. We will then begin to restrict our results to what are called locally small categories, categories whose morphisms between objects form a set. Many of the results would still generalize with the proper hypotheses, and many of our proofs can be modified to work in more general situations. Our main interest, besides showing the language, is to show much what we did for modules is much more general and is needed in other contexts, e.g., algebraic geometry.

Definition 134.1. A **category** $\mathcal{C}$ consists of all of the following:

1. A collection of *objects* $\mathrm{Ob}(\mathcal{C})$.
2. For each ordered pair (A, B) of objects, a collection of *morphisms* denoted by $\mathrm{Hom}_{\mathcal{C}}(A, B)$. If f lies in the collection $\mathrm{Hom}_{\mathcal{C}}(A, B)$, we denote it by $f : A \to B$ or $A \xrightarrow{f} B$ in $\mathcal{C}$. We call A the *domain* of f and denote it by $\mathrm{dom}\, f$ and call B the *codomain* of f and denoted it by $\mathrm{cod}\, f$. [Morphisms are also called *arrows*. They need not be functions.]
3. For all objects $A, B, C \in \mathrm{Ob}(\mathcal{C})$, a binary operation

$$\mathrm{Hom}_{\mathcal{C}}(B, C) \times \mathrm{Hom}_{\mathcal{C}}(A, B) \to \mathrm{Hom}_{\mathcal{C}}(A, C)$$

denoted by $(f, g) \mapsto fg$ (or $f \circ g$). We say that such f and g are *composable*. Moreover, such composable morphisms satisfy:

(*i*) If $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$, $g \in \mathrm{Hom}_{\mathcal{C}}(B, C)$, and $h \in \mathrm{Hom}_{\mathcal{C}}(C, D)$, the associative law holds, i.e.,
$$(hg)f = h(fg).$$
(*ii*) For every $A \in \mathrm{Ob}(\mathcal{C})$, there exists an *identity morphism* $1_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$ that satisfies for all $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$ and $g \in \mathrm{Hom}_{\mathcal{C}}(B, A)$
$$f\,1_A = f \text{ and } 1_A\,g = g$$
for all $B \in \mathrm{Ob}(\mathcal{C})$.

Notation 134.2. If $\mathcal{C}$ is a category, we sometimes write $A \in \mathcal{C}$ for $A \in \mathrm{Ob}(\mathcal{C})$ and *morph*$(\mathcal{C})$ for the totality of morphisms of $\mathcal{C}$.

It is often convenient to define a category $\mathcal{C}$ using the following notation:
$$\mathcal{C} := ((\text{objects in } \mathcal{C}, \text{ morphisms in } \mathcal{C})).$$

Definition 134.3. Let $\mathcal{C}$ be a category. We say $\mathcal{C}$ is a *locally small* category if $\mathrm{Hom}_{\mathcal{C}}(A, B)$ is a set for all $A, B \in \mathrm{Ob}(\mathcal{C})$ and $\mathcal{C}$ is *small* if the totality of morphisms in $\mathcal{C}$ is a set.

Note that if $\mathcal{C}$ is a category, there exists a bijection between objects and identity morphisms given by $A \mapsto 1_A$. In particular, it follows that $\mathrm{Ob}(\mathcal{C})$ is a set if $\mathcal{C}$ is small.

Examples 134.4. The following are categories:

1. **Set** := ((sets, maps)).
2. **Group** := ((groups, group homomorphisms)).
3. **Ab** := ((abelian groups, (abelian) group homomorphisms)).
4. **Ring**:= ((rings, ring homomorphisms)).
5. **ComRing**:= ((commutative rings, ring homomorphisms)).
6. ${}_R\mathcal{M}$:= ((left R-modules, R-homomorphisms)).
7. $\mathcal{M}_R$:= ((right R-modules, R-homomorphisms)).
8. ${}_R\mathcal{M}_S$:= (((R-S)-bimodules, (R-S)-homomorphisms)).
9. **Field** : = ((fields, field homomorphisms))
10. Let P be a poset under $\leq$. Then P can be viewed as a category with object P and morphisms given by a unique $f : a \mapsto b$ if $a \leq b$.
    For example,
    (*i*) $\mathbb{0}$ is the category with no objects and no morphisms.
    (*ii*) $\mathbb{1}$ is the category with one object and one morphism.
    (*iii*) $\mathbb{2}$ is the category with two objects and a single non-identity morphism between them.
    (*iv*) ω a category determined by $0 \to 1 \to 2 \to 3 \cdots$.
11. **Poset** := ((posets, order-preserving morphisms)) where a morphism $f : A \to B$ takes $x \leq y$ in A to $f(x) \leq f(y)$ in B. In particular, if β is any ordinal, then the poset $\{\alpha < \beta\}$ is a category.
12. **Top** := ((topological spaces, continuous maps)).

13. Let $*$ be a distinguished element in a topological space, i.e., a *basepoint.* $(\mathbf{Top})_* :=$ ((topological spaces with basepoints, continuous maps preserving basepoints)).
14. **Man**:= ((manifolds, continuous maps).
15. **Diff**:= ((differentiable manifolds, differentiable maps)).
16. **Anal**:= ((analytic manifolds, analytic maps)).
17. Let G be a group. Then $\mathbf{B}\mathrm{G} = ((\{*\}, \operatorname{Hom}(*,*)))$, where $\{*\}$ is the set with one element $*$, $G = \operatorname{Hom}(*,*)$ via the group operation, i.e., the composition of morphisms on $*$ corresponds to the binary relation on the group. So in this category every morphism is an isomorphism (in fact, an automorphism), where a morphism $f : A \to B$ in $\mathcal{C}$ is called an *isomorphism* if there exist a morphism $g : B \to A$ satisfying $gf = 1_A$ and $fg = 1_B$.
Of course, this corresponds to our notions of isomorphisms in the above examples. We have the obvious definitions for *automorphisms endomorphisms*, and *automorphisms*. In particular, $\mathbf{B}\mathrm{G}$ is an example of a *groupoid*, defined to be a category in which all morphisms are isomorphisms.

    [More generally, if M is a monoid, then $((\{*\}, \operatorname{Hom}(*,*)))$ is a category with $\operatorname{Hom}(*,*)$ is the given monoid. Of course, in this case not every morphism is an isomorphism.]
18. If $\mathcal{C}$ is a category, then a *subcategory* $\mathcal{B}$ of $\mathcal{C}$ is a category whose objects is a subcollection of $\operatorname{Ob}(\mathcal{C})$ with morphisms in $\mathcal{B}$ arising from $\mathcal{C}$, i.e., those morphisms in $\mathcal{C}$ whose domain and codomain lie in $\mathcal{B}$ and whose identity morphisms in $\mathcal{B}$ arise from those in $\mathcal{C}$.
19. Let $\mathcal{C}$ be a category. Define a *congruence relation* R on $\mathcal{C}$ as follows: For each pair of objects $X, Y \in \operatorname{Ob}(\mathcal{C})$, there is an equivalence relation $R_{X,Y}$ on $\operatorname{Hom}(X,Y)$, such that these equivalence relations respect composition of morphisms. That is, if $f_1, f_2 : X \to Y$ are related in $\operatorname{Hom}(X,Y)$ and $g_1, g_2 : Y \to Z$ are related in $\operatorname{Hom}(Y,Z)$, then g_1f_1 and g_2f_2 are related in $\operatorname{Hom}(X,Z)$. Given such a congruence relation R on $\mathcal{C}$, define the *quotient category* $\mathcal{C}/R :=$ $((\operatorname{Ob}(\mathcal{C})$, equivalence classes of $morph(\mathcal{C})$, i.e., $\operatorname{Hom}_{\mathcal{C}/R}(X,Y) = \operatorname{Hom}_{\mathcal{C}}(X,Y)/R$. Composition of morphisms in $\mathcal{C}/R$ is well-defined, since R is a congruence relation.
20. Let $\mathcal{C}$ be a category and $C \in \operatorname{Ob}(\mathcal{C})$. We construct two new categories from $\mathcal{C}$ and C. The first $C\backslash\mathcal{C}$ has objects that are morphisms $f : C \to X$ with $\operatorname{dom} f = C$ in $\mathcal{C}$ and morphisms of objects $f : C \to X$ and $g : C \to Y$ in $C\backslash\mathcal{C}$ to be a morphism $h : X \to Y$ in $\mathcal{C}$ of the codomains in $\mathcal{C}$ so that

$$\begin{tikzcd} & C \arrow[dl, "f"'] \arrow[dr, "g"] & \\ X \arrow[rr, "h"'] & & Y \end{tikzcd}$$

    commutes, i.e., $g = hf$. This category is called the *slice category of* $\mathcal{C}$ *under* C.
The second new category is $\mathcal{C}/C$ that consists of objects that are morphisms $f : X \to C$ with $\operatorname{cod} f = C$ in $\mathcal{C}$ and morphisms of objects $f : X \to C$ and $g : Y \to C$ in $\mathcal{C}/C$ to be a morphism $h : X \to Y$ in $\mathcal{C}$ with h a map of the domains in $\mathcal{C}$ so that

$$\begin{tikzcd} X \arrow[rr, "h"] \arrow[dr, "f"'] & & Y \arrow[dl, "g"] \\ & C & \end{tikzcd}$$

commutes. This category is called the *slice category of* $\mathcal{C}$ *over* C

21. Let $\mathcal{C}$ be a category. The *opposite category* is defined to be the category $\mathcal{C}^{op} := ((\mathrm{Ob}(\mathcal{C}), \mathrm{Hom}_{\mathcal{C}^{op}}))$, where the morphisms f^{op} in $\mathcal{C}^{op}$ are in one-to-one correspondence with the morphisms f in $\mathcal{C}$ with $\mathrm{dom}\, f^{op} = \mathrm{cod}\, f$, $\mathrm{cod}\, f^{op} = \mathrm{dom}\, f$, $(fg)^{op} = g^{op} f^{op}$ for composable morphisms in $\mathcal{C}$, and 1_A^{op} is the identity on each $A \in \mathrm{Ob}(\mathcal{C})$. In particular, $(\mathcal{C}^{op})^{op} = \mathcal{C}$. This leads to the Principle of Duality: A statement in a category $\mathcal{C}$ based on the axioms of the elementary theory of categories has an opposite statement true in its dual category.

Remarks 134.5. Let $\mathcal{C}$ be a category.

1. An object $I \in \mathrm{Ob}(\mathcal{C})$ is called an *initial object* if for every $A \in \mathrm{Ob}(\mathcal{C})$, there exists a unique morphism $I \to A$. (Of course, a category may have no such object.) If I is an initial object in $\mathcal{C}$, then $T = I^{op}$ exists in the dual category, and for all A^{op} in $\mathrm{Ob}(\mathcal{C}^{op})$, there exists a unique morphism $A \to T$. Such a T is called a *terminal object* in $\mathcal{C}^{op}$.
   For example, in **Set**, the empty set is an initial object and any singleton set is a terminal object and in the category ${}_R\mathcal{M}$, the zero R-module is both an initial and terminal object. If a category has both an initial and a terminal object, it is unique up to an isomorphism and is called a *zero object*.
2. A morphism $f : A \to B$ in $\mathcal{C}$ is called a *monomorphism* if whenever $fg = fh$ with $g, h \in \mathrm{Hom}_{\mathcal{C}}(X, A)$, then $g = h$. We denote a monomorphism by $f : A \hookrightarrow B$. Such a monomorphism is called a *split monomorphism* if there exists a morphism $g : B \to A$ such that $fg = 1_A$. A morphism $f : A \to B$ is called an an *epimorphism.* if f^{op} in $\mathcal{C}^{op}$ is a monomorphism, i.e., $gf = hf$ with $g, h \in \mathrm{Hom}_{\mathcal{C}}(B, X)$, then $g = h$. We denote an epimorphism by $f : A \twoheadrightarrow B$. An epimorphism f is called a *split epimorphism* if f^{op} is a split monomorphism.
3. In general, the definition of a morphism being a monomorphism or epimorphism above differs from our notion of monomorphism and epimorphism defined before. In a category whose objects are sets (possibly with additional structure), we have the usual notion of injective and surjective maps. In the category **Ring** the inclusion map $\mathbb{Z} \to \mathbb{Q}$ is an epimorphism in **Ring** but is not a surjection homomorphism of rings. In fact, it is also a monomorphism in **Ring**, but is not an isomorphism in **Ring**. In the category ((infinite groups, group homomorphisms)), the morphisms $\mathbb{C}^\times \to \mathbb{C}^\times$ in **Ab** given by $x \mapsto x^n$, $n > 1$, are monomorphisms in **Ab** but not injections group homomorphisms. It is true, however, that in the categories **Set**, **Group** (not so obvious), and ${}_R\mathcal{M}$, monomorphisms are injective and epimorphisms are surjective homomorphisms.

Exercises 134.6. 1. Show that if a category has a zero object, it is unique up to an isomorphism.

2. Prove that a morphism in **Group** is injective if and only if a monomorphism and is surjective if and only if it is surjective.

3. Show that the inclusion $\mathbb{Z} \to \mathbb{Q}$ in **Ring** is both a monomorphism and epimorphism, but not an isomorphism.
4. Let $\mathcal{C}$ be a category and $f : A \to B$ and $g : B \to C$ in $\mathcal{C}$. show Show
   (i) If f and g are monomorphisms ,so is$gf$is a mononmorphism.
   (ii) If fg is a monomorphism, so is f.
   (iii) If f and g are epimorphisms, so is gf.
   (iv) If fg is an epimorphism, so is g.

135. Functors

When studying sets or sets with additional structure, one must study morphisms between them to determine properties of the objects. In a category one must do the same thing, i.e., study morphisms between categories. This leads to the concept of a functor that we now investigate.

Definition 135.1. A *functor* (or *covariant functor*) $F : \mathcal{A} \to \mathcal{B}$ of categories associates:

1. To each $A \in \mathrm{Ob}(\mathcal{A})$, an object $FA \in \mathcal{B}$ (or written $F(A)$ for clarity).
2. To each morphism $f : A \to B$ in $\mathcal{A}$, a morphism $Ff : FA \to FB$ in $\mathcal{B}$ (or written $F(f)$ for clarity).
3. To each $A \in \mathrm{Ob}(\mathcal{A})$, we have $F1_A = 1_{FA}$.
4. If $f \in \mathrm{Hom}_{\mathcal{A}}(A, B)$ and $g \in \mathrm{Hom}_{\mathcal{A}}(B, C)$, then $F(gf) = (Fg)(Ff)$.

We also write the functor as

$$\mathcal{A} \xrightarrow{F} \mathcal{B}$$

and use commutative diagrams to mean the obvious.
A *contravariant functor* or (*cofunctor*) is a functor $F : \mathcal{A}^{op} \to \mathcal{B}$.
By duality, we often leave details of the contravariant case to the reader.

Two categories $\mathcal{A}$ and $\mathcal{B}$ are called *isomorphic* if there exists a functor $F : \mathcal{A} \to \mathcal{B}$ that is an *isomorphism*, i.e., there exists a functor $G : \mathcal{B} \to \mathcal{A}$ satisfying $GF = 1_{\mathcal{A}}$ and $FG = 1_{\mathcal{B}}$.

If $\mathcal{A}$ and $\mathcal{B}$ be categories, we let $\mathrm{Funct}(\mathcal{A}, \mathcal{B})$ denote the functors from $\mathcal{A}$ to $\mathcal{B}$.

Examples 135.2. 1. Let $\mathcal{C}$ be a category, then $1_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ given by $1_{\mathcal{C}}(C) = 1_C$ for all $C \in \mathrm{Ob}(\mathcal{C})$ is a functor.

2. Let $\mathcal{A}$ be a category. Then a functor $U : \mathcal{A} \to \mathcal{B}$ is called a *forgetful functor* if some of the structure of $\mathcal{A}$ is ignored, e.g., $U : {}_R\mathcal{M} \to \mathbf{Ab}$ by ignoring R-structure or $U : \mathbf{Ring} \to \mathbf{Set}$ by ignoring the algebraic structure of rings.
3. $F : \mathbf{Set} \to \mathbf{Group}$ that gives the free group on a set S is a functor.
4. Let $\mathcal{C}$ be a category and R a congruence relation on $\mathcal{C}$. Let $^{-} : \mathcal{C} \to \mathcal{C}/R$ sending each morphism to its equivalence class. This is a functor, called the *quotient functor*. It is bijective on objects and surjective on *morph*$(\mathcal{C})$. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor. Then F determines a congruence relation $\sim$ on $\mathcal{C}$ by defining $f \sim g$ if $F(f) = F(g)$ for $C, D \in \mathrm{Ob}(\mathcal{C})$ and $f, g \in \mathrm{Hom}_{\mathcal{C}}(C, D)$. Moreover, f factors through the associated quotient functor $^{-} : \mathcal{C} \to \mathcal{C}/{\sim}$ in a unique manner, i.e., it induces a canonical functor

$\overline{F} : \mathcal{C}/R_F \to \mathcal{D}$ so that that the diagram

$$\begin{tikzcd} \mathcal{C} \arrow[r, "F"] \arrow[d, "-"'] & \mathcal{D} \\ \mathcal{C}/\sim \arrow[ur, "\overline{F}"'] & \end{tikzcd}$$

commutes. This is the "First Isomorphism Theorem" for functors.

5. Let **ComRing** := ((commutative rings, ring homomorphisms)). Then we have a contravariant functor Spec : $(\mathbf{ComRing})^{op} \to \mathbf{Top}$ that takes a ring R to $\operatorname{Spec}(R)$ with the Zariski topology.

6. Let $\mathcal{C}$ be a small category. Then a functor $\mathcal{C}^{op} \to \mathbf{Set}$, i.e., a contravariant functor $\mathcal{C} \to \mathbf{Set}$, is called a *presheaf* on $\mathcal{C}$ with values in **Set**.
   For example, let X be a topological space. Let $\mathcal{U}$ be the category with objects the open sets in X and morphisms the partial order on open sets in X given by inclusion. Then the presheaf $\mathcal{U}^{op} \to \mathbf{Set}$ is called presheaf over X with values in **Set**. The category **Set** can be replaced by other categories, e.g., **Group**, **ComRing**, ${}_R\mathcal{M}$, or other algebraic, geometric, and topological categories.

7. Let G be a group and $\mathcal{C}$ a category. A functor $F : \mathbf{B}G \to \mathcal{C}$ gives a unique $X \in \operatorname{Ob}(\mathcal{C})$ in the image of F together with an endomorphism $\lambda_g : X \to X$ for each $g \in G$ satisfying $\lambda_g\lambda_h = \lambda_{gh}$ and $\lambda_{e_G} = 1_X$ for all $h \in G$. This defines a G-*action* on $\mathcal{C}$ and is a functor.

8. Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be categories and $F : \mathcal{A} \to \mathcal{C}$ and $G : \mathcal{B} \to \mathcal{C}$ be functors. One usually write this as $\mathcal{A} \xrightarrow{F} \mathcal{C} \xleftarrow{G} \mathcal{B}$. We create a new category $(F \downarrow G)$ called the *comma category* of F and G as follows:
   (*i*) Objects of $(F \downarrow G)$ are triples (A, B, h) with $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $h : F(A) \to G(B)$ a morphism in $\mathcal{C}$.
   (*ii*) Morphisms $(A, B, h) \to (A', B', h')$ in $(F \downarrow G)$ are all pairs (f, g) with $f : A \to A'$ a morphism in $\mathcal{A}$ and $g : B \to B'$ a morphism in $\mathcal{B}$ such that the following diagram commutes

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, "h"'] & F(A') \arrow[d, "h'"] \\ G(B) \arrow[r, "G(g)"'] & G(B'). \end{tikzcd}$$

   commutes in $\mathcal{C}$.
   (*iii*) Morphisms are composed by $(f', g') \circ (f, g) := (f' \circ f, g' \circ g)$ whenever the left hand side is defined. The identity morphism on the object (A, B, h) is $(1_A, 1_B)$.

For example, the comma category $\mathcal{A} \xrightarrow{1_{\mathcal{A}}} \mathcal{C} \xleftarrow{G} \mathbb{1}$ is the case where if $\mathcal{C} = \mathcal{A}$, $F = 1_{\mathcal{A}}$ and $\mathcal{B} = \mathbb{1}$ the category with one object and one morphism, i.e., $G(*) = A_*$ for some $A_* \in \mathcal{A}$. One writes this category as $(\mathcal{A} \downarrow A_*)$. In this case, the object (A, B, h) is usually written (A, π_A) where $*$ is omitted and the morphism $(f, h) = \pi_A$, In this case a morphism $(f, 1_*) : (A, \pi_A) \to (A', \pi'_A)$ is written simply as $f : A \to A'$ such that the diagram

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "\pi_i"'] & & A' \arrow[dl, "\pi_i'"] \\ & A_* & \end{tikzcd}$$

commutes. This is just the slice category $\mathcal{A}/A_*$ of $\mathcal{A}$ over A_* and the comma category $\mathbb{1} \xrightarrow{1_{\mathcal{B}}} \mathcal{C} \xleftarrow{G} B$ is just the slice category $B/\mathcal{B}$ over B.

Construction/Definition 135.3. Let $\mathcal{A}$ and $\mathcal{B}$ be categories. Then the *product category* $\mathcal{A} \times \mathcal{B}$ is the category with

$$\mathrm{Ob}(\mathcal{A} \times \mathcal{B}) := \{(A, B)\,(\text{ordered pair}) \mid A \in \mathcal{A},\ B \in \mathcal{B}\}$$

and morphisms ordered pairs $(f, g) : (A, B) \to (A', B')$ with $f : A \mapsto A'$ a morphism in $\mathcal{A}$ and $g : B \mapsto B'$ a morphism in $\mathcal{B}$. We have $1_{A \times B} = (1_A, 1_B)$ and if $f' : A' \to A''$ and $g' : B' \to B''$, then $(f', g')(f, g) = (f'f, g'g)$.
We have the *projection* functors $P_{\mathcal{A}} : \mathcal{A} \times \mathcal{B} \to \mathcal{A}$ given by $P_{\mathcal{A}}(f, g) = f$ and $P_{\mathcal{B}} : \mathcal{A} \times \mathcal{B} \to \mathcal{B}$ given by $P_{\mathcal{B}}(f, g) = g$.
Suppose that we have a category $\mathcal{C}$ together with functors

$$\mathcal{A} \xleftarrow{G} \mathcal{C} \xrightarrow{H} \mathcal{B}.$$

Then there exists a unique functor $F : \mathcal{C} \to \mathcal{A} \times \mathcal{B}$ satisfying $G = P_{\mathcal{A}}F$ and $H = P_{\mathcal{B}}F$, i.e., we have a commutative diagram:

$$\begin{tikzcd} & \mathcal{C} \arrow[dl, "G"'] \arrow[d, "F"] \arrow[dr, "H"] & \\ \mathcal{A} & \mathcal{A} \times \mathcal{B} \arrow[l, "P_{\mathcal{A}}"] \arrow[r, "P_{\mathcal{B}}"'] & \mathcal{B}. \end{tikzcd}$$

Suppose that we have functors $U : \mathcal{A} \to \mathcal{A}'$ and $V : \mathcal{B} \to \mathcal{B}'$. Then there exists a unique *product functor* $U \times V : \mathcal{A} \times \mathcal{B} \to \mathcal{A}' \times \mathcal{B}'$ such that the diagram

$$\begin{tikzcd} \mathcal{A} \arrow[d, "U"'] & \mathcal{A} \times \mathcal{B} \arrow[l, "P_{\mathcal{A}}"'] \arrow[r, "P_{\mathcal{B}}"] \arrow[d, "U \times B"'] & \mathcal{B} \arrow[d, "V"] \\ \mathcal{A}' & \mathcal{A}' \times \mathcal{B}' \arrow[l, "P_{\mathcal{A}'}"] \arrow[r, "P_{\mathcal{B}'}"'] & \mathcal{B}' \end{tikzcd}$$

commutes.

Let $\mathcal{E}$ be a category whose objects are small categories and morphisms the functors between the objects in $\mathcal{E}$. This is a locally small category, but not small. In this case, we would have a product functor $\times : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$ by $(\mathcal{A}, \mathcal{B}) \mapsto \mathcal{A} \times \mathcal{B}$. This is true for $\mathcal{E} =$ **Set**, **Group**, **Posets**. The category of all small categories **Cat** contains each of these special cases of $\mathcal{E}$ as proper subcategories, but none are objects in the category **Cat** := ((small categories, functors)).

We can also look at other functors $G : \mathcal{A} \times \mathcal{B} \to \mathcal{C}$. We call such a functor G a *bifunctor*. (on $\mathcal{A}$ and $\mathcal{B}$). If we fix a variable, we get a functor in the other variable. The two such functors so obtained then determine the bifunctor. Formally, we have

Proposition 135.4. *Let $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ be categories and for all $A \in \mathrm{Ob}(\mathcal{A})$, $B \in \mathrm{Ob}(\mathcal{B})$, and $D \in \mathrm{Ob}(\mathcal{D})$. Let*

$$\lambda_A : \mathcal{B} \to \mathcal{D} \text{ and } \rho_B : \mathcal{A} \to \mathcal{D}$$

be functors satisfying $\lambda_A(B) = \rho_B(A)$ for all A and B. Then there exists a bifunctor $G : \mathcal{A} \times \mathcal{B} \to \mathcal{D}$ satisfying $G(A, \ ) = \lambda_A$ for all A and $G(\ , B) = \rho_B$ for all B if and only if given any pair of morphisms $a : A \to A'$ in $\mathcal{B}$ and $b : B \to B'$ in $\mathcal{B}$, we have

$$\lambda_{A'} b \circ \rho_B a = \rho_{B'} a \circ \lambda_A b$$

i.e., if $G(A)(b) = \lambda_A(b)$, etc., then

$$\begin{tikzcd} G(A,B) \arrow[r, "G(A,b)"] \arrow[d, "G(a,B)"'] & G(A,B') \arrow[d, "G(a,B')"] \\ G(A',B) \arrow[r, "G(A',b)"'] & G(A',B'). \end{tikzcd}$$

commutes.

We leave the proof as an exercise. We shall leave simple results about bifunctors to the reader.

Examples 135.5. 1. Let R be a ring, then $\ \otimes_R \ : \mathcal{M}_R \times {}_R\mathcal{M} \to \mathbf{Group}$ is a bifunctor. If R is commutative ring, then $\ \otimes_R \ : {}_R\mathcal{M} \times {}_R\mathcal{M} \to {}_R\mathcal{M}$ is a bifunctor.

2. We can create a bifunctor for a mix of functors and contravariant functors (but still calling them bifunctors) by using the opposite category. For example, consider the following two functors on locally small categories $\mathcal{C}$ and $\mathcal{C}^{op}$ respectively, and an $X \in \mathrm{Ob}(\mathcal{C})$:

   Let $\mathrm{Hom}_{\mathcal{C}}(X, \ ) : \mathcal{C} \to \mathbf{Set}$ be the functor defined at each $A \in \mathrm{Ob}(\mathcal{C})$ by
   $$\mathrm{Hom}_{\mathcal{C}}(X, \ )(A) = \mathrm{Hom}_{\mathcal{C}}(X, A)$$
   and each morphism $f : A \to B$ in $\mathcal{C}$ by
   $$\mathrm{Hom}_{\mathcal{C}}(X, \ )(f) = \mathrm{Hom}(X, f) : \mathrm{Hom}_{\mathcal{C}}(X, A) \to \mathrm{Hom}_{\mathcal{C}}(X, B)$$
   So $g : X \to A$ in $\mathcal{C}$, $\mathrm{Hom}(X, f)(g) = fg$. We also denote $\mathrm{Hom}(X, f)$ by f_*.

   Let $\mathrm{Hom}_{\mathcal{C}}(\ , X) : \mathcal{C}^{op} \to \mathbf{Set}$ be the contravariant functor defined at each $A \in \mathrm{Ob}(\mathcal{C})$ by
   $$\mathrm{Hom}_{\mathcal{C}}(\ , X)(A) = \mathrm{Hom}_{\mathcal{C}}(A, X)$$
   and each morphism $f : A \to B$ in $\mathcal{C}$ by
   $$\mathrm{Hom}_{\mathcal{C}}(\ , X) \ \mathrm{Hom}(f, X) : \mathrm{Hom}_{\mathcal{C}}(B, X) \to \mathrm{Hom}(A, X).$$
   So if $g : A \to X$ in $\mathcal{C}$, $\mathrm{Hom}(f, X)(g) = gf$. We also denote $\mathrm{Hom}(f, X)$ by f^*

   Then the bifunctor $\mathrm{Hom}_{\mathcal{C}}(\ , \ ) : \mathcal{C}^{op} \times \mathcal{C} \to \mathbf{Set}$ is the functor that takes $(X, Y) \in \mathrm{Ob}(\mathcal{C})$ to $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ and a pair of morphisms (f, g) with $f : X \to A$ and $g : B \to Y$ to $\mathrm{Hom}(f, g)(h) = g \circ h \circ f$.

3. If R is a ring, then $\mathrm{Hom}_R(\ , \ ) : {}_R\mathcal{M} \times {}_R\mathcal{M} \to \mathbf{Group}$ given by $M \times N \mapsto \mathrm{Hom}_R(M, N)$ is a bifunctor. If R is a commutative ring, then $\mathrm{Hom}_R(\ , \ ) : {}_R\mathcal{M} \times {}_R\mathcal{M} \to {}_R\mathcal{M}$ is a bifunctor.

Remarks 135.6. 1. An isomorphism of functors takes isomorphisms of morphisms to isomorphisms.

2. The categories **Set** and $\mathbf{Set^{op}}$ are not isomorphic as the first has a unique initial object and the second does not.
3. If R is a ring, and R^{op} its opposite ring then $(\ )^{op} : {}_R\mathcal{M} \to \mathcal{M}_{R^{op}}$ via $rm \mapsto mr$ is an isomorphism of categories.
4. The definition of a product of two categories generalizes in the obvious way to a product of finitely many categories. Similarly the notion of bifunctors generalizes to n-functors.

Definition 135.7. Let $F : \mathcal{A} \to \mathcal{B}$ be a functor of categories. We say:

(1) F is *faithful* if for all $A, B \in \mathrm{Ob}(\mathcal{A})$, the map $\mathrm{Hom}_{\mathcal{C}}(A, B) \to \mathrm{Hom}_{\mathcal{B}}(FA, FB)$ is faithful, i.e., if $f_1, f_2 : A \to A'$ are morphisms in $\mathcal{C}$ satisfying $Ff_1 = Ff_2$ in $\mathcal{B}$, then $f_1 = f_2$.

(2) If F is a faithful functor, then F is called an *embedding* if it is also injective on objects. This means that the F identifies $\mathcal{A}$ with a subcategory of $\mathcal{B}$.

(3) F is *full* if for all $A, B \in \mathrm{Ob}(\mathcal{A})$, the map $\mathrm{Hom}_{\mathcal{C}}(A, B) \to \mathrm{Hom}_{\mathcal{B}}(FA, FB)$ is surjective, i.e., if $g \in \mathrm{Hom}_{\mathcal{B}}(FA, FA')$, there exist $f \in \mathrm{Hom}_{\mathcal{C}}(A, A')$ satisfying $Ff = g$.

(4) F is called *essentially surjective on objects* if for every $B \in \mathrm{Ob}(\mathcal{B})$ there exists $A \in \mathrm{Ob}(\mathcal{A})$ such that B is isomorphic to $F(A)$.

(5) A subcategory of a category is called a *full subcategory* if the inclusion functor is full.

(6) If F is full and faithful, it is also called a *fully faithful functor*. In this case, the image $F(\mathcal{A})$ is a full subcategory of $\mathcal{B}$ and we say that F is a *full embedding*.

(7) A category $\mathcal{A}$ is called a *concrete* category if there exists a faithful functor $G : \mathcal{A} \to \mathbf{Set}$.

Examples 135.8. 1. **Ab** is full subcategory of **Group**.

2. ${}_R\mathcal{M}_{fg} :=$ ((finitely generated R-modules, R-homomorphisms)) is a full subcategory of ${}_R\mathcal{M}$
3. Let F be a field and $\mathbf{Vector}_F$ the category of F-vector spaces. Then $(\mathbf{Vector}_F)_{fd} :=$ ((finite dimensional F-vector spaces, linear transformations.)) is a full subcategory of $\mathbf{Vector}_F$.
4. Let G be a group and H, H' subgroups. By Exercise 19.10(3), (4), (5), (6), we can identify H with the left G-set of cosets G/H and G-equivariant maps $\varphi : G/H \to G/H'$ with $\varphi = \lambda_x$, $xgH' = gxH'$ for an $x \in G$ satisfying $x^{-1}Hx \subset H'$ for all $g \in G$. Let $\mathbf{Orb}_G :=$ ((subgroups of G, G-equivariant maps)).

   Let K/F be a finite Galois extension and $\mathbf{Field} :=$ ((fields, field monomorphisms)) and $\mathbf{Field}_F^K$ be the subcategory of the slice category $F/\mathbf{Field}$ of intermediate fields $K/E/F$ with morphisms field automorphisms fixing F. As usual $G(K/E)$ is the group of F-automorphisms of E.

   Let $G = G(K/F)$ and $\Phi : \mathbf{Orb}_G \to \mathbf{Field}_F^K$ be the functor that sends an object $H \subset G$ to K^H and a morphism $G/H \to G/H'$ induced by λ_x, $xgH' = gxH'$ for all $g \in G$ to the field homomorphism $y \to yx$ takes and element $y \in K$ fixed by E to an element $xy \in E$ fixed by H. Then the Fundamental Theorem of Galois Theory says that the map $\Phi : \mathbf{Orb}_G^{op} \to \mathbf{Field}_F^K$ take $E \in \mathrm{Ob}(\mathbf{Field}_F^K)$ to $G(K/E)$. By the Fundamental Theorem of Galois Theory, Φ is a bijection of objects. In fact, $\mathbf{Orb}_G^{op} \cong \mathbf{Field}_F^K$ using Galois theory which we leave as an exercise.

The notion of isomorphism of categories is too strong. We define a weaker equivalence relation between categories. It is convenient to write morphisms, etc., using diagrams, as we did the main part of the book.

Definition 135.9. Let $F, G : \mathcal{A} \to \mathcal{B}$ be functors of categories. A *natural transformation* $\alpha : F \to G$ associates:

1. To each $A \in \mathrm{Ob}(\mathcal{A})$ a morphism $\alpha(A) : FA \to GA$ in $\mathcal{B}$.
2. If $f : A \to A'$ is a morphism in $\mathcal{A}$, then we have a commutative diagram

$$\begin{tikzcd} FA \arrow[r, "Ff"] \arrow[d, "\alpha(A)"'] & FA' \arrow[d, "\alpha(A')"] \\ G(A) \arrow[r, "Gf"'] & GA'), \end{tikzcd}$$

i.e., $Gf\big(\alpha(A)\big) = \alpha(A')\,Ff$. We call the collection $\alpha(A)$ (also written as α_A) the *components* of the natural transformation α.

A natural transformation $\alpha : F \to G$ is called a *natural isomorphism* if for every $A \in \mathrm{Ob}(\mathcal{A})$, the component $\alpha(A)$ is an isomorphism. In this case, the $\alpha(A)^{-1}$ in B are the components of a natural isomorphism $\alpha^{-1} : G \to F$. If there exists a natural isomorphism $\alpha : F \to G$, we write $F \cong G$.
We call two categories $\mathcal{A}$ and $\mathcal{B}$ *equivalent* if there exist functors $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ together with natural isomorphisms $\alpha : 1_{\mathcal{A}} \to GF$ and $\beta : FG \to 1_{\mathcal{B}}$. If this is the case, we write $\mathcal{A} \simeq \mathcal{B}$.

We leave the following characterization of equivalence of categories as an exercise.

Theorem 135.10. *A functor $F : \mathcal{A} \to \mathcal{B}$ defines an equivalence of categories if and only if it is full, faithful, and essentially surjective on objects.*

If $\mathcal{C}$ is a locally small category, we let $\mathrm{Nat}(F, G)$ denote the of natural transformations $F \to G$.

Examples 135.11. 1. If $V \in \mathbf{Vector}_F$, as usual let $V^* = \mathrm{Hom}_F(V, F)$ be its dual space. Let $D : V \to V^*$ be the duality functor, i.e., the functor sending objects $V \to V^*$ and morphisms $f : V \to W$ to $f^* : W^* \to V^*$ where $\beta \mapsto \beta f$. Let $D^2 = D \circ D$. Then $h : 1 \to D^2$ by $h(V) \mapsto V^{**}$ given by $h(V) : x \mapsto L_x$, where $L_x(f) = f(x)$, is a natural transformation and induces a natural isomorphism $D^2 : (\mathbf{Vector}_F)_{fd} \to (\mathbf{Vector}_F)_{fd}$. This property of the double dual was the motivation for the beginning of category theory.

2. **Morita equivalence**. There exists an equivalence of categories ${}_R\mathcal{M} \cong {}_{\mathbb{M}_n(R)}\mathcal{M}$:

Let A be an $(R\text{-}\mathbb{M}_n(R))$-bimodule of row vectors of length n with entries in R and B be an $(\mathbb{M}_n(R)\text{-}R)$-bimodule of column vectors of length n with entries in R. Define functors

$$F : {}_R\mathcal{M} \to {}_{\mathbb{M}_n(R)}\mathcal{M} \text{ by } F(M) = B \otimes_R M \text{ and } F(f) = 1_B \otimes f$$

for $f \in \operatorname{Hom}_R(M, M')$ with $M, M' \in {}_R\mathcal{M}$ and

$$G : {}_{\mathbb{M}_n(R)}\mathcal{M} \to {}_R\mathcal{M} \text{ by } G(N) = A \otimes_{\mathbb{M}_n(R)} N \text{ and } G(g) = 1_A \otimes g$$

for $g \in \operatorname{Hom}_{\mathbb{M}_n(R)}(N, N')$ with $N, N' \in {}_R\mathcal{M}$. Since multiplication induces isomorphisms

$$A \otimes_{\mathbb{M}_n(R)} B \xrightarrow{\sim} R \text{ and } B \otimes_R A \xrightarrow{\sim} \mathbb{M}_n(R),$$

both $G \circ F$ and $F \circ G$ are naturally isomorphic to identity functors.

Note that Morita equivalence is a generalization of the classification of simple left Artinian rings by Wedderburn Theory 102.

Let $F, G : \mathcal{A} \times \mathcal{B} \to \mathcal{D}$ be bifunctors. Suppose for each $A \in \operatorname{Ob}(\mathcal{A})$, $B \in \operatorname{Ob}(\mathcal{B})$, we have a morphism $\alpha(A, B) : F(A, B) \to G(A, B)$ in $\mathcal{D}$. We say α is *natural* in A, if for each $B \in \operatorname{Ob}(\mathcal{B})$, the components of $\alpha(A, B)$ for all A define a natural transformation $\alpha(\ \ , B) : F(\ \ , B) \to G(\ \ , B)$ of functors $A \to D$. Similarly, we define α natural B. We have

Proposition 135.12. *Let $F, G : \mathcal{A}\times\mathcal{B} \to \mathcal{D}$ be bifunctors. Let $\alpha(A, B) : F(A, B) \to G(A, B)$ in $\mathcal{D}$ be a morphism in $\mathcal{D}$ for every pair of objects A in $\mathcal{A}$ and B in $\mathcal{B}$. Then $\alpha : F \to G$ is a natural transformation (of bifunctors) if and only if $\alpha(A, B)$ is natural in A for each $B \in \operatorname{Ob}(\mathcal{B})$ and natural in B for each $A \in \operatorname{Ob}(A)$.*

We leave the proof as an exercise.

Exercises 135.13. 1. Prove Proposition 135.4.

2. Prove the assertions in Examples 135.5.
3. Show in Example 135.8(4), we have an isomorphism of categories.
4. Prove Theorem 135.10.
5. Prove Proposition 135.12.

136. Yoneda's Lemma

Definition 136.1. Let $\mathcal{C}$ and $\mathcal{D}$ be categories and $\operatorname{Funct}(\mathcal{C}, \mathcal{D})$ be the collection of functors $\mathcal{C} \to \mathcal{D}$. Then

$$\mathcal{D}^{\mathcal{C}} := \big(\big(\operatorname{Funct}(\mathcal{C}, \mathcal{D}), \operatorname{Nat}(\operatorname{Funct}(\mathcal{C}, \mathcal{D})\big)\big)$$

is called a *functor category*. The functor category $\mathcal{D}^{\mathcal{C}}$ is also denoted by $\mathbf{Funct}(\mathcal{C}, \mathcal{D})$. If we wish to deal with contravariant functors from $\mathcal{C}$, we will use functors in $\mathcal{D}^{\mathcal{C}^{op}} = \mathbf{Funct}(\mathcal{C}^{op}, \mathcal{D})$. If both $\mathcal{C}$ and $\mathcal{D}$ are small, then so is $\mathcal{D}^{\mathcal{C}}$. If $\mathcal{C}$ is small and $\mathcal{D}$ locally small, then $\mathcal{D}^{\mathcal{C}}$ is locally small. In general, this is not the case, in particular, if we only assume that $\mathcal{C}$ is locally small. One reason that we defined the concept of category instead of only defining locally small categories is that one wishes to view a functor category as a category.

In this section, we shall be interested in the functor category $\mathbf{Funct}(\mathcal{C}, \mathbf{Sets})$ when $\mathcal{C}$ is locally small. We call such functors *set valued functors* and the set $\mathrm{Hom}_{\mathcal{C}}(A, B)$ is usually called a *hom-set*. An object in $\mathbf{Funct}(\mathcal{C}, \mathbf{Sets})$ is called a *representable functor*. We also call an object of $\mathbf{Funct}(\mathcal{C}^{op}, \mathbf{Sets})$ when $\mathcal{C}$ is locally small a *representable functor*.

Notation 136.2. Since objects in our category $\mathcal{C}$ may not be sets, we must define the analog of $x \in S$, when S is a set. We do this as follows: If $T \in \mathrm{Ob}(\mathcal{C})$, call an arbitrary morphism $\alpha : T \to A$ in $\mathcal{C}$ a *variable element* of A parametrized by T. When α is treated as a variable element of A parametrized by T and f has domain A, we will write $f(\alpha)$ for $f \circ \alpha$. For example, in this notation, f is a monomorphism if and only if for any variable elements $x, y : T \to A$, if $x \neq y$, then $f(x) \neq f(y)$.

Let $\mathcal{C}$ be a locally small category. If $X, Y \in \mathrm{Ob}(\mathcal{C})$, then we define two functors. (Cf. Example 135.5(2).) The first is

$$h_X = \mathrm{Hom}_{\mathcal{C}}(X, \ \ ) : \mathcal{C} \to \mathbf{Set}$$

with $h_X(B) = \mathrm{Hom}_{\mathcal{C}}(X, B)$ for $B \in \mathcal{C}$ and a morphism $B \xrightarrow{g} B'$ in $\mathrm{Ob}(\mathcal{C})$ goes to

$$\mathrm{Hom}_{\mathcal{C}}(X, B) \xrightarrow{h_X(g)} \mathrm{Hom}_{\mathcal{C}}(X, B') \ \text{ by } \ \varphi :\mapsto g\varphi.$$

We also write $h_X(g)$ as $g_*(X)$ (or just g_* if X is clear). Then h_X is a functor in $\mathrm{Funct}(\mathcal{C}, \mathbf{Sets})$.

The second functor is

$$h^Y = \mathrm{Hom}_{\mathcal{C}}(\ , Y) : \mathcal{C}^{op} \to \mathbf{Set}$$

with $h^Y(A) = \mathrm{Hom}_{\mathcal{C}}(A, Y)$ for $A \in \mathcal{O}(\mathcal{C})$ and a morphism $A' \xrightarrow{f} A$ in $\mathcal{C}$ goes to

$$\mathrm{Hom}_{\mathcal{C}}(A', Y) \xrightarrow{h^Y(f)} \mathrm{Hom}_{\mathcal{C}}(A, Y) \ \text{ by } \varphi \mapsto \varphi f.$$

We also write $h^Y(f)$ as $f^*(Y)$ (or just f^* if Y is clear). This is a functor in $\mathrm{Funct}(\mathcal{C}^{op}, \mathbf{Sets})$ and can be viewed as a contravariant functor $\mathcal{C} \to \mathbf{Set}$.

Note that for all $X, Y \in \mathrm{Ob}(\mathcal{C})$,

$$h_X(Y) = h^Y(X). \tag{136.3}$$

Let $B \xrightarrow{g} B'$ and $A' \xrightarrow{f} A$ be morphisms in $\mathcal{C}$. If $A \xrightarrow{\varphi} B$ is also a morphism in $\mathcal{C}$, then $(g\varphi)f = g(\varphi f)$. That is, we have a commutative diagram

$$\begin{tikzcd} h_A(B) \arrow[r, "g_*(A)"] \arrow[d, "f^*(B)"'] & h_A(B') \arrow[d, "f^*(B')"] \\ h_{A'}(B) \arrow[r, "g_*(A')"'] & h_{A'}(B'). \end{tikzcd} \tag{136.4}$$

This means that we can define two natural transformations as follows:

Let $X_2 \xrightarrow{f} X_1$ in $\mathcal{C}$. For all $B \in \mathcal{C}$, define $h^f : h_{X_1} \to h_{X_2}$ by

$$h^f(B) = h_{X_2}(B)$$

on objects and

$$h_{X_1}(B) \xrightarrow{h^f(B)} h_{X_2}(B) \ \text{ by } \ \varphi \mapsto f^*(B)\varphi = \varphi f$$

on morphisms. This is a natural transformation using $X_1 = A$ and $X_2 = A'$ in Diagram 136.4. The natural transformation $h^f : h_X \to h_{X'}$ is called the *induced natural transformation* corresponding to f. We also write the component of $h^f(B)$ of h^f at B by $\mathrm{Hom}_{\mathcal{C}}(f, B)$. We also write h^f by $h(f)$ if we know that f is contravariant.

Let $Y_1 \xrightarrow{g} Y_2$ in $\mathcal{C}$. For all $A \in \mathcal{C}$, define $h_g : h^{Y_1} \to h^{Y_2}$ by

$$h_g(A) = h^{Y_2}(A)$$

on objects, and

$$h^{Y_1}(A) \xrightarrow{h_g(A)} h^{Y_2}(A) \text{ by } \varphi \mapsto g_*(A)\varphi = g\varphi$$

on morphisms. This is a natural transformation using $Y_1 = B$ and $Y_2 = B'$ in Diagram 136.4. The natural transformation $h_g : h^{Y_1} \to h^{Y_2}$ is called the *induced natural transformation* corresponding to g. We also write the component of $h_g(A)$ of h_g at A by $\mathrm{Hom}_{\mathcal{C}}(A, g)$. We also write h_g by $h(g)$ if we know g is covariant.

In particular, if $\mathcal{C}$ is a locally small category, we have constructed a functor, called the *Yoneda functor*,

$$h : \mathcal{C}^{op} \to \mathbf{Funt}(\mathcal{C}, \mathbf{Set})$$

that satisfies

(*i*) If $X \in \mathrm{Ob}(\mathcal{C})$, then $h(X) = h_X$.
(*ii*) If $f : X' \to X$ in $\mathcal{C}$ and $B \in \mathrm{Ob}(\mathcal{C})$, then a component $h^f(B)$ of $h^f : h_X \to h_{X'}$ is $h^f(B) = \mathrm{Hom}(f, B) : \mathrm{Hom}_{\mathcal{C}}(X, B) \to \mathrm{Hom}_{\mathcal{C}}(X', B)$.

So in this notation, Diagram 136.4 is

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{C}}(X, B) \arrow[r, "\mathrm{Hom}_{\mathcal{C}}(X,g)"] \arrow[d, "h^f(B)"'] & \mathrm{Hom}_{\mathcal{C}}(X, B') \arrow[d, "h^f(B')"] \\
\mathrm{Hom}_{\mathcal{C}}(X', B) \arrow[r, "\mathrm{Hom}_{\mathcal{C}}(X',g)"'] & \mathrm{Hom}_{\mathcal{C}}(X', B')
\end{tikzcd}$$

commutes where $X, X' \in \mathrm{Ob}(\mathcal{C})$ and $g : B \to B'$ in $\mathcal{C}$.

Note that $h(X)$ is a hom functor and $h^f(B)$ is a component of a contravariant hom functor.

Theorem 136.5. (Yoneda Embedding) *Let $\mathcal{C}$ be a locally small category. Then the functor $h : \mathcal{C}^{op} \to \mathbf{Funct}(\mathcal{C}, \mathbf{Set})$ is full and faithful.*

Proof. h is faithful: Let $f_1, f_2 : X' \to X$ in $\mathcal{C}$. Then the component $h(f_i)(X) : h_X(X) \to h_{X'}(X)$ of the natural transformation $h(f_i)$ at X, takes $1_X \mapsto f_i$ for $i = 1, 2$. If $f_1 \neq f_2$, then $h(f_1) \neq h(f_2)$. So h is faithful.

h is full: Let $\alpha : h_X \to h_{X'}$. Define $f : X' \to X$ by $f = \alpha(X)(1_X)$. The component of α at X is a function $\alpha(X) : \mathrm{Hom}_{\mathcal{C}}(X, X) \to \mathrm{Hom}_{\mathcal{C}}(X', X)$ in **Set**, so it is well-defined. We must show, if $k : X \to A$ in $\mathcal{C}$, then $\alpha(A)(k) = k \circ f : X' \to A$.

Consider the diagram

$$(*)\qquad \begin{array}{ccc} \mathrm{Hom}_{\mathcal{C}}(X,X) & \xrightarrow{h_X(k)_*} & \mathrm{Hom}_{\mathcal{C}}(X,A) \\ {\scriptstyle\alpha(X)}\big\downarrow & & \big\downarrow{\scriptstyle\alpha(A)} \\ \mathrm{Hom}_{\mathcal{C}}(X',X) & \xrightarrow[h_{X'}(k)_*]{} & \mathrm{Hom}_{\mathcal{C}}(X',A). \end{array}$$

Starting at the top left hand corner evaluating at 1_X gives $1_X \mapsto \alpha(A)(k)$ going along the top first and $1_X \mapsto k \circ f$ going down first. This proves that (*) commutes, hence h is full. □

Corollary 136.6. *Let $\mathcal{C}$ be a locally small category. Then every natural transformation $h_X \to h_{X'}$ is given by composition with a unique morphism $X' \to X$. Such a natural transformation is an isomorphism if and only if the corresponding morphism $X' \to X$ is an isomorphism. In particular, if $F : \mathcal{C} \to$ **Set** is represented by both X and X', then $X \cong X'$.*

This corollary means that one can construct a morphism in a locally small category by constructing a natural transformations of hom functors.

The dual statement to Corollary 136.6 also holds, i.e., if $\mathcal{C}$ is a locally small category, there exists a *Yoneda functor*

$$h' : \mathcal{C} \to \mathbf{Funct}(\mathcal{C}^{op}, \mathbf{Set})$$

that is full and faithful. If $Y \in \mathrm{Ob}(\mathcal{C})$, then $h'(Y) = h^Y$, a contravariant functor. If $g : Y \to Y'$ in $\mathcal{C}$ and $A \in \mathrm{Ob}(\mathcal{C})$, then the component $h'(g)(A) : h^Y(A) \to h^{Y'}(A)$ of the natural transformation $h'(g) : h^Y \to h^{Y'}$ is $\mathrm{Hom}(A, g) : h^Y(A) \to h^{Y'}(A)$.

Since h' is full and faithful, a morphism from $Y \to Y'$ in $\mathcal{C}$ can be uniquely defined by a natural transformation $\alpha : h^Y \to h^{Y'}$. The components of $\alpha(X) : h^Y(X) \to h^{Y'}(X)$ for each $X \in \mathrm{Ob}(\mathcal{C})$ are just $\alpha(X) : \mathrm{Hom}_{\mathcal{C}}(X,Y) \to \mathrm{Hom}_{\mathcal{C}}(X,Y_2)$. As $\mathcal{C}$ is locally small, it follows that each morphism $Y \to Y'$ in $\mathcal{C}$ can be prescribed by a variable element Y' parametrized by X, in such a way that for each $f : X' \to X$, the diagram

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{C}}(X,Y) & \xrightarrow{\alpha(X)} & \mathrm{Hom}_{\mathcal{C}}(X,Y) \\ {\scriptstyle h^X(A)(f)^*}\big\downarrow & & \big\downarrow{\scriptstyle h^B(f)^*} \\ \mathrm{Hom}_{\mathcal{C}}(X',Y) & \xrightarrow[\alpha(X')]{} & \mathrm{Hom}_{\mathcal{C}}(C',Y') \end{array}$$

commutes.

What we have shown is that any natural transformation $\alpha : h^X \to h^{X'}$ is given by a unique morphism $X' \to X$ in $\mathcal{C}$. We shall now show that we have a more general conclusion. Namely, we shall show that the same is true if we replace $h^{X'}$ by any set functor $F : \mathcal{C} \to$ **Set**.

Construction 136.7. Let $\mathcal{C}$ be a locally small category and $F : \mathcal{C} \to$ **Set** a functor. Let $X \in \mathrm{Ob}(\mathcal{C})$ and x an element of the set $F(X)$. Then we want to show that x induces a natural transformation $\tau_x : h_X \to F$ by the formula

$$f \mapsto F(f)(x). \tag{136.8}$$

If $f : X \to X'$ is a morphism in $\mathcal{C}$, i.e., an element in $\mathrm{Hom}_{\mathcal{C}}(X, X')$, and $F : \mathcal{C} \to \mathbf{Set}$ is a functor, we must have the induced morphism is a set map $F(f) : F(X) \to F(X')$ and this map can be evaluated at $x \in F(X)$. So we want to show this is a natural transformation.

Proposition 136.9. *Let $\mathcal{C}$ be a locally small category and $F : \mathcal{C} \to \mathbf{Set}$ a functor. Let $X \in \mathrm{Ob}(\mathcal{C})$. Then each $x \in F(X)$ determines a natural transformation $\tau_x : h_X \to F$ by equation (136.8).*

Proof. Let $\alpha(X') : h_X(X') \to F(X')$ take $f \mapsto F(f)(x)$ for $x \in F(X)$. We must show that for any $g : X' \to A$ in $\mathcal{C}$, the diagram

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(X, X') \arrow[r, "\alpha(X')"] \arrow[d, "h_X(A)(g)_*"'] & F(X') \arrow[d, "F(g)"] \\ \mathrm{Hom}_{\mathcal{C}}(X, A) \arrow[r, "\alpha(A)"'] & F(A) \end{tikzcd}$$

commutes. But if $f : X \to X'$ is a morphism in $\mathcal{C}$, we have

$$\begin{aligned} \alpha(A)\big(h_X(g)\big)(f) &= \alpha(A)(gf) = F(gf)(x) \\ &= F(g)(F(f)(x) = F(g)\big(\alpha(X')\big)(f) \end{aligned}$$

as required. □

Theorem 136.10. (Yoneda's Lemma) *Let $\mathcal{C}$ be a locally small category and $F : \mathcal{C} \to \mathbf{Set}$ a functor. For each $X \in \mathrm{Ob}(\mathcal{C})$, let $h_X \to F$ be the natural transformation induced by $f \mapsto F(f)(x)$ for $x \in F(X)$. Then this defines a bijection $\psi : \mathrm{Nat}(h_X, F) \to F(X)$. In particular, if $F : \mathcal{C} \to \mathbf{Set}$ is represented by both X and X', then $X \cong X'$. Moreover, this bijection is natural in X and F when both sides of ψ are viewed as functors $\mathcal{C} \times \mathbf{Funct}(\mathcal{C}, \mathbf{Set}) \to \mathbf{Set}$.*

Proof. If x_1, x_2 are distinct elements in $F(X)$, then the natural transformation corresponding to x_i takes 1_X to x_i for $i = 1, 2$. Therefore, the map is injective.
Suppose that $\alpha : h_X \to F$ is a natural transformation. Then we have $\alpha(X) : h_X \to F(X)$. Let $x = \alpha(X)(1_X) \in F(X)$. If $f : X \to X'$ is a morphism in $\mathcal{C}$, then

$$\alpha(X')\big(h_X(f)\big)(1_X) = F(f)\big(\alpha(X)\big)(1_X)$$

by the naturality of α. The left hand side is $\alpha(X')(f)$ and the right hand side is $F(f)(x)$. Therefore α is the natural transformation given by Proposition 136.9. It follows that the map in Yoneda's Lemma is surjective.

Naturality on functors asserts the following: Given a natural transformation $\beta : F \to G$ of functors in $\mathcal{C} \to \mathbf{Set}$, the element $x \in G(X)$ representing the composite natural transformation $\beta\alpha$ satisfies

$$\begin{tikzcd} \mathrm{Nat}(h_X, F) \arrow[r, "\psi(F)"] \arrow[d, "\beta_*"'] & F(X) \arrow[d, "\beta(X)"] \\ \mathrm{Nat}(h_X, G) \arrow[r, "\psi(G)"] & G(X) \end{tikzcd} \tag{136.11}$$

commutes in $\mathbf{Set}$ where $\beta(X) : F(X) \to G(X)$ is represented by $\alpha : h_X \to F$.

Let $\alpha\beta$ be the composite natural transformation defined in Exercise 136.19(1). By definition $\psi_G(\beta\alpha) = (\beta\alpha)(1_X)$. This is just $\beta(X)\big(\alpha(X)(1_X)\big)$ by the composition of natural transformations in Exercise 136.19(1).

Naturality on objects asserts the following: Let $A \xrightarrow{f} X$ in $\mathcal{C}$ be the element in $F(A)$ representing the composite natural transformation $h_A \xrightarrow{f_*} h_X \xrightarrow{\alpha} F$. Then it is the image under $Ff : F(X) \to \mathrm{F}(A)$ of the element in $F(X)$ representing α, i.e., the diagram

$$\begin{tikzcd}
\mathrm{Nat}(h_X, F) \arrow[r, "\psi(F)"] \arrow[d, "\beta_*"'] & F(X) \arrow[d, "\beta(X)"] \\
\mathrm{Nat}(h_A, F) \arrow[r, "\psi(G)"] & G(A)
\end{tikzcd}$$

commutes in **Set**.

The image of α in the diagram starting along the top-right is $Ff\big(\alpha_X(1_X)\big)$. The image starting along the left vertical is $(\alpha f_*)_A(1_A)$. By definition of the composition of natural transformations in Exercise 136.19(1),

$$h_X(X) \xrightarrow{f^*} h_X(A) \xrightarrow{\alpha_A} F(A) \text{ takes } 1_X \mapsto f \mapsto \alpha_A(f)$$

establishing the needed commutativity. Applying the Yoneda functor h gives the result. $\square$

Corollary 136.12. *If $\mathcal{C}$ is a locally small category, then $\mathbf{Funct}(\mathcal{C}^{op}, \mathbf{Set})$ is locally small.*

Of course, the contravariant version Yoneda's Lemma is also true.

Theorem 136.13. *Let $\mathcal{C}$ be a locally small category. For each $C \in \mathrm{Ob}(\mathcal{C})$ and $F : \mathcal{C}^{op} \to \mathbf{Set}$, let $h^X \to F$ be the natural transformation induced by $f \mapsto F(f)(x)$ for $x \in F(X)$. Then this defines a bijection $\phi : \mathrm{Nat}(h^X, F) \to F(X)$. In particular, if $F : \mathcal{C}^{op} \to \mathbf{Set}$ is represented by both X and X', then $X \cong X'$. Moreover, this bijection is natural in X and F when both sides of ϕ are viewed as functors $\mathcal{C}^{op} \times \mathbf{Funct}(\mathcal{C}^{op}, \mathbf{Set}) \to \mathbf{Set}$.*

Examples 136.14. 1. (Cayley's Theorem) Let G be a group. We view G as the category $\mathbf{B}G$. The Yoneda lemma 136.10, identifies the image $h' : \mathbf{B}G \to \mathbf{Funct}\big((\mathbf{B}G)^{op}, \mathbf{Set}\big)$ of the Yoneda Embedding Theorem 136.5 as the right G-set G with G acting by right multiplication. By the Yoneda embedding, the only G-invariant endomorphisms of the right G-set are those maps defined by left multiplication by a fixed element of G. In particular, any G-equivariant endomorphism of G must be an automorphism.
Therefore, we have an isomorphism between G and the automorphism group of the right G-sets G. This automorphism group is an object in $\mathbf{Funct}\big((\mathbf{B}G)^{op}, \mathbf{Set}\big)$. Composing this with the faithful forgetful functor $\mathbf{Funct}\big((\mathbf{B}G)^{op}, \mathbf{Set}\big) \to \mathbf{Set}$, we obtain an isomorphism between G and a subgroup of the automorphism group $\Sigma(G)$ of the set G.

2. Let R be a commutative ring. Let $\mathbf{Calg}_R$:= ((commutative R-algebras, R-algebra homomorphisms)). The tensor product $\otimes_R$ is a coproduct in this category. In Exercise 92.31(11), we defined an affine scheme $(\operatorname{Spec} R, R)$. If S is an R-algebra, We call the affine scheme $(\operatorname{Spec} S, S)$ an R-scheme when S is viewed as an R-algebra. Then

$(\mathbf{Calg}_R)^{op}$ is naturally equivalent to the category of ((affine R-schemes, scheme morphisms)). In the category $\mathbf{Funct}(\mathbf{Calg}_R, \mathbf{Set})$, let $\mathcal{R}_{\mathbf{Calg_R}}$ be the full subcategory of representable functors. Then it is naturally equivalent to the category of R-schemes in algebraic geometry.

3. Let $\mathcal{G}_{\mathbf{Calg}_R}$ be the full category in the category $\mathbf{Funct}(\mathbf{Calg}_R, \mathbf{Group})$ of representable functors. Then $\mathcal{G}_{\mathbf{Calg}_R}$ is naturally equivalent to the category of *(affine) group schemes over* R. If $h_X \in \mathbf{Funct}(\mathbf{Calg}_R, \mathbf{Set})$ and $T \in \mathbf{Calg}_R$, then $h_X(T)$ is called the set of T-valued points of h. This is quite useful when R is not an algebraically closed field. As a specific example, let $\mathbf{GL}_2 : \mathbf{Calg}_R \to \mathbf{Group}$ by $T \mapsto \mathrm{GL}_2(T)$. If $T \xrightarrow{f} T'$ in $\mathbf{Calg}_R$, let $\mathrm{GL}_2\, T \to GL_2 T'$ be the obvious map. Let
$$A = R[t_{11}, t_{12}, t_{21}, t_{22}, \frac{1}{t_{11}t_{222} - t_{12}, t_{21}}].$$
Then $\mathbf{GL}_2 \cong h_A$ is an (affine) group scheme.

Definition 136.15. Let $\mathcal{C}$ be a locally small category and $F : \mathcal{C} \to \mathbf{Set}$. An element $x \in F(X)$, $X \in \mathrm{Ob}(\mathcal{C})$, is called a *universal element* of F if the natural transformation $\alpha(X) : h_X \to F$ induced by x is an isomorphism.

If there exists a universal element for F, it is, of course, representable. By Yoneda's Lemma, every representable functor is induced by a unique element $x \in F(X)$. Moreover, α is an isomorphism if and only if x is a universal element. So the converse is also true. To compute x, we use the following:

Proposition 136.16. *Let $\mathcal{C}$ be a locally small category and $\alpha : h_X \to F$ a natural isomorphism. Then the unique universal element $x \in F(X)$ inducing α is $\alpha(X)(1_X)$.*

PROOF. Let $f : X \to X'$ be a morphism in $\mathcal{C}$. Then $\alpha(X')(f) = F(f)\big(\alpha(X)(1_X)\big)$, since the diagram
$$\begin{array}{ccc} h^X(X) & \xrightarrow{\alpha(X)} & F(X) \\ {\scriptstyle f(X)_*}\big\downarrow & & \big\downarrow{\scriptstyle f(X)_*} \\ h^X(X') & \xrightarrow[\alpha(X')]{} & F(X'). \end{array}$$
commutes. Then $\alpha(X)(1_X)$ is the required unique x $\square$

An alternative way of describing a universal element is given by the following:

Proposition 136.17. *Let $\mathcal{C}$ be a locally small category, $F : \mathcal{C} \to \mathbf{Set}$ a functor, $X \in \mathrm{Ob}(\mathcal{C})$, and $x \in F(X)$. Then x is a universal element of F if and only if given any object $A \in \mathrm{Ob}(\mathcal{C})$ and element $a \in F(A)$, there exists a unique morphism $f : X \to A$ in $\mathcal{C}$ satisfying $a = (Ff)(x)$.*

PROOF. ($\Rightarrow$): If $A \cong X$ in $\mathrm{Ob}(\mathcal{C})$, then $h_A \cong h_X$ by Proposition 136.9. Therefore all the components must be bijective. It follows that such an f exists and is unique.
($\Leftarrow$): The uniqueness of f for each $A \in \mathrm{Ob}(\mathcal{C})$ and $x \in F(A)$ determines a bijection $\alpha(A) : h_A \to F(A)$ for every A that takes $X \xrightarrow{f} A$ to $F(f)(x)$. By Proposition 136.9, this determines a natural transformation, and it is then a natural isomorphism. $\square$

We leave the following uniqueness consequence as an exercise.

Corollary 136.18. *Let $\mathcal{C}$ be a locally small category, $F : \mathcal{C} \to$ **Set** a functor. If $X, X' \in \operatorname{Ob}(\mathcal{C})$ have universal elements x, x', respectively, then there exists a unique isomorphism $f : X \to X'$ satisfying $F(f)(x) = x'$*

Exercise 136.19. 1. Let $\alpha : F \to G$ and $\beta : G \to H$ be natural transformations between functors $F, G, H : \mathcal{C} \to \mathcal{D}$. Then there is a natural transformation $\alpha\beta : F \to H$ satisfying $(\alpha\beta)(C) = \beta(C)\alpha(C)$ for all $C \operatorname{Ob}(\mathcal{C})$, the components of β and α.

2. Prove Corollary 136.18.
3. Show how the covariant and contravariant Yoneda embeddings are really different incarnations of a common bifunctor $C^{op} \times \mathcal{C} \to$ **Sets** for any locally small category $\mathcal{C}$.
4. Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be categories. Prove the following:
   (i) Let $G : \mathcal{C} \times \mathcal{B} \to \mathcal{A}$ be a functor such that the functor
$$\operatorname{Hom}_{\mathcal{B}}\big(A, G(X,\ \ )\big) : \mathcal{B} \to \mathbf{Set}$$
   is representable for every object $A \in \operatorname{Ob}(\mathcal{A})$ and $X \in \operatorname{Ob}(\mathcal{C})$. Then there exists a unique functor $F : \mathcal{A} \times \mathcal{C}^{op} \to \mathcal{B}$ satisfying
$$\operatorname{Hom}_{\mathcal{A}}\big(\ \ , G(\ \ ,\ \ )\big) \cong \operatorname{Hom}_{\mathcal{B}}\big(F(\ \ ,\ \ ),\ \ \big)$$
   as functors $\mathcal{A}^{op} \times \mathcal{C} \times \mathcal{B} \to$ **Set**.
   (ii) Let $F : \mathcal{A} \times \mathcal{C}^{op} \to \mathcal{B}$ be a functor such that the functor
$$\operatorname{Hom}_{\mathcal{B}}\big(F(\ \ , Y), B,\big) : \mathcal{A} \to \mathbf{Set}$$
   is representable for every object $B \in \operatorname{Ob}(\mathcal{B})$ and $Y \in \operatorname{Ob}(\mathcal{C})$. Then there exists a unique functor $G : \mathcal{C} \times \mathcal{B} \to \mathcal{A}$ satisfying
$$\operatorname{Hom}_{\mathcal{A}}\big(\ \ , G(\ \ ,\ \ )\big), \cong \operatorname{Hom}_{\mathcal{B}}\big(F(\ \ ,\ \ ),\ \ \big)$$
   as functors $\mathcal{A}^{op} \times \mathcal{C} \times \mathcal{B} \to$ **Set**.

137. Adjoints

Let $\mathcal{A}$ and $\mathcal{B}$ be locally small categories and $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ two functors. We say that F is a *left adjoint* of G and G is a *right adjoint* of F (or (F, G) is an *adjoint pair*), if there exists a natural transformation of bifunctors
$$\tau : \operatorname{Hom}_{\mathcal{B}}\big(F(\ \ ),\ \ \big) \to \operatorname{Hom}_{\mathcal{A}}\big(\ \ , G(\ \ )\big)$$
that is a natural isomorphism of bifunctors $\mathcal{A}^{op} \times \mathcal{B} \to$ **Set** , i.e., for all $A \in \mathcal{A}$ and for all $B \in \mathcal{B}$,
$$\tau_{A,B} : \operatorname{Hom}_{\mathcal{B}}\big(F(A), B\big) \to \operatorname{Hom}_{\mathcal{A}}\big(A, G(B)\big)$$
is a bijection of sets and if $A \xrightarrow{f} A'$ in $\mathcal{A}$ and $B \xrightarrow{g} B'$, then we have a commutative diagram:

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{B}}(FA', B) \arrow[r, "(Ff)^*(B)"] \arrow[d, "\tau_{A',B}"'] & \mathrm{Hom}_{\mathcal{B}}(FA, B) \arrow[r, "g_*(FA)"] \arrow[d, "\tau_{A,B}"'] & \mathrm{Hom}_{\mathcal{B}}(FA, B') \arrow[d, "\tau_{A,B'}"'] \\
\mathrm{Hom}_{\mathcal{C}}(A', GB) \arrow[r, "f^*(GB)"] & \mathrm{Hom}_{\mathcal{C}}(A, GB) \arrow[r, "(Gg)_*(A)"] & \mathrm{Hom}_{\mathcal{C}}(A, GB').
\end{tikzcd}$$

In particular,

$$\begin{tikzcd}
h_{FA'}(B) \arrow[r, "(Ff)^*(B)"] \arrow[d, "\tau_{A',B}"'] & h_{FA}(B) \arrow[d, "\tau_{A,B}"] \\
h_{A'}(GB) \arrow[r, "f^*(GB)"] & h_A(GB),
\end{tikzcd}$$

and

$$\begin{tikzcd}
h^B(FA) \arrow[r, "g_*(FA)"] \arrow[d, "\tau_{A,B}"'] & h^{B'}(FA) \arrow[d, "\tau_{A,B'}"] \\
h^{GB}(A) \arrow[r, "(Gg)_*(A)"] & h^{GB'}(A)
\end{tikzcd}$$

commute. The data (F, G, τ) is called an *adjunction*.

Notation 137.1. We shall write f^* for $f^*(X)$ and g_* for $g_*(Y)$, etc.

Remark 137.2. One can define adjoints for arbitrary categories as follows: Let $\mathcal{A}$ and $\mathcal{B}$ be categories. An *adjunction* consists of a pair of functors $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ and an isomorphism

$$(*) \qquad \mathrm{Hom}_{\mathcal{B}}(FA, B) \cong \mathrm{Hom}_{\mathcal{C}}(A, GB)$$

for each $A \in \mathrm{Ob}(\mathcal{A})$ and $B \in \mathrm{Ob}(\mathcal{B})$ that is natural in both variables. We call F the *left adjoint* of G and G the *right adjoint* of F. The morphism

$$FA \xrightarrow{f^\sharp} B \qquad \leftrightsquigarrow \qquad B \xrightarrow{f^\flat} GB$$

correspond under the isomorphism (*) and are called *adjuncts* or *transposes* of each other.

However, as we are assuming that they are categories are locally small, we can use Yoneda's Lemma that the isomorphism $\tau(A, B)$'s arising from the natural transformation τ above.

Examples 137.3. In the examples below, let R be a commutative ring and U the appropriate forgetful functor. Then (F, U) is an adjoint pair if

1. $F : \mathbf{Set} \to \mathbf{Group}$ by $X \mapsto$ the free group on X.
2. $F : \mathbf{Set} \to {}_R\mathcal{M}$ by $X \mapsto$ the free R-module on X (R need not be commutative).
3. $F : \mathbf{Set} \to \mathbf{Calg}_R$ by $X \mapsto$ the free commutative R-algebra on X, i.e., $R[X] =$ polynomials on X.
4. $F : {}_R\mathcal{M} \to \mathbf{Calg}_R$ by $M \mapsto S(M)$, the symmetric R-algebra on M.
5. $F : \mathbf{Group} \to \mathbf{Ab}$ by $G \mapsto G/[G, G]$.
6. $F : {}_R\mathcal{M} \to {}_A\mathcal{M}$ by $m \mapsto A \otimes_R M$ if $A \in \mathbf{Calg}_R$.

Examples 137.4. In the examples below, let R be a commutative ring, M a fixed R-module. Then (F, G) is an adjoint pair if

1. $G : \mathbf{Ring} \to \mathbf{Group}$ by $S \mapsto S^\times$ and $F : \mathbf{Group} \to \mathbf{Ring}$ by $G \mapsto \mathbb{Z}[G]$, the group ring on G.
2. $G : {}_R\mathcal{M} \to {}_R\mathcal{M}$ by $N \mapsto h_M(N)$ and $F : {}_R\mathcal{M} \to {}_R\mathcal{M}$ by $N \mapsto M \otimes_R N$.

   More generally, suppose that R and S are arbitrary rings and that $M \in \mathrm{Ob}({}_R\mathcal{M}_S)$. If $G : \mathcal{M}_S \to \mathbf{Ab}$ by $C \mapsto h_M(C)$ for $C \in \mathrm{Ob}(\mathcal{M}_S)$ and $F : \mathcal{M}_S \to \mathbf{Ab}$ by $A \mapsto A \otimes_R M$ for $A \in \mathrm{Ob}(\mathcal{M}_R)$, then $(\otimes_R M, h_M)$ is an adjoint pair, as
   $$\mathrm{Hom}_S(A \otimes_R M, C) \cong \mathrm{Hom}_R\big(A, \mathrm{Hom}_S(M, C)\big)$$
   (naturally) as abelian groups by 129.9.
3. $G : \mathbf{Ab} \to {}_R\mathcal{M}$ by $X \mapsto \mathrm{Hom}_{\mathbb{Z}}(R, X)$ and $F : {}_R\mathcal{M} \to \mathbf{Ab}$, the forgetful functor.
4. Let $\mathbf{Dom} = ((\mathit{domains}, \mathit{monomorphisms}))$. Then $F : \mathbf{Dom} \to \mathbf{Field}$ taking $R \mapsto qf(R)$ and G the forgetful functor.

Construction 137.5. Let $\mathcal{A}$ and $\mathcal{B}$ be two locally small categories with $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{C}$ functors. Suppose there exists a natural transformation $\eta : 1_{\mathcal{C}} \to GF$ satisfying the following: For all $A \in \mathrm{Ob}(A)$, $B \in \mathrm{Ob}(\mathcal{B})$, and morphism $f : A \to GB$, there exists a unique morphism $g : FA \to B$ satisfying

$$\begin{tikzcd} A \arrow[r, "\eta(A)"] \arrow[dr, "f"'] & GFA \arrow[d, "Gg"] \\ & GB \end{tikzcd}$$

commutes. We wish to show this is equivalent to (F, G) being an adjoint pair. The property of η in the above definition is called a *universal mapping property*. It says for each object $A \in \mathcal{A}$, the morphism $\eta(A)$ is a universal element for $\mathrm{Hom}_{\mathcal{A}}\big(A, G(\ )\big)$. We call such an η a *unit* for the pair (F, G).

Of course, we would also want this be equivalent to: Let $\mathcal{A}$ and $\mathcal{B}$ be two locally small categories with $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{C}$ functors. Suppose there exists a natural transformation $\varepsilon : FG \to 1_{\mathcal{B}}$ satisfy the following: For all $A \in \mathrm{Ob}(A)$, $B \in \mathrm{Ob}(\mathcal{B})$, and morphism $g : FA \to B$, there exists a unique morphism $g : A \to GB$ satisfying

$$\begin{tikzcd} & FGB \arrow[d, "\varepsilon(B)"] \\ FA \arrow[ur, "Ff"] \arrow[r, "g"'] & B \end{tikzcd}$$

commutes. In this case ε is called a *counit*.

Theorem 137.6. *Let $\mathcal{A}$ and $\mathcal{B}$ be locally small categories with $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{C}$ functors. Then (F, G) is an adjoint pair if and only if there exists a natural transformation $\eta : 1_{\mathcal{A}} \to GF$ a unit for the pair (F, G).*

PROOF. ($\Leftarrow$): Let $A \in \mathrm{Ob}(\mathcal{A})$ and $B \in \mathrm{Ob}(\mathcal{B})$. Define
$$\tau_{A,B} : \mathrm{Hom}_{\mathcal{B}}(FA, B) \to \mathrm{Hom}_{\mathcal{A}}(A, GB) \ \text{ by } \ \tau_{A,B}(g) = Gg \circ \eta(A)$$

and define

$$\sigma_{A,B} : \mathrm{Hom}_{\mathcal{A}}(A, GB) \to \mathrm{Hom}_{\mathcal{B}}(FA, B)$$

by letting $\sigma_{A,B}(f)$ be the unique morphism g that satisfies $f = Gg \circ \eta(A)$ given by the definition of the unit η. Uniqueness then implies that $\sigma_{A,B}\big(\eta_{A,B}(g)\big) = g$. The definition of $\eta_{A,B}$ and $\sigma_{A,B}$ then imply that $\eta_{A,B}\big(\sigma_{A,B}(f)\big) = f$. We leave the proof of naturality as an exercise.

($\Rightarrow$): Let

$$\tau_{A,B} : \mathrm{Hom}_{\mathcal{B}}(FA, B) \to \mathrm{Hom}_{\mathcal{A}}(A, GB)$$

be the natural isomorphism giving the adjunction of (F, G). Let $A \in \mathrm{Ob}(\mathcal{A})$ and $B \in \mathrm{Ob}(\mathcal{B})$ such that $B = FA$. Therefore,

$$(*) \qquad \mathrm{Hom}_{\mathcal{B}}(FA, FA) \cong \mathrm{Hom}_{\mathcal{A}}(A, GFA).$$

Let $\tau(A) \in \mathrm{Hom}_{\mathcal{A}}(A, GFA)$ be the morphism corresponding to 1_{FA}. If $f \in \mathrm{Hom}_{\mathcal{A}}(A, GB)$ is a morphism, let $g \in \mathrm{Hom}_{\mathcal{B}}(FA, B)$ corresponding to it via the isomorphism (*). Since τ is a natural transformation, we have a commutative diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{B}}(FA, FA) \arrow[r, "\tau_{A,FA}"] \arrow[d, "g_*(FA)"'] & \mathrm{Hom}_{\mathcal{B}}(A, GFA) \arrow[d, "g_*(FA)"] \\
\mathrm{Hom}_{\mathcal{C}}(FA, B) \arrow[r, "\tau_{A,B}"'] & \mathrm{Hom}_{\mathcal{C}}(A, GB).
\end{tikzcd}$$

As $\tau(A)$ corresponds to 1_{FA}, we have $1_{FA} \mapsto \tau(A) \mapsto g_*(FA)(1_{FA})$. In the composition, we see that $1_{FA} \mapsto g$. As g corresponds under the isomorphism to f, we have $f = g_*(FA)(1_{FA})$ as needed. If $h \in \mathrm{Hom}_{\mathcal{A}}(FA, B)$, also satisfies $f = h_*(FA)(1_{FA})$, then both g and h correspond to f, hence $g = h$. □

Corollary 137.7. *Let $G : \mathcal{B} \to \mathcal{A}$ be a functor of locally small categories. If G has a left adjoint, then it is unique up to a natural isomorphism.*

For locally small categories, one sometimes build functors from objects. One such is the following:

Theorem 137.8. (Pointwise Adjointness Theorem) *Let $G : \mathcal{B} \to \mathcal{A}$ be a functor of locally small categories. Suppose for each $A \in \mathrm{Ob}(\mathcal{A})$, there exists $F(A) \in \mathrm{Ob}(\mathcal{B})$ satisfying $\mathrm{Hom}_{\mathcal{B}}(FA, \ )$ is naturally equivalent to $\mathrm{Hom}_{\mathcal{A}}\big(A, G(\ )\big)$. Then the definition of F on objects can be extended on morphisms so that F becomes a functor and is a left adjoint of G.*

Proof. First define a function of objects $F : \mathrm{Ob}(\mathcal{A}) \times \mathrm{Ob}(\mathbf{Set}) \to \mathcal{B}$ to satisfy $\mathrm{Hom}_{\mathcal{A}}\big(A, G(X, B)\big) \cong \mathrm{Hom}_{\mathcal{B}}\big(F(A, X), B\big)$ for all $A \in \mathcal{A}$, $X \in \mathrm{Ob}(\mathbf{Set})$, $B \in \mathrm{Ob}(\mathcal{B})$. We want to extend this to a functor. Let $f : A \to A'$ in $\mathcal{A}$ and $g : X' \to X$ in $\mathbf{Set}$. Then for any $B \in \mathcal{B}$, we have a diagram

$$(*) \qquad \begin{tikzcd}
\mathrm{Hom}_{\mathcal{A}}\big(A', G(X', B)\big) \arrow[r, "\sim"] \arrow[d, "\mathrm{Hom}_{\mathcal{A}}(f, G(h, B))"'] & \mathrm{Hom}_{\mathcal{B}}\big(F(A', X'), B\big) \\
\mathrm{Hom}_{\mathcal{A}}\big(A, G(X, B)\big) \arrow[r, "\sim"] & \mathrm{Hom}_{\mathcal{B}}\big(F(A, X), B\big).
\end{tikzcd}$$

There exists a unique morphism $\varphi(f,g,B) : \mathrm{Hom}_{\mathcal{B}}\big(F(A',X'),B\big) \to \mathrm{Hom}_{\mathcal{B}}\big(F(A,X),B\big)$ so that placing this as the right hand vertical of (*) results in a commutative diagram. As both of the horizontals in the diagram are isomorphisms and $\mathrm{Hom}_{\mathcal{A}}\big(f, G(g,B)\big)$ are natural with respect to B, the map $\varphi(f,g,B)$ is also natural. It follows by Yoneda's Lemma 136.10 that there exists a unique morphism $F(f,g) : F(A,X) \to F(A',X')$ satisfying $\varphi(f,g,B) = \mathrm{Hom}_{\mathcal{B}}\big(F(f,g),B\big)$. If we have other morphisms $f' : A' \to A''$ in $\mathcal{C}$ and $g' : X'' \to X'$ in **Set**, respectively, we can put these diagrams together to see that $F(f,g) \circ F(f',g') = G(f \circ f', g' \circ g)$. We leave it as an exercise to show F preserves identities. $\square$

Corollary 137.9. *Let $G : \mathcal{B} \to \mathcal{A}$ be a functor of locally small categories. If G has a left adjoint, then it is unique up to a natural isomorphism.*

Proposition 137.10. *Let $G : \mathcal{B} \to \mathcal{A}$ be a functor of locally small categories. Then G has a left adjoint if and only if for each $A \in \mathcal{A}$, the functor $\mathrm{Hom}_{\mathcal{A}}\big(A, G(\ \ )\big) : \mathcal{B} \to \mathbf{Set}$ has a universal element.*

PROOF. If $b : A \to GB$ is a universal element, then $FA = B$ and $b : A \to G(B) = GF(A)$ is the component of A of the natural transformation $\eta : 1_{\mathcal{A}} \to GF$. $\square$

Remark 137.11. Of course, the dual statements for the last five results are valid for a functor $F : \mathcal{A} \to \mathcal{B}$ of locally small categories. We leave the statements to the reader.

We give further examples of adjoints. These are easier to check using the equivalent definitions of adjoint pairs.

Examples 137.12. 1. Let S be a set, the set of subsets of S is a poset via inclusion. This becomes a category $\mathbf{Sub}(S)$ as follows: If $S_0, S_1 \subset S$, then there exists precisely one morphism $S_0 \to S_1$ if and only if $S_0 \subset S_1$. Let $f : S \to T$ be a function of sets. If $T_0 \subset T$, we have the inverse image of T_0 via f. As $f^{-1}(T_0) \subset f^{-1}(T_1)$ if $T_0 \subset T_1$, $f^{-1} : \mathbf{Sub}(T) \to \mathbf{Sub}(T)$ is a functor. This functor has a left adjoint f_* defined by $f_*(S_0) = \{f(x) \mid x \in S_0\}$, called the *direct image functor*. So $f_*(S_0) \subset T_0$ if and only if $S_0 \subset f^{-1}(T_0)$. This says $y \in f_*(S)$ if and only if some element of $f^{-1}(y)$ lies in S_0.

2. Let $f : S \to T$ be a set map. Then f^{-1} has a right adjoint, the functor $f_!$ defined by $y \in f_!(S_0)$ if and only if $f^{-1}(\{y\}) \subset S_0$, i.e., every element of the inverse image of y is in S_0.

Exercises 137.13. 1. Prove the naturality of $\sigma_{A,B}$ in the proof of Theorem 137.6.

2. Prove Corollary 137.7.
3. Prove the naturality if Theorem 137.6
4. Fill in details of the proof of the Pointwise Adjointness Theorem 137.8.
5. Prove Corollary 137.9.

138. Limits

Let $\mathcal{J}$ be a small category and $\mathcal{C}$ a category. A functor $F : \mathcal{J} \to \mathcal{C}$ is called a *diagram* of *shape* $\mathcal{J}$. Let $C \in \mathrm{Ob}(\mathcal{C})$. Then C defines a functor $|C| : \mathcal{J} \to \mathcal{C}$ that sends every object of $\mathcal{J}$ to C and every morphism of $\mathcal{J}$ to the identity morphism 1_C. This functor is called the *constant functor*. Recall that we also denote the category $\mathbf{Funct}(\mathcal{J},\mathcal{C})$ by $\mathcal{C}^{\mathcal{J}}$. We shall use this notation. The constant functor define an embedding $\Delta : \mathcal{C} \to \mathcal{C}^{\mathcal{J}}$ called the *diagonal functor* that sends an object C to the constant functor $|C| : \mathcal{J} \to C$ and a morphism $f : C \to C'$ to the *constant natural transformation* which on each component is defined to be the morphism f.

Definition 138.1. Let $F : \mathcal{J} \to \mathcal{C}$ be a *diagram* of *shape* $\mathcal{J}$ and $C \in \mathrm{Ob}(\mathcal{C})$. We call a natural transformation $\lambda : C \to F$ a *cone over* F with *apex* C whose domain is the constant functor $|C|$. We write the components of λ by $\{\lambda_j : C \to F\}_{j\in J}$. Therefore, λ is a collection of morphisms $\{\lambda_j\}_{\mathcal{J}}$ satisfying for each morphism $f : i \to j$ in $\mathcal{J}$, a commutative diagram

$$\begin{tikzcd} & C \arrow[dl, "\lambda_i"'] \arrow[dr, "\lambda_j"] & \\ Fi \arrow[rr, "Ff"'] & & Fj. \end{tikzcd} \tag{138.2}$$

Proposition 138.3. *Let $F : \mathcal{J} \to \mathcal{C}$ be a diagram of shape $\mathcal{J}$ and $C \in \mathrm{Ob}(\mathcal{C})$. Then the following are equivalent:*

1. *$\lambda : C \to F$ is a cone over F with apex C.*
2. *$\lambda : \Delta(C) \to F$ is a natural transformation.*
3. *For each morphism $f : i \to j$ in $\mathcal{J}$, the diagram in equation (138.2) commutes.*
4. *(C, λ) is an object in the comma category $(\Delta \downarrow F)$.*

We leave the proof as an exercise.

This proposition allows us to define a category of cones as the natural map between $\Delta(C)$ and $\Delta(C')$ correspond to morphisms between C and C', since the diagonal morphism acts trivially on morphisms.

Definition 138.4. Let $\mathcal{J}$ be a small category and $\mathcal{C}$ a category and $F : \mathcal{J} \to \mathcal{C}$ a diagram of shape $\mathcal{J}$. Then the *category of cones to* F is defined to be the comma category $(\Delta \downarrow F)$. We then view morphisms of cones in this category, with objects the cones with an apex. In particular, a morphism of cones $\lambda : C \to F$ and $\lambda' : C' \to F$ is a morphism $f : C \to C'$ such that

$$\begin{tikzcd} & C \arrow[d, "f"] \arrow[ddl, "\lambda_i"'] \arrow[ddr, "\lambda_j"] & \\ & C' \arrow[dl, "\lambda'_i"] \arrow[dr, "\lambda'_j"'] & \\ Fi \arrow[rr, "Ff"'] & & Fj. \end{tikzcd}$$

commutes for all $i, j \in \mathcal{J}$.

We now turn to the the additional assumption that $\mathcal{C}$ is locally small, so that we can use Yoneda's Lemma. As J is small and $\mathcal{C}$ is locally small, $\mathcal{C}^{\mathcal{J}}$ is locally small, so the collection of such cones is a set. In particular, for any diagram $F : \mathcal{J} \to \mathcal{C}$ there exists a functor

$$\text{cone}(\ \ , F) : \mathcal{C}^{op} \to \mathbf{Set}$$

sending $C \in \mathcal{C}$ to the set of cones over F with apex C. This is a set as $\mathcal{J}$ is small. A *limit* of F is a representation for $\text{cone}(\ \ , F)$. By Yoneda's Lemma, a limit consists of an object $\lim F \in \mathcal{C}$ together with a universal cone $\lambda : \lim F \to F$, called the *limit cone* defining a natural isomorphism $h^{\lim F} \cong \text{cone}(\ \ , F)$. We often just write this as $\lim_{\mathcal{J}} F$ or just $\lim F$.

We relate limits to universal objects.

Theorem 138.5. *Let $\mathcal{J}$ be a small category, $\mathcal{C}$ a locally small category, and $F : \mathcal{J} \to \mathcal{C}$ a diagram of shape $\mathcal{J}$. Then the following are equivalent:*

(1) A limit for F exists.
(2) A terminal object $\text{cone}(D)$ *exists.*
(3) A universal object for the functor $\text{cone}(\ \ , F)$ *exists*

Proof. We leave the proof as an exercise. □

If $C = \lim F$ is a terminal object for a universal cone, then it comes with a natural transformation $\tau : \Delta(X) \to F$ which is universal with respect to all natural transformations $\nu : \Delta(X) \to F$ for all $X \in \text{Ob}(\mathcal{C})$.

We now look at the dual of the above. Let $\mathcal{J}$ be a small category and functor $F : \mathcal{J}^{op} \to C^{op}$ of shape $\mathcal{J}^{op}$ with $C \in \text{Ob}(\mathcal{C})$. A natural transformation $\lambda : F \to C$ is called a *cocone* over F with apex C whose domain is the constant functor $|C|$. Since $(\mathcal{C}^{\mathcal{J}})^{op} = (\mathcal{C}^{op})^{\mathcal{J}^{op}}$, we see that is equivalent to a diagram $F : \mathcal{J} \to \mathcal{C}$ a diagram of shape $\mathcal{J}$ with $\lambda : F \to C$ whose components satisfy

$$\begin{tikzcd} Fi \arrow[rr, "Ff"] \arrow[dr, "\lambda_i"'] & & Fj. \arrow[dl, "\lambda_j"] \\ & C & \end{tikzcd}$$

So we call this cocone a *cone* with *nadir* C.

We also have the category of cocones. It is the dual of the category of cones, so by the dual of Proposition 138.3, it is given by

Definition 138.6. Let $\mathcal{J}$ be a small category and $\mathcal{C}$ a category with $\mathcal{J} \to \mathcal{C}$. Then the *category of cocones to* F is defined to be the comma category $(F \downarrow \Delta)$. We then view morphisms of cocones in this category. In particular, analogous as before, a morphism of cocones $\lambda : C \to F$ and $\lambda' : C' \to F$ is a morphism $f : C' \to C$ such that

$$\begin{tikzcd} & C & \\ & C' \arrow[u, "f"] & \\ Fi \arrow[uur, "\lambda_i"] \arrow[ur, "\lambda'_i"'] \arrow[rr, "Ff"'] & & Fj. \arrow[uul, "\lambda_j"'] \arrow[ul, "\lambda'_j"] \end{tikzcd}$$

commutes for all $i, j \in \mathcal{J}$.

Now, in addition, assume that $\mathcal{C}$ locally small (hence $\mathcal{C}^{\mathcal{J}}$ is locally small), so the collection of cones is a set. For any diagram $F : \mathcal{J} \to \mathcal{C}$, there exists a functor $\mathrm{cone}(F, \ ) : \mathcal{C} \to \mathbf{Set}$ sending $C \in \mathcal{C}$ to the set of cones over F with nadir C. Therefore, Yoneda's Lemma 136.10 is applicable. A *colimit* of F is a representation for $\mathrm{cone}(F, \ ) : \mathcal{C} \to \mathbf{Set}$. It consists of an object $\operatorname{colim} F \in \mathrm{Ob}(\mathcal{C})$ together with a universal cone $\lambda : F \to \operatorname{colim} F$ called the *colimit cone* that defines the natural isomorphism $h_{\operatorname{colim} F} \cong \mathrm{cone}(F, \ )$. We often write $\operatorname{colim} F$ by $\operatorname{colim}_{\mathcal{J}} F$ when $\mathcal{J}$ is unclear.

Definition 138.7. We say that $\mathcal{J}$ is a *discrete* category, i.e., the only morphisms in $\mathcal{J}$ are the needed identity morphisms. A *product* is a limit of a diagram indexed by a discrete category with only identity morphisms. A diagram in $\mathcal{C}$ indexed by a discrete category $\mathcal{J}$ is simply a collection of objects $F_j \in \mathrm{Ob}(\mathcal{C})$ indexed by $j \in \mathcal{J}$. A cone over this diagram is just a family of morphisms $(\lambda_i : C \to F_i)_{\mathcal{J}}$, some $C \in \mathrm{Ob}(\mathcal{C})$, with no other relations. The limit is called the *product* of the diagram and denoted by $\prod_{\mathcal{J}} F_j$ with components of the limit cone $(\pi_i : \prod_{\mathcal{J}} F_j \mapsto F_i)_{i \in \mathcal{J}}$ called *projections*. The universal property of limit cones means for all $C \in \mathrm{Ob}(\mathcal{C})$, we have natural isomorphisms

$$\mathrm{Hom}_{\mathcal{C}}(C, \prod_{j \in \mathcal{J}} F_j) \xrightarrow[\cong]{(\pi_k)_*} \prod_{k \in \mathcal{J}} \mathrm{Hom}_{\mathcal{C}}(C, F_k) = \mathrm{cone}(C, F).$$

Examples 138.8. Let $\mathcal{J}$ be a small category, $\mathcal{C}$ a locally small category.

1. If $\mathcal{C}$ is a category and $f, g : A \to B$ in $\mathcal{C}$, we call f and g a *parallel pair*. Write such a parallel pair of morphisms as $f, g : A \rightrightarrows B$ in $\mathcal{C}$. Suppose that $\mathcal{J}$ is the category of two elements and two non-identity morphisms. Then the shape of $\mathcal{C}$ indexed by $\mathcal{J}$ is just parallel pair of morphisms $f, g : A \rightrightarrows B$ in $\mathcal{C}$. A cone over this diagram with apex C is a pair of morphisms $a : C \to A$ and $b : C \to B$ satisfying $fa = b$ and $ga = b$. By naturality, this implies that $fa = gb$, hence $b = fa$. It follows that the cone over the parallel pair $f, g : A \rightrightarrows B$ in $\mathcal{C}$ is represented by the single morphism $a : C \to A$ such that $fa = ga$. We call the universal morphism $h : E \to A$ with this property, the *equalizer* of f and g and denote the diagram of this by

$$\begin{tikzcd} E \arrow[r, "h"] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B. \end{tikzcd}$$

   The universal property says if $a : C \to A$ satisfies $fa = ga$, then there exists a unique morphism $k : C \to E$ such that

$$\begin{tikzcd} C \arrow[d, "k"'] \arrow[dr, "a"] & & \\ E \arrow[r, "h"] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \end{tikzcd}$$

   commutes. Moreover, if $h' : E' \to A$ is also an equalizer, then there exists a unique isomorphism $i : E \to E'$ satisfying $h = h'i$. We shall write $h = \mathrm{equalizer}(f, g)$

2. Suppose that $\mathcal{C}$ has a zero object 0. If $g : 0 \to A$, it is unique, called it the *zero map* 0. Then the equalizer
$$E \xrightarrow{h} A \underset{0}{\overset{f}{\rightrightarrows}} B.$$
is called the *kernel* of f.
3. Let $\mathcal{J}$ be a poset category consisting of two non-identity morphisms, which we can picture as $\bullet \to \bullet \leftarrow \bullet$, i.e., the two nonidentity morphisms have the same codomain. Let $\mathcal{C}$ be a category with shape $\mathcal{J}$. A limit of such a diagram is called a *pullback*. Write f and g for the morphisms in $\mathcal{C}$ defining the image of a diagram of this shape in $\mathcal{C}$. Then a cone with apex E is a triple of morphisms, one for each object of $\mathcal{J}$ so that both triangles in

(138.9)
$$\begin{tikzcd} E \arrow[r, "c"] \arrow[d, "b"'] \arrow[dr, "a"] & C \arrow[d, "g"] \\ B \arrow[r, "f"'] & A \end{tikzcd}$$

commute by the naturality in the definition of cones. The component a says that gc and fb have a common composite. We just will say this data of the cone over $B \xrightarrow{f} A \xleftarrow{g} C$ is a pair of morphisms $B \xleftarrow{b} E \xrightarrow{c} C$ defining the commutative square (138.9).

The *pullback* is the universal cone over $B \xrightarrow{f} A \xleftarrow{g} C$ satisfying $fh = gh$ and the following universal property: Given any commutative diagram (138.9), there exists a commutative diagram of the components

$$\begin{tikzcd} E \arrow[drr, "c"] \arrow[dr, "d"] \arrow[ddr, "b"'] & & \\ & P \arrow[r, "k"'] \arrow[d, "h"'] & C \arrow[d, "g"] \\ & B \arrow[r, "f"'] & A \end{tikzcd}$$

with $hf = kg$ and $d : E \to P$ unique. The pullback P is also called the *fiber product* and denoted by $B \times_A C$.

In the above, let $A \in \mathrm{Ob}(\mathcal{C})$ and $1 \in \mathrm{Ob}(\mathcal{C})$ with 1 representing a set functor $\mathcal{C} \to$ **Set** variable element parametrizing $1 \to C$ by Yoneda's Lemma 136.10. Then the pullback

$$\begin{tikzcd} P \arrow[r, "k"] \arrow[d, "h"'] & C \arrow[d, "g"] \\ 1 \arrow[r, "f"'] & A \end{tikzcd}$$

is called the *fiber* of $g : C \to A$ over f. For example, if $\mathcal{C}$ has a unique zero element and $f : 0 \to A$ in $\mathcal{C}$. Then $\ker g = \mathrm{equalizer}(f, g) = 0 \times_A C$.

4. Let $\mathcal{J}$ be a *directed poset*, i.e., a poset (viewed as a category) also satisfying the condition that for all $i, j \in \mathrm{Ob}(\mathcal{J})$, there exists a $k \in \mathrm{Ob}(\mathcal{J})$ such that $i \leq k$ and $j \leq k$. Then the limit of the diagram $F : \mathcal{J} \to \mathcal{C}$ is called the *inverse* or *projective limit* of F and denoted by $\varprojlim F_j$.

Of course, we have the dual of the notions of a diagram $\mathcal{J} \to \mathcal{C} \to F$ above with apex replaced with *nadir* and terminal object replaced by an initial object: $C(F, \ ) : \mathcal{C}^{op} \to \mathbf{Set}$ sending cone$(F, \ ) \to \mathbf{Set}$ sending C to the set of cones with nadir C, colimit, coproduct coequalizer, cokernel, and *direct limit* (the dual of inverse limit). We leave the accompanied diagrams to the reader.

Definition 138.10. Let $\mathcal{J}$ be a small category, $\mathcal{C}$ a locally small category, and $\{G : \mathcal{J} \to \mathcal{C}\}$ a collection of diagrams. Suppose that $F : \mathcal{C} \to \mathcal{D}$ is a functor of locally small categories. We say that

1. F *preserves limits* of the collection $\{G : \mathcal{J} \to \mathcal{C}\}$ if for any limit cone over $G : \mathcal{J} \to \mathcal{C}$ in the collection, the image of this cone defines a limit cone over $FG : \mathcal{J} \to \mathcal{D}$.
2. F *reflects limits* of the collection $\{G : \mathcal{J} \to \mathcal{C}\}$ if whenever a diagram a $G : \mathcal{J} \to \mathcal{C}$ in the collection has image $FG : \mathcal{J} \to \mathcal{D}$ defines a limit cone over $G : \mathcal{J} \to \mathcal{C}$.
3. F *creates limits* of the collection $\{G : \mathcal{J} \to \mathcal{C}\}$ if whenever $FG : \mathcal{J} \to \mathcal{D}$ has a limit, there exists a a limit cone over FG that can be lifted to a limit cone over G and, moreover, F reflects the limits in the collection $\{G : \mathcal{J} \to \mathcal{C}\}$.

Of course, we have the dual notions for colimits. In particular, a fully faithful functor preserves, reflects, and creates isomorphisms (whose proof we leave as an exercise).

Proposition 138.11. *Let $\mathcal{C}$ and $\mathcal{D}$ be locally small categories and $F : \mathcal{C} \to \mathcal{D}$ a functor. If F creates limits for a collection of diagrams in $\mathcal{C}$ and $\mathcal{D}$ has limits of those diagrams, then $\mathcal{C}$ admits those limits and F preserves them.*

PROOF. Let $G : \mathcal{J} \to \mathcal{C}$ be a diagram in the collection. Then, by hypothesis, there exists a limiting cone $\mu : D \to FG$ in $\mathcal{D}$. As F creates these limits, there must be a limit cone $\lambda : C \to G$ in $\mathcal{C}$ whose image under F is isomorphic to μ by Proposition 138.5. Therefore, $\mathcal{C}$ admits this collection of limits. To see that it preserves them, suppose that $\lambda' : C' \to G$ is another limit cone. Then by Proposition 138.5, they are isomorphic in $\mathcal{C}$ and composing the isomorphisms, shows $\lambda' : FC' \to FG$ is isomorphic to the limit cone $\mu : D \to FG$. Therefore, $F\lambda' : FC' \to FG$ is isomorphic to the limit cone $\mu : D \to FG$. Consequently, $F\lambda' : FC' \to FG$ is a limit cone and F preserves these limits. $\square$

Definition 138.12. Let $\mathcal{C}$ be a locally small category. We say that $\mathcal{C}$ is *complete* if every diagram $F : \mathcal{J} \to \mathcal{C}$ (with $\mathcal{J}$ small) has a limit. We say it is *cocomplete* if every such diagram has a colimit.

Construction 138.13. Let $F : \mathcal{J} \to \mathbf{Set}$. Then a limit is a representation $\mathrm{Hom}_{\mathbf{Set}}(S, \lim F) \cong \mathrm{cone}(S, F)$ of the functor that sends a set S to a set of cones over F with apex S. Let $1 \in \mathrm{Ob}(\mathbf{Set})$ be a singleton set representing the identity functor $\mathbf{Set} \to \mathbf{Set}$. Then $\lim F \cong \mathrm{Hom}_{\mathbf{Set}}(1, \lim F) = \mathrm{cone}(1, F)$, i.e., we may view $\lim F$ as $\mathrm{cone}(1, F)$. The components of this limit cone $\mathrm{cone}(1, F)$ are then functions $\lambda_j : \lim F \to Fj$ indexed by objects in $j \in \mathrm{Ob}(\mathcal{J})$ via taking a cone μ with summit 1 and the component μ_j to an element in the set Fj defining the natural transformation $\mu : 1 \to F$ via $\mu_j : 1 \to Fj$

We leave as an exercise the following result.

Theorem 138.14. *Let $\mathcal{C}$ be a locally small category. Then*

1. *All representable functors h_X preserve all limits that exist in $\mathcal{C}$.*
2. *The covariant Yoneda embedding $h : \mathcal{C} \to \mathbf{Set}^{\mathcal{C}^{op}}$ both reserves and reflects limits, i.e., a cone over a diagram in $\mathcal{C}$ is a limit cone if and only if its image defines a limit cone in $\mathbf{Set}^{\mathcal{C}^{op}}$.*

The existence of limits is reflected by the notion of adjoints. In particular, the following is true.

Theorem 138.15. *Let $\mathcal{C}$ be a locally small category and $\mathcal{J}$ a set. Then*

(1) All diagrams indexed by J have a have a limit if and only if the constant diagram functor $\Delta : \mathcal{C} \to \mathcal{C}^{\mathcal{J}}$ has a right adjoint.
(2) All diagrams indexed by J have a have a colimit if and only if the constant diagram functor $\Delta : \mathcal{C} \to \mathcal{C}^{\mathcal{J}}$ has a left adjoint.

Proof. It suffices to prove (1) by duality and the universal property of limits and of colimits, respectively. Let $C \in \mathcal{C}$ and $F \in \mathcal{C}^{\mathcal{J}}$. As $\mathcal{J}$ is small $\mathrm{Hom}_{\mathcal{C}^{\mathcal{J}}}$ is the set $(\mathrm{Nat}\Delta, F)$. But this is the set of cones over F with apex C. Therefore, there exists $\lim F \in \mathrm{Ob}(\mathcal{C})$ and an isomorphism $\mathrm{Hom}_{\mathcal{C}}(\Delta(A), F) \cong \mathrm{Hom}_{\mathcal{C}}(A, \lim F)$ with the isomorphism natural in $C \in \mathcal{C}$ if and only if the limit exists by Theorem 137.8 and Theorem 138.14. □

The proof is essentially that given in

Theorem 138.16. *The category* **Set** *is complete.*

Proof. To show that **Set** is complete, we must show that given the components defined in the Construction 138.13 defines a cone, i.e., given $F : \mathcal{J} \to \mathbf{Set}$ and $: i \to j$ in $\mathrm{Ob}(\mathcal{J})$, then the diagram

$$\begin{tikzcd} & \lim F \arrow[dl, "\lambda_i"'] \arrow[dr, "\lambda_j"] & \\ Fi \arrow[rr, "Ff"'] & & Fj. \end{tikzcd}$$

commutes. In the notation of Construction 138.13, as μ is the cone with apex 1, we have

$$Ff\big(\lambda_i(\mu)\big) = Ff(\lambda_i) = \mu_j = \lambda_k(\mu).$$

where $f : i \to j$ in $\mathrm{Ob}(\mathcal{J})$. Therefore the diagram commutes. It follows that $\{\lambda_j\}_J \in \mathcal{J}$ defines a cone over F.

We must show that $\lambda : \lim F \to F$ is the universal cone. Let $S \in \mathbf{Set}$ and $\varepsilon : S \to F$ be a cone with apex S. We must show that there exists a unique function $g : S \to \lim F$

such that

(*)
$$\begin{tikzcd} & S \arrow[d, "g"] \arrow[ddl, "\varepsilon_i"'] \arrow[ddr, "\varepsilon_j"] & \\ & \lim F \arrow[dl, "\lambda_i"] \arrow[dr, "\lambda_j"'] & \\ Fi \arrow[rr, "Ff"'] & & Fj. \end{tikzcd}$$

Viewing $s \in S$ as the variable element parametrized by the function $1 \to S$ ($1 = \star$ for some fixed element $\star$), we see that there is a cone $\varepsilon s : 1 \to F$ defined by restricting the cone ε along the variable element s. Define $g(s) \in \lim F = \mathrm{cone}(1, F)$ to be the cone εs. By the definition of the components of the limit cone λ, we have

$$\lambda_j\big((g(x)\big) = \lambda_j(\varepsilon s) = (\varepsilon s)_j = \varepsilon_j(s).$$

Therefore, (*) commutes. Moreover, the definition of $g(s) = \varepsilon x$ is necessary, so the map is unique. The result that **Set** is complete follows. □

Examples 138.17. Let $\mathcal{J}$ be a small. In this notation, Examples 138.8(1),(2) follow in the following way (whose proofs we leave as exercises):

1. The product of sets (A_j) indexed by the objects in a small category $\mathcal{J}$ is the set of cones over this collection of sets with summit 1 with $\mathcal{J}$ discrete. The object in the category of **Set** is the usual product of sets in **Set**.
2. Given a parallel pair of morphisms $x \rightrightarrows y$ in $\mathcal{J}$ their equalizer is the set of maps $x : 1 \to X$ satisfying $f(x) = g(x)$ for all $x \in X$. So the equalizer of f and g is $\{x \in X \mid f(x) = g(x)\}$. If $F : \mathcal{J} \to$ **Sets** is a functor, we have an equalizer diagram

$$\lim_{\mathcal{J}} F \hookrightarrow \prod_{j \in \mathrm{Ob}(\mathcal{J})} F_j \underset{h}{\overset{g}{\rightrightarrows}} \prod_{f \in \mathrm{morph}(\mathcal{J})} F(\mathrm{cod}\, f).$$

Theorem 138.18. *$\mathcal{C}$ admits all limits of diagrams indexed by a small category $\mathcal{J}$ if and only if the constant functor $\Delta : \mathcal{C} \to \mathcal{C}^{\mathcal{J}}$ admits a right adjoint. If $\mathcal{J} \to$ **Set** is a diagram, then there exists an equalizer diagram*

(138.19)
$$\lim_{\mathcal{J}} F \hookrightarrow \prod_{j \in \mathrm{Ob}(\mathcal{J})} F_j \underset{h}{\overset{g}{\rightrightarrows}} \prod_{f \in \mathrm{morph},(\mathcal{J})} F(\mathrm{cod}\, f).$$

In particular, any limit in **Set** *can be expressed as an equalizer as in equation (138.19).*

Proof. In Construction 138.13, let $F : \mathcal{J} \to$ **Set** be a diagram corresponding to cones with apex 1 with components $\{\lambda_j\}_{\mathcal{J}}$ with each cone λ_j corresponding to the set Fj. So for each $f : i \to j$ in $\mathcal{J}$, we have

$$\begin{tikzcd} & 1 \arrow[dl, "\lambda_{\mathrm{dom} f}"'] \arrow[dr, "\mathrm{cod} f"] & \\ F(\mathrm{dom} f) \arrow[rr, "Ff"'] & & F(\mathrm{cod} f) \end{tikzcd}$$

commutes, i.e., $Ff(\lambda_{\mathrm{dom}f}) = \lambda_{\mathrm{cod}f}$. This describes the domain of the equalizer in equation (138.19) for the product $\prod_{j\in \mathrm{Ob}(\mathcal{J})} Fj$ and the conditions about the codomain and the parallel pair in $\prod_{f\in\mathcal{J}} F(\mathrm{cod}f)$.

So we need to define the parallel pair g and h. View $\{\lambda_j\}_{\mathcal{J}}$ as the components $\prod_{\mathcal{J}} F_j$, i.e., of a cone with apex 1 over F with c maps to $(\lambda_{\mathrm{cod}f})_{f\in\mathcal{J}} \in \prod_{f\in\mathrm{morph}\mathcal{J}} F(f_{\mathrm{cod}f})$. CHECK The equalizer is the subset of $\prod_{j\in\mathrm{Ob}(\mathcal{J})} Fj$ whose elements are equal to the set of $(\lambda_j : 1 \to Fj)_j \in \mathrm{Ob}(j)$ that satisfy the compatibility conditions of the cones. Since this, together with the explicit Examples 138.17 shows that $lim_{\mathcal{J}} F$ is the equalizer of g and h. $\square$

Theorem 138.20. *Let U : **Group** $:\to$ **Set** be the forgetful functor. Then U creates limits.*

Proof. Let $G : \mathcal{J} \to$ **Group** be a diagram with $\mathcal{J}$ small. As **Set** is complete, we can define the product of two cones σ, τ in the cone $\mathrm{cone}(1, UG)$ via the components $(\alpha\beta)_j = \alpha_j\beta_j$ and $(\alpha^{-1})_j = \alpha_j^{-1}$ in the group H_j. This defines a group structure on $\mathrm{cone}(1, UG)$. Moreover, each component of $\tau : \mathrm{cone}(1, UG) \to UG$ is a group homomorphism. Conversely, if each component of $\lambda : \mathrm{cone}(1, UG) \to UG$, $\alpha \to \alpha_j$, is a group homomorphism, then the product $\sigma, \tau \in \mathrm{cone}(1, UG)$ must be given by this formula.

Let H is a group and $\lambda : H \to G$ a cone in **Group** with components $\lambda_j : H \to G_j$ a group homomorphism. If $\lim UH = \mathrm{cone}(1, UH)$ (identifying the isomorphism as an identity as before), then $U\lambda : UH \to UG$ is a cone in **Set**. By universality, $U\lambda = (U\tau))h$ for a unique function $f : UH \to \mathrm{cone}(1, UH)$. If $h_1, h_2 \in H$, then we have

$$\big(f(h_1h_2)\big) = \lambda_j(h_1h_2) = (\lambda_j h_1)(\lambda_j h_2) = \big((\lambda_j h_1)(\lambda_j h_2)\big)_j,$$

as λ is a homomorphism of groups. Therefore, f is also a homomorphism of groups and $\lim UH$ is a limit in **Group**. $\square$

As Set is complete, and using the analogue of of Theorem 138.20 for the categories **Ring**, **Ab**, ${}_R\mathcal{M}$, we have

Corollary 138.21. **Ring**, **Ab**, ${}_R\mathcal{M}$ *are complete.*

Theorem 138.22. *The category* **Set** *is cocomplete.*

We leave the proof that **Set** is cocomplete as an exercise. The proof is essentially formalizing the proof that we gave for modules ${}_R\mathcal{M}$ in Proposition 130.2 but using quotients arising from epimorphisms of sets and coproducts as disjoint unions of sets which gives us the coequalizer version of Theorem 138.18 for the explicit case, i.e,

As the analogue of Theorem 138.20 holds for colimits, we have

Corollary 138.23. **Ring**, **Ab**, ${}_R\mathcal{M}$ *are cocomplete.*

The analogue of Theorem 138.18 a

Theorem 138.24. *$\mathcal{C}$ admits all colimits of diagrams indexed by a small category $\mathcal{J}$ if and only if the constant functor $\Delta : \mathcal{C} \to \mathcal{C}^{\mathcal{J}}$ admits a right adjoint. If $\mathcal{J} \to$* **Set** *is a* *diagram, then there exists an equalizer diagram*

$$\coprod_{f \in morph(\mathcal{J})} F(\mathrm{dom} f) \overset{g}{\underset{h}{\rightrightarrows}} \coprod_{\mathcal{J}} F_j \twoheadrightarrow \mathrm{colim}_{\mathcal{J}} F \tag{138.25}$$

In particular, any colimit in **Set** *can be expressed as a coequalizer as in equation (138.25).*

also holds.

We further relate limits and adjoints in the following:

Theorem 138.26. *Let $\mathcal{C}$ and $\mathcal{D}$ be locally small categories. Suppose that $F : \mathcal{C} \to \mathcal{D}$ be the left adjoint to $G : \mathcal{B} \to \mathcal{C}$. Then G preserves limits and F preserves colimits.*

Proof. We give a sketch of the full proof. Let $\mathcal{J} \to \mathcal{C} \to \mathcal{D}$ be the functor by cone($, D) : \mathcal{C}^{op} \to$ **Set** with apex $C \in \mathcal{C}$ with $\mathcal{C}$ locally small) with limit given by a cone $V \to D$. Then we have equivalences of contravariant functors whose proof we leave as an exercise.

$$\mathrm{cone}(\ \ , GD) \cong \mathrm{cone}(F(\ \ ), D \cong \mathrm{Hom}(F(\ \ ), V) \cong \mathrm{Hom}((\ \ ), GV).$$

It follows that G preserves limits. We leave the proof that F preserves colimits as an exercise. □

Example 138.27. Let X be a topological space. Let $\mathcal{O}$ be the poset of open subsets of X ordered by inclusion indexed by $\mathcal{J}$. So if $U \subset U'$, there exists a unique morphism $\iota_{U,U'}$ in $\mathcal{O}$ giving the inclusion $U \hookrightarrow U'$. A family $(U_i)_{\mathcal{J}}$ of open subsets of an open set U in X is called a *cover* of U if the totality of diagrams comprising the cover of the set U and the associated inclusion maps $\iota_{U_j,U} : U_j \to U$ of pairwise intersections $U_{ij} = U_i \cap U_j$ (ordered) has U as a colimit. Suppose this is true. As before, a *presheaf* of sets on X is a functor $F : \mathcal{O}(X)^{op} \to$ **Set**. If $U \subset U'$, let $\rho_{U,U'} = F(\iota_{U.U'})$. A presheaf $F : \mathcal{O}(X)^{op} \to$ **Set**, is called a *sheaf* if it preserves these colimits, sending them to limits in **Sets**, i.e., for any open cover $(U_i)_{\mathcal{J}}$ of an open set U in X, we have an equalizer diagram

$$F(U) \overset{\rho}{\rightarrowtail} \prod_{i \in \mathcal{J}} F(U_i) \overset{\rho_1}{\underset{\rho_2}{\rightrightarrows}} \prod_{i,j \in \mathcal{J}} F(U_i \cap U_j)$$

in **Set**, where $\rho = \prod_{i \in \mathcal{J}} \rho_{U,U_i}$, $\rho_1 = \prod_{i \in \mathcal{J}} \rho_{U_i,U_{ij}}$, and $\rho_2 = \prod_{i \in \mathcal{J}} \rho_{U_i,U_{ji}}$. If **Set** is replaced by **Group**, respectively, **ComRing**, ${}_R\mathcal{M}$, etc., then F is called a *sheaf* of *groups* respectively, of *rings, modules*, etc.

If X is a topological space and F is a sheaf of rings on X which is written by $\mathcal{O}_X$ (different from the poset above), then the pair $(X, \mathcal{O}_X)$ is called a *ringed space*. This includes manifolds $(X, \mathcal{O}_X)$ where, e.g., $\mathcal{O}(U)$ is the real-valued functions $U \to \mathbb{R}$ and U is an open euclidean subspace of X. In algebra, if R is a a commutative ring, $X = \mathrm{Spec}(R)$ has a sheaf of rings determined by $\mathcal{O}\big(D(f)\big) = R_f$, the localization at $\{\mathfrak{p} \in X \mid f \notin D(f)\}$ for all $f \in R$. The ringed space $(X, \mathcal{O}_X)$ is called an *affine scheme*. (Cf. Exercise 92.31(11).) To get a category of ringed spaces, we must define morphisms. Let $f : X \to Y$ be a continuous map of topological spaces. Let F be a sheaf on X. For all $V \subset Y$ open, let $V \mapsto F\big(f^{-1}(V)\big)$ defines a sheaf f_*F on X. Then a morphism of ringed spaces is $(f, f^{\#}) : (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ with $f : X \to Y$ a continuous map and $f^{\#} : \mathcal{O}_Y \to f_*\mathcal{O}_X$.

Now for modules and rings, we can also use limits and colimits. In particular, if $x \in U$, then $F_x := \varinjlim_{x\in U} F(U)$ is defined for any sheaf F and called the *stalk* of F at x. For the sheaf $\mathcal{O}$, we write $\mathcal{O}_{X,x} := (\mathcal{O}_X)_x$. Let $(f, f^\#) : (X, \mathcal{O}_X) \to (y, \mathcal{O}_Y)$ be a morphism of ringed spaces and G a sheaf on Y. Then $\varinjlim_{f(U)\subset V} G(V)$ is a presheaf on X and the sheafification lies in the reflective sheaf category $\mathbf{Sheaf}_X$. (Cf. Exercise 138.29(12).) It is written as $f^{-1}(G)$. Then the stalk $(f^{-1}G)_x = G_{f(x)} := \varinjlim_{f(x)\in V} G(V)$ for every $x \in X$. So $f^{\#-1}(\mathfrak{m}_x) = \mathfrak{m}_{f(x)}$ where $\mathfrak{m}_x$ is the germ at x. Therefore, the induced map $\mathcal{O}_{Y,f(x)} := (\mathcal{O}_X)_x$ is a local ring homomorphism for all $x \in X$. We call these ringed spaces satisfying this additional property *local ring spaces*. It forms a full subcategory of the category of ring spaces.

Examples 138.28. 1. ADD some examples

Exercises 138.29. 1. Prove Theorem 138.3 (Think of cones as natural tranformations, then they are just morphisms in $\mathcal{C}^{\mathcal{J}}$ with domain (or codomain) a constant functor.)

2. Prove Theorem 138.5.
3. Let $\mathcal{C}$ be a locally small category. Call a monomorphism $f : A \to B$ in $\mathcal{C}$ *regular* if it is an equalizer. Prove that every monomorphism in **Set** is regular. Show if $f : A \to B$ is a regular equalizer and also an epimorphism, then f is an isomorphism.
4. Let $f, g : A \to B$ in ${}_R\mathcal{M}$. Show the the equalizer
$$E \xrightarrow{h} A \underset{g}{\overset{f}{\rightrightarrows}} B$$
is $\ker(f - g)$.
5. Prove Theorem 138.14.
6. Show Example 138.17 is valid.
7. Prove Theorem 138.26.
8. Let $F : \mathcal{C} \to \mathcal{D}$ be a full and faithful functor. Prove that F reflects any limits and colimits that are present in its codomains.
9. Any equivalence of categories preserves, reflects, and creates any limits and colimits that are present in its domain or codomain.
10. Prove he category **Set** is cocomplete.
11. Let $\mathcal{J}$ be a small category and $\mathcal{C}$ a locally small category. Prove that there exists a natural isomorphism $h_X(\lim_{\mathcal{J}} F) \cong \lim_{\mathcal{J}} h_X(F)$.
12. $\mathcal{C}$ be a locally small category. A full subcategory $\mathcal{D} \hookrightarrow \mathcal{C}$ is called *reflective*: if the inclusion functor admits a left adjoint called a *reflector* or *localization*. In Example 138.27, show that sheaves on X define a reflexive category $\mathbf{Sheaf}_X$ of the category $\mathbf{Presheaf}_X$. The reflector is called *sheafication*.
13. Use Remark 137.2 for the general definition of adjoints to prove the following: Let $\mathcal{J}$ be a small category, $\mathcal{C}$, $\mathcal{D}$ arbitrary categories. If $G : \mathcal{C} \to \mathcal{D}$ is a functor having a left adjoint and $T : \mathcal{J} \to \mathcal{A}$ has a limiting cone $\tau : C \to T$ in $\mathcal{C}$, then GT has the limiting cone $G\tau : GA \to GT$ in $\mathcal{D}$.

139. Additive and Abelian Categories

In this section, we generalize material that we did in our studies of modules and homological algebra. Many of the results have been shown in the particular case of modules, so even more results are just stated with proofs left to the reader (who can look at the the corresponding proofs done before). In addition, proofs about new elementary statements also left as exercises.

Definition 139.1. A category $\mathcal{C}$ is called a *preadditve* category if for all objects $A, B, C \in \mathrm{Ob}(\mathcal{C})$, we have

1. $\mathrm{Hom}_{\mathcal{C}}(A, B)$ is a abelian group with the identity morphism called the 0-morphism and written 0.
2. $\mathrm{Hom}_{\mathcal{C}}(A, B) \times \mathrm{Hom}_{\mathcal{C}}(B, C) \to \mathrm{Hom}_{\mathcal{C}}(A, C)$ satisfies the distributive laws, i.e., is *biadditive.*

Remarks 139.2. Let $\mathcal{C}$ be a preadditive category.

1. $\mathcal{C}$ is locally small.
2. The 0-morphism has the usual properties, i.e.,

$$\begin{tikzcd} & B \arrow[dr, "g"] & \\ A \arrow[ur, "0"] \arrow[rr, "0"'] & & C \end{tikzcd}$$

   is a commutative diagram of morphisms in $\mathcal{C}$.
3. **Group** is not preadditive.
4. If $\mathrm{Ob}(\mathcal{C}) = \{A\}$, then $\mathcal{C}$ is the "same thing" as a ring, with 1_A the one.
5. If $A \in \mathrm{Ob}(\mathcal{C})$ satisfies $|\mathrm{Hom}_{\mathcal{C}}(A, A)| = 1$, then A has a zero object, which is unique up to isomorphism.
6. ${}_R\mathcal{M}$ is a preadditive category with 0 the same as the zero object in **Ab**.

Definition 139.3. A preadditive category $\mathcal{C}$ is called an *additive* category if it has a unique zero object, written 0, and a coproduct.
[Of course, having a coproduct means that it has finite coproducts.]

We leave the proofs of the next four results as exercises.

Lemma 139.4. *Let $\mathcal{C}$ be an additive category, $A_1, \dots, A_n \in \mathrm{Ob}(\mathcal{C})$. Then A is a coproduct of $A_1, \dots, A_n$ if and only if there exist morphisms $A_j \xrightarrow{\iota_j} A$ and $A \xrightarrow{\pi_j} A_j$ in $\mathcal{C}$ for $j = 1, \dots n$ satisfying for all $j, k = 1, \dots, n$*

$$\pi_k \iota_j = \delta_{kj} 1_{A_j} \ \text{ and } \ \sum_j \iota_j \pi_j = 1_A.$$

Proposition 139.5. *Let $\mathcal{C}$ be an additive category. Then $\mathcal{C}$ has a product. Moreover, in an additive category, we can identify the product and the coproduct.*

Of course, an additive category may be neither cocomplete nor complete as it may not have infinite products or coproducts.

Examples 139.6. 1. If $\mathcal{C}$ is a preadditive category with a unique 0 object and a product, then $\mathcal{C}$ is additive. In particular, $\mathcal{C}$ is additive if and only if $\mathcal{C}^{op}$ is additive.

2. If $\mathcal{C}$ is additive category, then a morphism $A \xrightarrow{f} B$ in $\mathcal{C}$ is a monomorphism if and only if whenever $C \xrightarrow{g} B$ is a morphism satisfying $fg = 0$, then $g = 0$. The analogous result holds for epimorphisms.

Proposition 139.7. *If $\mathcal{C}$ is additive and we have morphisms*

$$A \xrightarrow{f} B \xrightarrow{h} D \text{ and } A \xrightarrow{g} C \xrightarrow{j} D$$

in $\mathcal{C}$, then there exists unique morphisms

$$A \xrightarrow{(f,g)} B \coprod C \xrightarrow{h \coprod j} D$$

satisfying

(*i*) $(f, g) = \iota_B f + \iota_C g$.
(*ii*) $h \coprod j = j\pi_B + j\pi_C$.

We give an interesting special case of Proposition 139.7. Let $\Delta = (1_A, 1_A) : A \to A \coprod A$ be the *diagonal morphism.* If $A = B = C$ and $f = g = 1_A$, we have the following:

Proposition 139.8. *Let $\mathcal{C}$ be an additive category. Then*

$$\begin{tikzcd}
A \arrow[rr, "h+j"] \arrow[dr, "\Delta"'] & & D \arrow[dl, "h \coprod j"] \\
& A \coprod A &
\end{tikzcd}$$

is a commutative diagram of morphisms in $\mathcal{C}$. In particular, addition in $\mathrm{Hom}_{\mathcal{C}}(A, D)$ *is determined by $\mathcal{C}$.*

Definition 139.9. A functor $F : \mathcal{A} \to \mathcal{B}$ of additive categories is called *additive* if for all morphisms $f, g : A \to B$ in $\mathcal{A}$, we have $F(f + g) = Ff + Fg$. In particular, such a functor F must take the 0-morphism to the 0-morphism.

We leave as an exercise a proof of the following (which of course has an obvious analog for contravariant functors):

Proposition 139.10. *Let $F : \mathcal{A} \to \mathcal{B}$ be a functor of additive categories. Then F is an additive functor if and only if F preserves coproducts (equivalently, products).*

Definition 139.11. Let $\mathcal{C}$ be a locally small category. Then

1. If a morphism $\alpha : X \to A$ in $\mathcal{C}$ is a monomorphism, then (X, α) is called a *subobject* of A. Two subobjects (X, α) and (X', α') of A are called *equivalent* if there exists a

commutative diagram

$$\begin{tikzcd} X \arrow[dd, "\cong"'] \arrow[dr, "\alpha"] & \\ & A \\ X' \arrow[ur, "\alpha'"'] & \end{tikzcd}$$

2. If a morphism $\beta : A \to X$ in $\mathcal{C}$ is an epimorphism, then (X, β) is called a *quotient object* of A. Two quotient objects (X, β) and (X', β') of A are called *equivalent* if they are equivalent subobjects in $\mathcal{C}^{op}$.

So we see if $\mathcal{C}$ is an additive category (or has a zero object and 0 morphism) and $\alpha : X \to Y$ is a morphism in $\mathcal{C}$, then, as before, $(\ker \alpha, \gamma)$ is the kernel of the equalizer $(\alpha, 0)$

$$\ker \alpha \xrightarrow{\gamma} X \underset{0}{\overset{\alpha}{\rightrightarrows}} Y$$

unique up to a unique isomorphism. So it is the largest subobject of (X, α) of X mapping to zero by α (and unique up to a unique isomorphism). We say (X, α) *kills* α. We write $\ker \alpha$ for it (if it exists) for this subobject of (X, α) or $\ker \alpha \to X$ in $\mathcal{C}$. Similarly, the cokernel is the coequalizer$(\alpha, 0)$. It is unique up to a unique isomorphism and written $\operatorname{coker} \alpha$ and is a quotient object (Y, α) of Y (if it exists). It is the smallest quotient object (Y, α) of Y *killed* by Y (unique up to a unique isomorphism). and also written as $Y \to \operatorname{coker} \alpha$.

Definition 139.12. Let $\mathcal{C}$ be an additive category and $\alpha : X \to Y$ a morphism in $\mathcal{C}$.

1. If $\ker \alpha$ exists, then the *coimage* of α, if it exists, is $\operatorname{coker}(\ker \alpha)$ and written $\operatorname{coim} \alpha$. So it is a quotient object of X.
2. If $\operatorname{coker} \alpha$ exists, then the *image* of α, if it exists, is $\ker(\operatorname{coker} \alpha)$ and written $\operatorname{im} \alpha$. So it is a subobject of X.

Construction 139.13. Let $\mathcal{C}$ be an additive category and $\alpha : X \to Y$ in $\mathcal{C}$. Suppose that $\ker \alpha, \operatorname{coker} \alpha, \operatorname{coim} \alpha$, and $\operatorname{im} \alpha$ all exist. Then we have a commutative diagram

$$\begin{tikzcd} \ker \alpha \arrow[r] & X \arrow[r, "\alpha"] \arrow[d] & Y \arrow[r] & \operatorname{coker} \alpha \\ & \operatorname{coim} \alpha \arrow[r, dashed, "\alpha'"] \arrow[ur, dashed, "\delta"] & \operatorname{im} \alpha \arrow[u] & \end{tikzcd}$$

with the dotted arrows to be defined. We first show the existence of the morphism δ. Since $\operatorname{coim} \alpha$ is the largest quotient of X that kills $\ker \alpha$ and $\alpha \ker \alpha = 0$. We see that δ exists and we have a commutative diagram

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d] & Y \\ \operatorname{coim} \alpha. \arrow[ur, "\delta"] & \end{tikzcd}$$

Next we show that α' exists. By the commutativity and definition, we have $\operatorname{coker}\alpha \circ \delta \circ \operatorname{coim}\alpha = \operatorname{coker}\alpha \circ \alpha = 0$. Since $\operatorname{coim}\alpha = \operatorname{coker}(\ker\alpha)$ is an epimorphism, we have $\operatorname{coker}\alpha \circ \delta = 0$. Since $\operatorname{im}\alpha$ is the largest subobject of y killed by $\operatorname{coker}\alpha$, there exists a unique $\alpha' : \operatorname{coim}\alpha \to \operatorname{im}\alpha$ in $\mathcal{C}$ such that

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d] & Y \\ \operatorname{coim}\alpha \arrow[ur, "\delta"'] & \end{tikzcd} \quad \text{and hence} \quad \begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d] & Y \\ \operatorname{coim}\alpha \arrow[r, "\alpha'"'] & \operatorname{im}\alpha \arrow[u] \end{tikzcd}$$

commutes. If we started the above construction with $\operatorname{im}\alpha$ instead of $\operatorname{coim}\alpha$, we would have constructed another morphism $\alpha'' : \operatorname{coim}\alpha \to \operatorname{im}\alpha$ in $\mathcal{C}$ making

$$\begin{tikzcd} X \arrow[r, "\alpha"] \arrow[d] & Y \\ \operatorname{coim}\alpha \arrow[r, "\alpha''"'] & \operatorname{im}\alpha \arrow[u] \end{tikzcd}$$

commute. But

$$\operatorname{im}\alpha \circ \alpha' \circ \operatorname{coim}\alpha = \operatorname{im}\alpha = \operatorname{im}\alpha \circ \alpha'' \circ \operatorname{coim}\alpha,$$

with $\operatorname{coim}\alpha$ an epimorphism and $\operatorname{im}\alpha$ a monomorphism. It follows that $\alpha' = \alpha''$.

Definition 139.14. An additive category $\mathcal{A}$ is called an *abelian category* if it satisfies all of the following:

1. Every morphism in $\mathcal{A}$ has a kernel and a cokernel.
2. If $\alpha : X \to Y$ in $\mathcal{A}$, the natural induced map $\operatorname{coim}\alpha \to \operatorname{im}\alpha$ is an isomorphism in $\mathcal{A}$, i.e., the First Isomorphism Theorem holds. We write $\operatorname{coim}\alpha$ by $X/\ker\alpha$.

Examples 139.15. 1. If $\mathcal{A}$ is an abelian category, then so is $\mathcal{A}^{op}$. However, in general, an abelian category is not equivalent to its dual.

2. ${}_R\mathcal{M}$, $\mathcal{M}_S$ and ${}_R\mathcal{M}_S$ are abelian categories for all rings R and S.
3. Let R be a left Noetherian ring. Then ((finitely generated left R-modules, R-homomorphisms)) is an abelian category.
4. The category ((abelian topological groups, continuous homomorphisms)) is not abelian as the First Isomorphism Theorem fails.
5. Let $(X, \mathcal{O}_X)$ be a ringed space, e.g., a differential manifold, complex analytic manifold, or scheme which are locally ringed spaces by Example 138.27. Let

   $$\mathbf{Mod}(X) = ((\text{sheaves of } \mathcal{O}_X\text{-modules, sheaf morphisms})),$$

   i.e., if F is a sheaf and $F(U)$ is an $\mathcal{O}_X(U)$-module for all open sets $U \subset X$. Then $\mathbf{Mod}(X)$ is abelian.

Definition 139.16. Let $\mathcal{A}$ be an abelian category and

$$(*) \qquad A \xrightarrow{f} B \xrightarrow{g} C$$

in $\mathcal{A}$. Then we have a commutative diagram

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "p"] & B \arrow[r, "g"] & C \\ \operatorname{coim} f \arrow[r, "\sim", "\bar{f}"'] & \operatorname{im} f. \arrow[u, "j"'] & \end{tikzcd}$$

We say that (*) is *exact* if $j : \operatorname{im} f \to B$ is a kernel of $g : B \to C$, i.e., $\ker g = \operatorname{im} f$. Exact sequences and short exact sequences are now defined in the usual way. If $\mathcal{B}$ is also an abelian category and $F : \mathcal{A} \to \mathcal{B}$ an additive functor, we say

1. F is *exact* if $A \xrightarrow{f} B \xrightarrow{g} C$ is exact in $\mathcal{A}$ implies $FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC$ is exact in $\mathcal{B}$.
2. F is *left exact* if $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact in $\mathcal{A}$ implies $0 \to FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC$ is exact in $\mathcal{B}$.
3. F is *right exact* if $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact in $\mathcal{A}$ implies $FA \xrightarrow{Ff} FB \xrightarrow{Fg} FC \to 0$ is exact in $\mathcal{B}$.

If F is contravariant, we use the same terminology given from $F : \mathcal{A}^{op} \to \mathcal{B}$.

We also call a sequence $X^* : \cdots \to X^{n_1} \xrightarrow{d^{n-1}} X^n \xrightarrow{d^n} X^{n+1} \to \cdots$ a *cochain complex* if $d^n d^{n-1} = 0$ for all n and the *cohomology* $H^n(X) = \ker d^n / \operatorname{im} d^{n-1}$ for all n.

Proposition 139.17. *Let $F : \mathcal{A} \to \mathcal{B}$ be an additive functor of abelian categories. Then F is exact if and only if F preserves short exact sequences, i.e., it is both left and right exact.*

PROOF. Exercise. □

Yoneda's Lemma 136.10 implies

Lemma 139.18. *Let $\mathcal{A}$ be an abelian category and $A \xrightarrow{\alpha} B$ and $B \xrightarrow{\beta} C$. Suppose that*

$$(\dagger) \qquad h_X(A) \xrightarrow{\alpha_*} h_X(B) \xrightarrow{\beta_*} h_X(C)$$

is exact in $\mathcal{A}$ for all $X \in \mathcal{A}$, i.e.,

$$(*) \qquad \operatorname{Hom}_{\mathcal{A}}(X, A) \xrightarrow{\alpha_*} \operatorname{Hom}_{\mathcal{A}}(X, B) \xrightarrow{\beta_*} \operatorname{Hom}_{\mathcal{A}}(X, C)$$

is exact in $\mathcal{A}$. Then $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ is exact in $\mathcal{A}$.

PROOF. (*) is a zero sequence: Let $X = A$, then $\beta\alpha = \beta_*\alpha_*(1_A) = 0$.

$\ker \beta \subset \operatorname{im} \alpha$: Let $X = \ker \beta$, so we have $\ker \beta \xrightarrow{\iota} B \xrightarrow{\beta} C$. So $\beta_* \iota = \beta\iota = 0$. By the exactness of (†), there exists $\sigma \in \operatorname{Hom}_{\mathcal{A}}(\ker \beta, A)$ satisfying $\iota = \alpha_*\sigma = \alpha\sigma$, i.e.,

$$\begin{tikzcd} & & X \arrow[d, "\alpha"] & \\ 0 \arrow[r] & \ker \beta \arrow[ur, "\sigma"] \arrow[r, "\iota"] \arrow[rr, bend right, "0"'] & B \arrow[r, "\beta"] & C. \end{tikzcd}$$

So $\ker \beta = \operatorname{im} \iota \rightarrowtail \operatorname{im} \alpha$ as needed. □

Theorem 139.19. *Let $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ be additive functors of abelian categories with (F, G) an adjoint pair. Then*

(1) F is right exact.
(2) G is left exact.

Proof. We first show (2). Let $0 \to B' \xrightarrow{f} B \xrightarrow{g} B'' \to 0$ be exact in $\mathcal{B}$. Then we have a commutative diagram

$$\begin{array}{ccccc} 0 \to \operatorname{Hom}_{\mathcal{B}}(FA, B') & \xrightarrow{h_{FA}(f)} & \operatorname{Hom}_{\mathcal{B}}(FA, B) & \xrightarrow{h_{FA}(g)} & \operatorname{Hom}_{\mathcal{B}}(FA, B'') \\ \Big\downarrow{\scriptstyle \tau_{A,B'}} & & \Big\downarrow{\scriptstyle \tau_{A,B}} & & \Big\downarrow{\scriptstyle \tau_{A,B''}} \\ 0 \to \operatorname{Hom}_{\mathcal{A}}(A, GB') & \xrightarrow{h^{GA}(f)} & \operatorname{Hom}_{\mathcal{A}}(A, GB) & \xrightarrow{h^{GA}(g)} & \operatorname{Hom}_{\mathcal{A}}(A, GB'') \end{array}$$

in $\mathcal{A}$ for all $A \in \mathcal{A}$ with verticals isomorphisms by adjointness. Therefore, the bottom row is exact. By Yoneda's Lemma 136.10, $0 \to GB' \xrightarrow{g_*} GB \xrightarrow{g_*} B''$ is exact.
(1) follows as (G^{op}, F^{op}) is an adjoint pair. □

By the proposition and Theorem 138.19, we have:

Corollary 139.20. *Let $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ be additive functors of abelian categories with (F, G) an adjoint pair. Then*

(1) F preserves direct limits when they exist. In particular, if $\mathcal{A}$ is cocomplete, it preserves all direct limits.
(2) G preserves inverse limits when they exist. In particular, if $\mathcal{B}$ is complete, G preserves all inverse limits.

We have proven many things about such exact sequences and zero sequences. The proofs also work for ${}_R\mathcal{M}$. In fact, the following fact is true.

Fact 139.21. (Mitchell Embedding Theorem) *Let $\mathcal{A}$ be a small abelian category. Then there exists a ring R and a fully faithful exact functor $F : \mathcal{A} \to {}_R\mathcal{M}$ that is injective on objects.*

This means if we work on a specific problem, one can often ignore abelian categories and assume the problem is in ${}_R\mathcal{M}$ for some ring R and prove it there. In particular, many of our proofs for the module case carry over. We shall use the Mitchell Embedding Theorem implicity below when applcable.

Exercises 139.22. 1. Prove Lemma 139.4.

2. Prove Proposition 139.5.
3. Prove Proposition 139.10.
4. Prove Proposition 139.8.
5. Let $\mathcal{A}$ be an abelian category and $\alpha : X \to Y$ a morphism in $\mathcal{A}$. Prove all of the following:
   (i) α is a monomorphism if and only if $\ker \alpha = 0$
   (ii) α is an epimorphism if and only if $\operatorname{coker} \alpha = 0$
   (iii) α is an isomorphism if and only if α is a monomorphism and an epimorphism.

(*iv*) Let $\mathcal{A}$ be an abelian category and $\alpha : X \to Y$ a morphism in $\mathcal{A}$. Prove that there exists a unique *factorization* (up to isomorphism) such that the diagram

$$\begin{tikzcd} X \arrow[rr, "\alpha"] \arrow[dr, two heads] & & Y \\ & Z \arrow[ur, hook] & \end{tikzcd}$$

commutes in $\mathcal{A}$.

6. Prove 139.17.
7. Let $F : \mathcal{A} \to \mathcal{B}$ be an additive functor (respectively, contravariant functor) of abelian categories. Prove that F is a left exact (respectively, right exact) if and only if whenever

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \qquad (\text{respectively, } A \xrightarrow{f} B \xrightarrow{g} C \to 0)$$

is exact, then

$$0 \to FA \xrightarrow{Ff} B \xrightarrow{Fg} C \qquad (\text{respectively, } FA \xrightarrow{f} FB \xrightarrow{g} FC \to 0)$$

is exact.

140. Derived Functors

In this section, we indicate which of our results from Chapter XXI are valid for certain abelian category using the Mitchell Embedding Theorem 139.21. One can prove these results and others in this section without using this theorem, but working without elements (if they exist) can be tedious. This section is really a discussion of how one sets up a (co)homology theory in an abelian category satisfying suitable conditions. We set up the omenclature and results that can be proven, but omit proofs. We begin by using the same language in an abelian category as we did in a for modules, i.e., in ${}_R\mathcal{M}$. .

Definition 140.1. Let $\mathcal{A}$ be an abelian category. A *chain complex* in $\mathcal{A}$ is *sequence*

$$\cdots \xrightarrow{d_{n+2}} A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \xrightarrow{d_{n-2}} \cdots$$

in $\mathcal{A}$ with the morphisms d_i satisfying $d_{n-1}d_n = 0$ for all n, called *differentials*, i.e., *a zero sequence*. Denote such a chain complex by (A_*, d_*) or simply A_* if d_* is clear. If (A_*, d_*) is a chain complex, then we set

$$H_n(A) := \ker d_n / \operatorname{im} d_{n+1}$$

the quotient in $\mathcal{A}$ and call it the *nth homology of* (A_*, d_*). A chain complex is called *acyclic* if it is an exact sequence, i.e., $H_n(A) = 0$ for all n. If (A_*, d_*), $n \geq 0$, is acyclic and $M \in \operatorname{Ob}(\mathcal{A})$, satisfies

$$\cdots \xrightarrow{d_2} A_2 \xrightarrow{d_1} A_1 \xrightarrow{d_0} A_0 \xrightarrow{\varepsilon} M \to 0$$

is acyclic, we call it an *acyclic resolution* of M with *augmentation* ε. We write this as $A_* \xrightarrow{\varepsilon} M \to 0$. The chain complex $A_* \to 0$ is called the *deleted resolution* of A. Note that the deleted resolution is not exact in general (at A_0).

If (A_*, d_*) is a chain complex, the objects A_n are called the *n-chains* and

$$Z_n(A) := \ker d_n \text{ the } n\text{-}cycles \text{ of } A_n$$
$$B_n(A) := \operatorname{im} d_{n+1} \text{ the } n\text{-}boundaries\ of\ A_n.$$

If we write the indices of the A_* and the d_n to go up, i.e.,

$$\cdots \xrightarrow{d_{n-2}} A_{n-1} \xrightarrow{d_{n-1}} A_n \xrightarrow{d_n} A_{n+1} \xrightarrow{d_{n+1}} \cdots,$$

then (A_*, d_*) is called a *cochain complex* with the differentials written as d^n. We use the notation (A^*, d^*) for a cochain complex. Analogously, A^n are called the *n-cochains* and

$$Z^n(A) := \ker d^n \text{ the } n\text{-}cocycles \text{ of } A_n$$
$$B^n(A) := \operatorname{im} d^{n-1} \text{ the } n\text{-}coboundaries \text{ of } A_n$$
$$H^n(A) := \ker d^n / \operatorname{im} d^{n-1} = Z^n(A)/B^n(A) \text{ the } nth\text{-}cohomology \text{ of } A_n.$$

An acyclic resolution with augumentation ε of $M \in \operatorname{Ob}\mathcal{A}$ is an exact sequence $0 \to M \xrightarrow{\varepsilon} A^*$, with A_n, $n \geq 0$, acyclic.

We have a category

$$\mathbf{Chain}(\mathcal{A}) := ((\text{chains in } \mathcal{A},\ \text{chain homomorphisms})),$$

where a morphism of chains, $f : (A_*, (d_A)_*) \to (B_*(d_B)_*)$ is a commutative diagram

$$\begin{tikzcd}
\cdots \arrow[r, "d_{n+2}"] & A_{n+1} \arrow[r, "d_{n+1}"] \arrow[d, "f_{n+1}"'] & A_n \arrow[r, "d_n"] \arrow[d, "f_n"'] & \cdots \\
\cdots \arrow[r, "d'_{n+2}"'] & A'_{n+1} \arrow[r, "d'_{n+1}"'] & A'_n \arrow[r, "d_n"'] & \cdots .
\end{tikzcd}$$

An exact sequence of chains is a morphism of chains

$$\cdots \to (A_*, (d_A)_*) \xrightarrow{(f_A)_*} (B_*, (d_B)_*) \xrightarrow{(f_B)_*} (C_*, (d_C)_*) \xrightarrow{(f_C)_*} \cdots$$

with

$$\cdots \to A_n \xrightarrow{(f_A)_n} B_n \xrightarrow{(f_B)_n} C_n \xrightarrow{(f_C)_n} \cdots$$

exact at every n. We often omitting the maps f_n, etc. when clear.

Usually will be interested in chains starting at 0.

Since the Snake Lemma 122.1 holds in an abelian category, we see that the following are true:

Theorem 140.2. *Let $\mathcal{A}$ be an abelian category. Suppose that*

$$0 \to (A_*, d_A) \xrightarrow{f_*} (B_*, d_B) \xrightarrow{g_*} (C_*, d_C) \to 0$$

is an exact sequence of chain complexes in $\mathcal{A}$. Then for all $n \geq 0$, there exist morphisms $\partial_{n+1} : H_{n+1}(C) \to H_n(A)$, called the connecting morphisms of the sequence, that give rise to a long exact sequence in homology

$$\cdots \to H_{n+1}(C) \xrightarrow{\partial_{n+1}} H_n(A) \xrightarrow{\overline{f}_n} H_n(B) \xrightarrow{\overline{g}_n} H_n(C) \xrightarrow{\partial_n} H_{n-1}(A) \to \cdots$$

where $\overline{f}_n$ and $\overline{g}_n$ are the induced maps for all n.

Theorem 140.3. *Let $\mathcal{A}$ be an abelian category and*

$$\begin{tikzcd}
0 \arrow[r] & (A_*, d_A) \arrow[r, "f"] \arrow[d, "\alpha_*"'] & (B_*, d_B) \arrow[r, "g"] \arrow[d, "\beta_*"'] & (C_*, d_C) \arrow[r] \arrow[d, "\gamma_*"'] & 0 \\
0 \arrow[r] & (A'_*, d_{A'}) \arrow[r, "f'"] & (B'_*, d_{B'}) \arrow[r, "g'"] & C'_*, d_{C'}) \arrow[r] & 0
\end{tikzcd}$$

be a commutative diagram of chain complexes in $\mathcal{A}$. Then there exists a commutative diagram with exact rows

$$\begin{tikzcd}
\cdots \to H_n(A) \arrow[r, "f_n"] \arrow[d, "\overline{\alpha}_n"'] & H_n(B) \arrow[r, "g_n"] \arrow[d, "\overline{\beta}_n"'] & H_n(C) \arrow[r, "\partial_n"] \arrow[d, "\overline{\gamma}_n"'] & H_{n-1}(A) \to \cdots \arrow[d, "\overline{\alpha}_{n-1}"'] \\
\cdots \to H_n(A') \arrow[r, "f'_n"] & H_n(B') \arrow[r, "g'_n"] & H_n(C') \arrow[r, "\partial'_n"] & H_{n-1}(A') \to \cdots ,
\end{tikzcd}$$

where ∂_n and ∂'_n are the corresponding connecting morphisms and $\overline{f}_n$, $\overline{g}_n$ $\overline{\alpha}_n$, $\overline{\beta}_n$, and $\overline{\gamma}_n$ the induced maps for all n.

We are interested in the following question: Let $F : \mathcal{A} \to \mathcal{B}$ be a functor (or contravariant functor) of abelian categories. If $A_* \to M \to 0$ in $\mathcal{A}$ is an acyclic resolution of M, what can you say about the chain complex $\big(F(A_*)\big) \to F(M) \to 0$, in particular about its homology $H_n\big(F(A)\big)$?

Definition 140.4. Let $T_n : \mathcal{A} \to \mathcal{B}$, $n \geq 0$, be additive functors (respectively, $T^n : \mathcal{A} \to \mathcal{B}$, $n \geq 0$, be additive functors) satisfying the following:

Given a short exact sequence $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ in $\mathcal{A}$, there exist morphisms $\partial_n : T_n(C) \to T_{n-1}(A)$ (respectively $\delta^n : T^n(C) \to T^{n+1}(A)$) with the following properties:

1. We have a long exact sequence

$$\cdots \xrightarrow{g_{n-1}} T_{n+1}(C) \xrightarrow{\partial_n} T_n(A) \xrightarrow{f_n} T_n(B) \xrightarrow{g_n} T_n(C) \xrightarrow{\partial_{n-1}} \cdots$$

(respectively,

$$\cdots \xrightarrow{g^{n-1}} T^{n-1}(C) \xrightarrow{\delta^n} T^n(A) \xrightarrow{f^n} T^n(B) \xrightarrow{g^n} T_n(C) \xrightarrow{\delta_{n+1}} \cdots)$$

in $\mathcal{B}$.

As we are assuming that T_n (respectively T^n) are trivial for $n < 0$, this also means that T_0 is right exact (respectively, T^0 is left exact).

2. If $0 \to A' \xrightarrow{f'} B' \xrightarrow{g'} C' \to 0$ is another exact sequence in $\mathcal{A}$ and there exists a morphism from it to $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$, then we have commutative diagrams

$$\begin{tikzcd}
T_n(C') \arrow[r, "\partial'_n"] \arrow[d] & T_{n-1}(A') \arrow[d] \\
T_n(C) \arrow[r, "\partial_n"'] & T_{n-1}(A)
\end{tikzcd}
\quad \text{(respectively,} \quad
\begin{tikzcd}
T^n(C') \arrow[r, "\delta'^n"] \arrow[d] & T^{n+1}(A') \arrow[d] \\
T^n(C) \arrow[r, "\delta^n"'] & T^{n+1}(A)
\end{tikzcd}
\text{).}$$

We call this data $T_* = (T_n)_{n\geq 0}$ a *homological ∂-functor* (respectively, $T^* = (T^n)_{n\geq 0}$ a *cohomological δ-functor*) between $\mathcal{A}$ and $\mathcal{B}$.

A morphism $\tau : (T_n)_{n\geq 0} \to (S_n)_{n\geq 0}$ (respectively, $\tau : (T^n)_{n\geq 0} \to (S^n)_{n\geq 0}$) are natural transformations such that

$$\begin{tikzcd} T_n(C) \arrow[r, "\partial_n"] \arrow[d] & T_{n-1}(A) \arrow[d] \\ S_n(C) \arrow[r, "\partial_n"'] & S_{n-1}(A) \end{tikzcd} \quad \text{(respectively, } \begin{tikzcd} T^n(C) \arrow[r, "\delta^n"] \arrow[d, "\tau_{n,C}"'] & T^{n+1}(A) \arrow[d, "\tau_{n-1,A}"] \\ S^n(C) \arrow[r, "\delta^n"'] & S^{n+1}(A) \end{tikzcd} \text{)}$$

commute.

A homological ∂-functor T is called a *homological universal ∂-functor* if given any other ∂-functor S and a natural transformation $f_0 : S_0 \to T_0$, then there exists a unique morphism $f_n : T_n \to S_n$ of ∂-functors extending f_0.

A cohomological δ-functor T is called a *universal cohomological δ-functor* if given any other δ-functor S and a natural transformation $f^0 : T^0 \to S^0$, then there exists a unique morphism $f^n : T_n \to S^n$ of δ-functors extending f^0.

Definition 140.5. An additive functor $F : \mathcal{A} \to \mathcal{B}$ of abelian categories is called *effaceable* if for every $A \in \mathrm{Ob}(\mathcal{A})$ there exists a monomorphism $f : A \to E$ satisfying $F(f) = 0$. It is called *coeffeaceable* if for every if for every $A \in \mathrm{Ob}(\mathcal{A})$ there exists an epimorphism $f : E \to A$ satisfying $F(f) = 0$.

The proof of the following is left as a (long and tedious) exercise.

Lemma 140.6. *Let $T_* : \mathcal{A} \to \mathcal{B}$ be a homological ∂-functor (respectively, T^* be a cohomological δ-functor) that is effaceable (respectively coeffaceable) for all $n > 0$, then T_* (respectively, T^*) is a universal homological ∂-functor (respectively, T^* is a universal cohomological δ-functor).*

Definition 140.7. Let $\mathcal{A}$ be an abelian category. An object $I \in \mathrm{Ob}(\mathcal{A})$ is called an *injective object* if given a diagram

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "g"'] & B \\ & I & \end{tikzcd} \quad , \quad \text{there exists } h \text{ such that} \quad \begin{tikzcd} 0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[dl, "h"] \\ & I & \end{tikzcd}$$

commutes in $\mathcal{A}$.

We say that $\mathcal{A}$ has *enough injectives* if for every object $M \in \mathcal{A}$ there exists an injective object $I \in \mathrm{Ob}(\mathcal{A})$ and a monomorphism $0 \to M \xrightarrow{\varepsilon} I$.

Remark 140.8. If $(X, \mathcal{O}_X)$ is a ringed space, then it can be shown that the category $\mathbf{Mod}(X)$ contains enough injectives.

If $\mathcal{A}$ is an abelian category with enough injections, then every $M \in \mathcal{A}$ has an injective resolution $0 \to M \to I^*$ in $\mathcal{A}$.

The main facts about injective modules in Section **??** hold for injective objects in a abelian category $\mathcal{A}$, viz., if $A \in \mathcal{A}$ then there exists an injective object $I \in \mathrm{Ob}(\mathcal{A})$, and if $I \in \mathrm{Ob}(\mathcal{A})$ is injective in an abelian category $\mathcal{A}$ if and only if whenever $0 \to I \to B \to C \to 0$ is exact, is split exact, i.e., $I \to B$ is a split monomorphism.

Let $\mathcal{A}$ be an abelian category with enough injectives and $F : \mathcal{A} \to \mathcal{B}$ be a left exact functor (or contravariant functor). Then we define the *right derived functors* R^iF of F, $i \geq 0$, as follows: For each $M \in \mathrm{Ob}(\mathcal{A})$, let $0 \to M \xrightarrow{\varepsilon} I^*$ be an injective resolution. Set

$$R^iF(M) := H^i\big(F(I^*)\big)$$

for the deleted complex I^*. These cohomology groups are called the *right derived functors* of F. In particular, F is effaceable.
We are now in the situation studied in Chapter XXI and we see that we have

Theorem 140.9. *Let $\mathcal{A}$ be an abelian category.*

(1) The Comparison Theorem 125.2 holds in $\mathcal{A}$.
(2) The Horseshoe Lemma 125.6 holds in $\mathcal{A}$.

It follows that $H^iF(M) := H^iF(I^*)$ is independent of the injective resolution.

Let $\mathcal{A}$ is a abelian category with enough injectives and $F : \mathcal{A} \to \mathcal{B}$ a left exact functor of abelian categories. An object $M \in \mathrm{Ob}(\mathcal{A})$ is called *F-acyclic* if $R^iF(M) = 0$ for all $i > 0$, e.g., if M is itself injective, as then $0 \to I \to I$ is an injective resolution. We do not need to have an injective resolutions to compute these cohomology groups. Indeed

Proposition 140.10. *Let $F : \mathcal{A} \to \mathcal{B}$ be a left exact functor of abelian categories. Suppose that $\mathcal{A}$ has enough injectives. Let $0 \to M \to X^*$ be exact with $X^* = (X^*, d)$ satisfying each X^i is F-acyclic. Then $H^iF(X) \cong R^iF(M)$ (naturally).*

PROOF. Exercise □

Lemma 140.11. *Let $F : \mathcal{A} \to \mathcal{B}$ be a left exact functor of abelian categories. Suppose that $\mathcal{A}$ has enough injectives. Then F is effaceable if and only if $F(I) = 0$ for all injective objects $I \in \mathrm{Ob}(\mathcal{A})$.*

PROOF. ($\Rightarrow$): Suppose that $I \in \mathrm{Ob}(\mathcal{A})$ is injective and $0 \to I \xrightarrow{i} A$ is exact with $F(I) = 0$. As i is a split monomorphism, $A = i(I) \coprod B$ for some $B \in \mathrm{Ob}(\mathcal{C})$. Since F preserves split exact sequences and $F(i) = 0 : F(I) \to F(I) \coprod F(B)$, we conclude that $F(I) = 0$.

($\Leftarrow$): If $M \in \mathrm{Ob}(\mathcal{A})$, choose $I \in \mathrm{Ob}(\mathcal{A})$ injective such that $0 \to M \xrightarrow{i} I$ is exact. As $F(I) = 0$, we have $F(i) = 0$. □

Corollary 140.12. *Let $\mathcal{A}$ and $\mathcal{B}$ be abelian categories with $\mathcal{A}$ having enough injectives. If $F : \mathcal{A} \to \mathcal{B}$ is a left exact functor, then (R^iF, δ) is a universal δ-functor.*

Summary 140.13. Let $F : \mathcal{A} \to \mathcal{B}$ be a left exact functor of abelian categories. Suppose that $\mathcal{A}$ has enough injectives. Then

(1) For each $i > 0$, $R^iF : \mathcal{A} \to \mathcal{B}$ is an additive functor independent of the injective resolutions (up to a natural isomorphism of functors).
(2) There is a natural isomorphism $F \to R^0F$.

(3) For each short exact sequence $0 \to A' \to A \to A'' \to 0$ in $\mathcal{A}$, there is a natural morphism $\delta^i : R^iF(A'') \to R^{i+1}(A')$ such that we obtain a long exact sequence

$$\cdots \to R^iF(A') \to R^iF(A'') \xrightarrow{\delta^i} R^{i+1}F(A) \to \cdots .$$

(4) Given a morphism of exact of short exact sequences $0 \to A' \to A \to A'' \to 0$ to $0 \to B' \to B \to B'' \to 0$, we have commutative diagrams

$$\begin{tikzcd} R^iF(A'') \arrow[r, "\delta^n"] \arrow[d] & R^{i+1}F(A') \arrow[d] \\ R^iF(B'') \arrow[r, "\delta^n"] & R^{i+1}(FB'). \end{tikzcd}$$

(5) For each injective $I \in \mathrm{Ob}(\mathcal{C})$ and $i > 0$, $R^iF(I) = 0$.
(6) If $0 \to M \to X^*$ is exact with each X^i F-acyclic for $i \geq 0$, then there is a natural isomorphism $R^iF(M) \cong H^i(X)$.
(7) $(R^iF)_{i\geq 0}$ is a universal cohomological δ-functor.

We have can also look at abelian categories with projective objects.

Definition 140.14. An object $P \in \mathrm{Ob}(\mathcal{A})$ in an abelian category is called an *projective object* if given a diagram

$$\begin{tikzcd} & P \arrow[d, "g"] & \\ B \arrow[r, "f"'] & A \arrow[r] & 0 \end{tikzcd}, \quad \text{there exists } h \text{ such that} \quad \begin{tikzcd} & P \arrow[d, "g"] \arrow[dl, "h"'] & \\ B \arrow[r, "f"'] & A \arrow[r] & 0 \end{tikzcd}$$

commutes in $\mathcal{A}$.

So a projective object $P \in \mathrm{Ob}(\mathcal{A})$ is just an object such that $P^{op} \in \mathrm{Ob}(\mathcal{A}^{op})$ is injective.

We say that $\mathcal{A}$ has *enough projectives* if for every object $M \in \mathcal{A}$ there exists an projective object $P \in \mathrm{Ob}(\mathcal{A})$ and an epimorphism $P \xrightarrow{\varepsilon} M \to 0$. In particular, if $\mathcal{A}$ has enough projectives, then $M \in \mathcal{A}$ has a projective resolution $P^* \xrightarrow{\varepsilon} M \to 0$ in $\mathcal{A}$.

Remark 140.15. It is not as common for an abelian category to have enough projective objects.

We also have all the analogues of the results above for left exact contravariant functors $F : \mathcal{A} \to \mathcal{B}$ of abelian categories when $\mathcal{A}$ has enough injectives or projectives in Summary 140.13. We leave the statements of these results to the reader.

In the case of modules we also have some results involving products and coproducts as additive functors preserve split short exact sequences, e.g., they take preserve finite products and coproducts. However, if we want the result for arbitrary products or coproducts, we would also need that the category is complete and/or cocomplete.

Moreover, we still have for abelian categories dimension shifting when an abelian category has a universal δ-functor (or dually a universal ∂-functor), i.e.,

Proposition 140.16. (Dimension Shifting) *Let (T^*, δ) be a δ-functor $\mathcal{A} \to \mathcal{B}$ of abelian categories and $E \in \mathcal{A}$ such that $T^i E = 0$ for all $i > 0$. If $0 \to A \to E \to B \to 0$ is a short exact sequence, then $T^i B \xrightarrow{\sim} T^{i+1} A$ for all $i > 0$. More generally, if*

$$0 \to A \to E_0 \to E_1 \to \cdots \to E_n \to B \to 0$$

is exact, with $T^i(E_j) = 0$ for all j and all $i > 0$, then $T^i B \cong T^{i+n+1} A$.

whose proof we leave as an exercise.

We also have the dual notion for homology. Let $\mathcal{A}$ be an abelian category with enough projectives and $F : \mathcal{A} \to \mathcal{B}$ a right exact functor of abelian categories. Then define the *left derived functors* $L_i F$ of F, $i \geq 0$, as follows: For each $M \in \mathrm{Ob}(\mathcal{A})$, let $(P^*) \xrightarrow{\varepsilon} M$ be a projective resolution of M in $\mathcal{A}$. Set

$$L_i F(M) := H_i(F(P^*))$$

for the deleted complex P^* called the *left derived functors* of F. In particular, F is coeffaceable. We then get the results analogous to the right derived functor case whose statements we leave to the reader.

Examples 140.17. 1. Let $M, N \in {}_R\mathcal{M}$. Then we have a universal cohomological δ-functor and a universal cohomological contravariant ${}_R\mathcal{M} \to \mathcal{A}$ given by $\mathrm{Ext}^i_R(\ \ , N) := R^i h^N$ and $\underline{\mathrm{Ext}}^i_R(M,\ \ ) = R^i h_M$, respectively. If R is a commutative ring, these are functors ${}_R\mathcal{M} \to {}_R\mathcal{M}$ as ${}_R\mathcal{M}$ is a full and faithful subcategory of $\mathcal{A}$. We also have $\mathrm{Ext}^i_R(M, N) \cong \underline{\mathrm{Ext}}^i_R(M, N)$ naturally for all $M, N \in \mathrm{Ob}({}_R\mathcal{M})$.

2. Let $M \in \mathcal{M}_R$ and $N \in {}_R\mathcal{M}$. Then we have universal homology ∂-functors $\mathrm{Tor}^R_i(\ \ , N) := L_i(\ \ \otimes_R N) : \mathcal{M}_R \to \mathbf{Group}$ and $\underline{\mathrm{Tor}}^i_R(M,\ \ ) = L_i(M \otimes_R\ \ ) : {}_R\mathcal{M} \to \mathcal{A}$. If R is a commutative ring, then both functors take ${}_R\mathcal{M} \to {}_R\mathcal{M}$. We also have $\mathrm{Tor}^R_i(M, N) \cong \underline{\mathrm{Tor}}^R_i(M, N)$ naturally for all $M \in \mathrm{Ob}(\mathcal{M}_R)$ and $N \in \mathrm{Ob}({}_R\mathcal{M})$.

Exercises 140.18. 1. Prove Lemma 140.6.

2. Prove Prop 140.16.

Appendices

APPENDIX A

Axiom of Choice and Zorn's Lemma

We have alluded to the Axiom of Choice in the text, but have used Zorn's Lemma 28.5 for the most part. These two axioms of set theory are equivalent. In this appendix, we show this. We shall also show that Zorn's Lemma is also equivalent to the Well-ordering Principle.

Recall from Section 28.5 the following: A pair $(S, \leq)$ (or S if $\leq$ is clear) is called a *partially ordered set* or *poset* (under $\leq$) if the following holds: For all a, b, c in S, we have $a \leq a$; if $a \leq b$ and $b \leq a$, then $a = b$; and if $a \leq b$ and $b \leq c$, then $a \leq c$. Let $(S, \leq)$ be a poset. We call $(S, \leq)$ a *chain* if for all a and b in S, we have either $a \leq b$ or $b \leq a$. Let T be a subset of S. An element a in S is called an *upper bound* of T if $x \leq a$ for all $x \in T$, and s is called a *maximal element* of S if $s \leq y$ with $y \in S$ implies that $s = y$. We say that the poset S is *inductive* if every chain in S has an upper bound in S. We also shall write $a < b$ for $a \leq b$ with $a \neq b$. We shall also write $\geq$ and $>$ when it enhances the situation.

[Note: In this appendix for clarity, we write $\subseteq$ instead of $\subset$ as in the text.]

Zorn's Lemma is the axiom that we assume to always be true. It says:

Lemma A.1. (Zorn's Lemma) *Let S be a nonempty inductive poset. Then S contains a maximal element.*

Call an inductive poset S *strongly inductive* if every chain $\mathcal{C}$ in S has a *least upper bound*, i.e., an upper bound $a \in S$ such that $a \leq x$ for every upper bound x in S of $\mathcal{C}$. If the least upper bound for a chain lies in the chain, we shall call it a *last element* of the chain. We have similar definitions for lower bounds, least lower bounds, and first elements. As every strongly inductive poset is inductive, we have the following that be called a strong form of Zorn's Lemma.

Lemma A.2. *Assume that Zorn's Lemma holds. Let S be a nonempty poset. If S is strongly inductive, then S contains a maximal element.*

The converse is also true.

Proposition A.3. *Every nonempty strongly inductive poset has a maximal element if and only if Zorn's Lemma is valid.*

Proof. Suppose that S is a partially ordered set and every chain in S has an upper bound. We must show that S has a maximal element. Let $\mathcal{P}(S) = \{A \mid A \subseteq S\}$, be the power set of S. It is a poset via set inclusion $\subseteq$. Let $\mathfrak{C} = \{\mathcal{C} \mid \mathcal{C} \text{ is a chain in } S\}$. Since $\mathfrak{C} \subseteq \mathcal{P}(S)$, it is also partially ordered by $\subseteq$. Let $\mathcal{C} \in \mathfrak{C}$ and $C_0 = \bigcup_{C \in \mathcal{C}} C$. Then C_0 is a least upper bound for $\mathcal{C}$. Therefore, $\mathfrak{C}$ is strongly inductive, so has a maximal element C'

in $\mathfrak{C}$. It follows that there is an upper bound x_0 for C', viz., the union of the sets in C' and this $x_0 \in C_0$, so is a maximal element of S. Therefore, Zorn's Lemma holds. $\square$

We generalize the notion of the Well-ordering Principle for the integers.

Definition A.4. Let $(S, \leq)$ be a poset. We say that $(S, \leq)$ is *well-ordered* if every nonempty subset A of S has a first element. This is equivalent to every nonempty subset of a chain in S has a first element.

Definition A.5. The *Well-ordering Axiom* says if S is a nonempty set, then there exists a partial ordering $\leq$ on S such that $(S, \leq)$ is well-ordered.

Proposition A.6. *Suppose every nonempty strongly inductive poset has a maximal element. Then the Well-ordering Axiom holds.*

PROOF. Let S be a nonempty set. We must show that S can be well-ordered. We know that any partial ordering on S is inductive by Zorn's Lemma, hence strongly inductive. For each $A \subseteq S$ that can be well ordered, let W_A be the set of partial orderings on A that are well-orderings and set $W = \{(A, \leq_A) \mid \emptyset \neq A \subseteq S,\ \leq_A \in W_A\}$. Clearly, W is not the empty set as $(\{a\}, =)$ lies in W for $a \in A$. Define a partial ordering $\leq$ on W by setting $(A, \leq_A) \leq (B, \leq_B)$ in W, if $A \subseteq B$, $\leq_B|_A = \leq_A$, and if $a \in A$, $b \in B$, then $a \leq_B b$.
Let $\mathcal{C}$ be a chain in W. So the set $\mathcal{C}$ consists of all chains in $\mathcal{C}$. Set $C_0 = \bigcup_{(C, \leq_C) \in \mathcal{C}} C$ and define $\leq_{C_0}$ by $a \leq_{C_0} b$ if there exists $(C, \leq_C) \in \mathcal{C}$ with $a, b \in C$ and $a \leq_C b$. As $\mathcal{C}$ is a chain this is independent of C with $a, b \in C$. Therefore, C_0 is well-ordered by $\leq_{C_0}$ and $(C_0, \leq_{C_0})$ is the least upper bound for $\mathcal{C}$. Since $(C_0, \leq_{C_0})$ is strongly inductive, there exist a maximal element $(A, \leq_A)$ for W. If $(A, \leq_A) < W$, let $b \in W \setminus A$ and $B = A \cup \{b\}$. Define $\leq_B$ by $x \leq_B b$ for all $a \in A$ and $\leq_B|_A = \leq_A$. Then $(B, \leq_B) \in W$ and $(A, \leq_A) < (B, \leq_B)$. This contradicts the maximality of $(A, \leq_A)$. Hence $W = S$, and the Well-ordering Axiom holds. $\square$

There are a number of variants for the statement of the Axiom of Choice. We show:

Proposition A.7. *The following are equivalent:*

(1) *If $\mathcal{S}$ is a nonempty set of nonempty disjoint sets A, then there exists a set B whose elements consist of precisely one element from each set $A \in \mathcal{S}$.*
(2) *If $\mathcal{S}$ is a nonempty set of nonempty sets A, then there exists a function $\varphi : \mathcal{S} \to \bigcup_{\mathcal{S}} A$ satisfying $\varphi(A) \in A$ for all $A \in \mathcal{S}$.*
(3) *If I is a nonempty set and A_i is a nonempty set for each $i \in I$, then the cartesian product $\bigtimes_I A_i : \{f : I \to \bigcup_I A_i \mid f(i) \in A_i \text{ for all } i \in I\}$ is nonempty.*

PROOF. (1) $\Rightarrow$ (3): Let $B_i = \{(i, x) \mid x \in A_i\}$, a nonempty set for each $i \in I$. Set $\mathcal{S} = \{B_i \mid i \in I\}$, a nonempty collection of disjoint sets. By (1) there exists a set

$$B = \{(i, a_i) \mid \text{ precisely one element } a_i \text{ in } A_i \text{ for each } i \text{ in } I\}.$$

Define $\varphi : I \to \bigcup_I A_i$ by $i \mapsto a_i$. Then φ works.

(3) $\Rightarrow$ (2): By (3), the set $\bigtimes_{\mathcal{S}} A$ is nonempty. Therefore, there exists a function $\varphi : \mathcal{S} \to \bigcup_{\mathcal{S}} S$ such that $\varphi(A) \in A$ for all $A \in \mathcal{S}$.

(2) $\Rightarrow$ (1): Let $\mathcal{S}$ be a nonempty collection of nonempty disjoint sets. By (2) there exist a function $\varphi : \mathcal{S} \to \bigcup_{A\in\mathcal{S}} S$ satisfying $\varphi(A) \in A$ for all $A \in \mathcal{S}$. Then the image of φ has precisely one element from each $A \in \mathcal{S}$. $\square$

Definition A.8. We say the *Axiom of Choice* holds if the equivalent conditions of Proposition A.7 hold for all nonempty sets.

Proposition A.9. *Suppose the Well-ordering Axiom holds. Then the Axiom of Choice holds.*

PROOF. Given a set B, let $\mathcal{S} = \{A \mid A \text{ is a nonempty set}\}$ and set $B = \bigcup_{\mathcal{S}} A$. As the Well-ordering Axiom holds, there exists a well-ordering $\leq_B$ on B. In particular, each $A \in \mathcal{S}$ is a nonempty subset of the well-ordered set B. For each $A \in \mathcal{S}$, let a be the first element of A under $\leq_B$. Then $\varphi : \mathcal{S} \to B$ by $A \mapsto a$ satisfies $\varphi(A) \in A$ for all $A \in \mathcal{S}$. This establishes (2) of Proposition A.7. $\square$

Lemma A.10. *Suppose that S is a strongly inductively ordered set and $\varphi : S \to S$ a function satisfying $\varphi(x) \geq x$ for every $x \in S$. If $a \in S$, then there exists an $x_0 \in S$ satisfying $x_0 \geq a$ and $\varphi(x_0) = x_0$.*

PROOF. Let $\mathcal{B}_a$ be the set of all subsets B in S that satisfy all of the following:

(*i*) $a \in B$.
(*ii*) $\varphi(B) \subseteq B$.
(*iii*) If $\mathcal{C}$ is a chain in B, then the least upper bound of $\mathcal{C}$ lies in B.

Since the set $\{x \mid x \geq a\}$ lies in $\mathcal{B}_a$. $\mathcal{B}_a$ is a nonempty set. Let $A = \bigcap_{\mathcal{B}_a} B$. Then check that $A \in \mathcal{B}_a$. In particular, if $z \in A$, then $z \geq a$.

Claim. A is a chain.

Suppose that we show the claim is valid. Let x_0 be the least upper bound for A. Then x_0 lies in A, since S is strongly inductive and $\varphi(x_0)$ lies in B by *iii*). Thus $\varphi(x_0) \leq x_0 \leq \varphi(x_0)$, so $x_0 = \varphi(x_0)$ and the lemma would be proven. So we need only check that A is a chain, i.e., satisfies (*iii*). To do so let

$$P = \{p \in A \mid \text{if } y \in A \text{ and } y < p, \text{ then } \varphi(y) \leq p.\}$$

We have $P \subseteq A$ and wish to show that $A = P$. To show this, fix $p \in P$ and define

$$B(p) = \{z \in A \mid \text{either } z \leq p \text{ or } z \geq \varphi(p)\},$$

i.e., $B(p)$ contains no $z \in A$ properly between $\varphi(p)$ and p.

Step 1. $A = B(p)$:

We show that $(i), (ii), (iii)$ hold for $B(p)$. Let $z \in A$.

(*i*): For all $y \in A$, we have $a \leq y$ and by (*ii*) for A, we know that $\varphi(y) \in A$. So $a \in B(p)$.

(*ii*): If $z < p$, then $\varphi(z) \leq p$ by definition of P. If $z = p$, then $\varphi(z) = \varphi(p) \geq p$. If $z \geq \varphi(p)$, then $\varphi(z) \geq z \geq \varphi(p) \geq p$. It follows that if $z \in B(p)$, then $\varphi(z)$ lies in $B(p)$.

(*iii*): Let $\mathcal{C}$ be a chain in $B(p)$ and c_0 the least upper bound of $\mathcal{C}$. Then c_0 lies in A. Suppose that $c \leq p$ for all $c \in \mathcal{C}$. Then $c_0 \leq p$, hence $\varphi(c_0) \leq p$ by (*ii*), so c_0 lies in $B(p)$. Otherwise, there exists a $c \in \mathcal{C}$ satisfying $c \geq \varphi(p)$. Then $c_0 \geq c \geq \varphi(p)$. This also implies that c_0 lies in $B(p)$ using (*ii*). It follows that the least upper bound of $\mathcal{C}$ lies in $B(p)$.

We have shown that $B(p) \in \mathcal{B}_a$. Since $B(p) \subseteq A$, we must have $B(p)$. This establishes Step 1.

Step 2. $P = A$.

We show that $(i), (ii), (iii)$ hold for A and P. Let $z \in A$.

(i): We have $a \in P$, since no $y \in A$ can satisfy $y < a$.

(ii): Let $p \in P$. We want to show that $\varphi(p) \in P$, i.e., if $z \leq \varphi(p)$, then $\varphi(z) \leq \varphi(p)$. We know that $\varphi(p) \in A$ by (ii) for A. Suppose that $z < \varphi(p)$. Since $A = B(p)$, we must have $z \leq p$. If $z < p$, then by definition of P, we have $\varphi(z) \leq p \leq \varphi(p)$; and if $z = p$, then $\varphi(z) = \varphi(p) \leq \varphi(p)$. Hence $\varphi(p) \in P$ and $\varphi(P) \subseteq P$.

(iii): Let $\mathcal{C}$ be a chain in P. Then the least upper bound c_0 of $\mathcal{C}$ lies in A. Let $z \in A$. Suppose that we have $z < c_0$. If $z \in P$, then $c_0 \geq \varphi(z)$ by (ii) for P. If $z \notin P$ then for each $c \in \mathcal{C}$, either $z < c$ or $z \geq \varphi(c) \geq c$ as $A = B(p)$. But c_0 is the least upper bound of $\mathcal{C}$, so we cannot have $z \geq c$ for all $c \in \mathcal{C}$. Therefore, $c_0 \in P$.

It follows that $P \in \mathcal{B}_a$. Since $P \subseteq A$, we must have $P = A$. Finally, if $x \in A$ and $y \in A$, then either $x \leq y$ or $x \geq \varphi(z) \geq z$. Therefore, A is a chain. This proves the lemma. □

Proposition A.11. *Suppose the Axiom of Choice holds. Then Zorn's Lemma holds.*

Proof. Let S be strongly inductive and $\mathcal{S}$ the collection of all nonempty subsets of S. Let $\psi : \mathcal{S} \to S$ be a function satisfying $\psi(A) \in A$ for all $A \in \mathcal{S}$. Define $\varphi : S \to S$ by $\varphi(x) =$

$$\begin{cases} x & \text{if } x \text{ is a maximal element in } S \\ \psi(\{z \mid z > x\}) & \text{if } x \text{ is not a maximal element in } S. \end{cases}$$

Then $\varphi(x) \geq x$ for all $x \in S$. Hence by Lemma A.10, there is an x_0 in S such that $\varphi(x_0) = x_0$. Then x_0 is a maximal element of S. Therefore, every nonempty strongly inductive poset has a maximal element. By Proposition A.3, Zorn's Lemma holds. □

Definition A.12. Two sets A and B are said to have the same *cardinality* if there exists a bijection $f : A \to B$. If there exists an injection $f : A \to B$, we write $|A| \leq |B|$ and write $|A| < |B|$ if, in addition, there exists no bijection $g : A \to B$.

Theorem A.13. (Schroeder-Bernstein Theorem) *Let X and Y be two sets. Then there exists a bijection between X and a subset of Y or a bijection between Y and a subset of X. In particular, if $|X| \leq |Y|$ and $|Y| \leq |X|$, then $|X| = |Y|$.*

Proof. Let X and Y be sets. Let Z be the subset of the power set, i.e., the set of all the subsets, of $X \times Y$ consisting of all $C \subset X \times Y$ satisfying:

(i) For each $x \in X$, there exists at most one element $y \in Y$ such that $(x, y) \in C$

(ii) For each $y \in Y$, there exists at most one element $x \in X$ such that $(x, y) \in C$.

This set is not empty by Proposition A.7. Order Z by inclusion. Then Zorn's Lemma gives a maximal element C in Z. Such an element either defines an injection $X \to Y$ if it satisfies (i) or an injection $Y \to X$ if it satisfies (ii). □

Corollary A.14. *Any two bases of a vector space have the same cardinality.*

Remarks A.15. In fact, the Schroeder-Bernstein Theorem is equivalent to the Axiom of Choice.

APPENDIX B

Bertand's Hypothesis

In this appendix, we give a full prove Bertrand's Hypothesis as well as giving Chebyshev's approximation of the Prime Number Theorem as indicted in Section 3. As in that section, $[x]$, $x \in \mathbb{R}$, is the greatest integer in x. As mentioned in Section 3, the proof relies on the analysis of the binomial coefficient $\binom{2n}{n}$, Euclid's Lemma 4.11, and the Fundamental Theorem of Arithmetic 4.16, i.e, Section 4.

If $n > 0$ is an integer and $p > 0$ is a prime integer, we write $p^r \,||n$ for $r \geq 0$ in $\mathbb{Z}$ if $p^r \mid n$ and $p^{r+1} \nmid n$.

Theorem B.1. (Bertrand's Hypothesis) *Let n be a positive integer. Then there exists a prime integer p satisfying $n < p \leq 2n$. [If $n > 1$, then for this prime $n < p < 2n$.]*

The proof will begin with a number of computations. Throughout p will denote a positive prime. In particular, $p \leq x$ will mean all positive primes less than a real number x.

Lemma B.2. *For all real numbers y, $0 \leq [2y] - 2[y] \leq 1$.*

Proof. From Properties 3.2, we have $[x] \leq x \leq [x] + 1$ for all $x \in \mathbb{R}$. Letting $x = y$ and multiplying this by 2 gives $2[y] \leq 2y < 2[y] + 2$ and setting $x = 2y$ gives $[2y] \leq 2y, 2[y] + 1$ for all $y \in \mathbb{R}$. Subtracting the second equation from the first yields $-1 < [2y] - 2[y] < 2$ in $\mathbb{Z}$. In particular, $0 \leq [2y] - 2[y] < 2$ as needed. $\square$

Lemma B.3. *Suppose that $n > 1$ is an integer. Then $\binom{2n}{n}$ is the largest of the $2n + 1$ binomial coefficients in $(1 + 1)^n$ and $\binom{2n}{n} \geq 2^n$.*

Proof. This follows from the following computations:

$$\binom{2n}{n} = \frac{2n \cdot (2n-1) \cdots (n+1)}{n \cdot (n-1) \cdots 1} = 2\big(2 + \frac{1}{n-1}\big) \cdots \big(2 + \frac{n-1}{1}\big) \geq 2^n;$$

and, as $\binom{2n}{i} = \binom{2n}{n-i}$,

$$\binom{2n}{n} \geq \frac{2n \cdots i}{(n-i) \cdots 1} = \binom{2n}{i}.$$

for $i \leq n$, by pairing the appropriate terms in the numerator and denominator. $\square$

Lemma B.4. *Let n be a positive integer. Then $\prod_{p \leq n} p < 4^n$.*

Proof. We prove this by the Second Principle of Finite Induction on n. We may assume that $n > 2$, since the cases $n = 1, 2$ are immediate.

Case 1. n is even.

As $n > 2$, n is not a prime, so $\prod_{p \leq n} p < \prod_{p \leq n-1} p < 4^{n-1} < 4^n$, by induction.

Case 2. $n = 2m + 1$ is odd.

Let p be any prime satisfying $m+2 \leq p \leq 2m+1$. Then p and i are relatively prime for all i with $1 \leq i \leq m+1$. Hence p and $(m+1)!m!$ are also relatively prime, i.e., $p \not| \, (m+1)!m!$. For every such prime p, we have $p \mid (2m+1)!$, so $p \mid \dfrac{(2m+1)!}{(m+1)!m!} = \dbinom{2m+1}{m+1}$ in $\mathbb{Z}$. It follows by induction that

$$\prod_{p \leq n} p = \prod_{p \leq m+1} p \prod_{m+2 < p \leq n} p < 4^{m+1} \binom{2m+1}{m+1}.$$

As $2m+1$ is odd, in the binomial expansion of $(1+1)^{m+1}$, both $\binom{2m+1}{m}$ and $\binom{2m+1}{m+1}$ occur and are equal. So $2\dbinom{2m+1}{m+1} \leq (1+1)^{2m+1} = 2^{2m+1}$. Therefore, $\dbinom{2m+1}{m+1} \leq 2^{2m} = 4^m$ and

$$\prod_{p \leq n} p < 4^{m+1} \binom{2m+1}{m+1} \leq 4^{m+1} 4^m = 4^{2m+1}$$

as needed. □

Lemma B.5. *Let $n > 3$ be an integer and p a prime satisfying $\frac{2}{3}n < p \leq n$. Then $p \not| \dbinom{2n}{n}$.*

Proof. Since $2n < 3p$ and $2 < p \leq n$, we have $2p \leq 2n < 3p$. Hence $2 \leq \dfrac{2n}{p} < 3$ and $\dfrac{2}{p} \leq \dfrac{2n}{p^2} < \dfrac{3}{p} \leq 1$ in $\mathbb{Q}$. Thus $\displaystyle\sum_{n=1}^{\infty} [\frac{2n}{p^i}] = 2$. Therefore, $p^2 \,||\, (2n)!$.

Now suppose that $2p \leq n$. Then $p \leq \dfrac{n}{2} < \dfrac{2}{3}n < p$, a contradiction. It follows that $p \leq n \leq 2p$ and $\dfrac{1}{p} \leq \dfrac{n}{p^2} < \dfrac{2}{p} < 1$. Therefore, $\displaystyle\sum_{n=1}^{\infty} [\frac{n}{p^i}] = 1$ and $p \,||\, n!$. By Theorem 3.3 $p^2 \,||\, (2n)!$ and $p \mid \,||\, n!$. Consequently, $p \not| \dbinom{2n}{n} = \dfrac{(2n)!}{n!n!}$ by the Fundamental Theorem of Arithmetic **??**. □

Lemma B.6. *Let $n > 1$ and $p > 2$ be a prime. Set*

$$\mu_p := \sum_{i=1}^{\infty} [\frac{2n}{p^i}] - 2\sum_{i=1}^{\infty} [\frac{n}{p^i}].$$

Then the following are true:

(*a*) $\displaystyle\binom{2n}{n} = \prod_{p \le 2n} p^{\mu_p}$.

(*b*) *$p^{\mu_p} \le 2n$. In particular, if $\mu_p > 1$, then $p \le \sqrt{2n}$.*

(*c*) *Let $n \ge 5$. Then $\sqrt{2n} < \frac{2}{3}n$. If $\sqrt{2n} < p < \frac{2}{3}n$, then $\mu_p \le 1$.*

(*d*) *If $n \ge 3$ and $\frac{2}{3}n < p < n$, then $\mu_p = 0$.*

PROOF. For each prime p, let

$$e_p = \sum_{i=1}^{\infty}[\frac{2n}{p^i}] \quad \text{and} \quad f_p = \sum_{i=1}^{\infty}[\frac{n}{p^i}].$$

(*a*): By Theorem 3.3, we have $p^{e_p} \,||\, (2n)!$ and $p^{f_p} \,||\, n!$. Therefore, $\mu_p = e_p - 2f_p$.

(*b*): Choose r_p to be the unique integer satisfying $p^{r_p} \le 2n < p^{r_p+1}$. By Lemma B.2, we have

$$\mu_p = \sum_{i=1}^{\infty}[\frac{2n}{p^i}] - 2\sum_{i=1}^{\infty}[\frac{n}{p^i}] = (\sum_{i=1}^{r_p}[\frac{2n}{p^i}] - [\frac{n}{p^i}]) \le \sum_{i=1}^{r_p} 1 = r_p.$$

Therefore, $p^{\mu_p} \le p^{r_p} \le 2n$.

(*c*), (*d*): Suppose that $\sqrt{2n} < p$, so $2n < p^2$. By (*b*), we have $p^{\mu_p} \le 2n < p^2$. Hence $\mu_p \le 1$ that gives (*c*) as the first statement is clear. If $\frac{2}{3}n \le p$, then by Lemma B.5 and (*a*), we have $\mu_p = 0$, which gives (*d*). □

We shall need the following corollary that follows from our proof of Lemma B.6, when we prove Chebyshev's Theorem below.

Corollary B.7. *Let $n > 1$ an integer and $p > 2$ a prime. If r_p to be the unique integer such that $p^{r_p} \le 2n < p^{r_p+1}$, then*

(*a*) $\mu_p \le r_p$.

(*b*) $\displaystyle\binom{2n}{n} \mid \prod_{p \le 2n} p^{r_p}$.

(*c*) $\displaystyle\prod_{n < p \le 2n} p \mid \binom{2n}{n}$.

(*d*) $\displaystyle\prod_{n < p \le 2n} p \le \binom{2n}{n} \le \prod_{p \le 2n} 2n$.

We now proceed to the proof of Bertrand's Hypothesis.

PROOF. The primes 2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 557 are each less than two times the previous. Therefore, the result holds for $n \le 512$. We proceed by induction on n.

Set $n > 128$ and suppose that there exists no prime

(*) $$p \text{ satisfying } n < p \le 2n.$$

Using Lemma B.5, Lemma B.6, and (*), we have

$$(\dagger)\qquad \begin{aligned}\binom{2n}{n} &= \prod_{p\le 2n} p^{\mu_p} = \prod_{p\le n} p^{\mu_p} = \prod_{p\le \frac{2}{3}n} p^{\mu_p}\\ &\le \prod_{p\le\sqrt{2n}} p^{\mu_p} \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p^{\mu_p} \le \prod_{p\le\sqrt{2n}} p^{\mu_p} \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p\\ &\le \prod_{p\le\sqrt{2n}} 2n \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p.\end{aligned}$$

Clearly, $\pi(x) := |\{p \mid p \text{ a prime } p \le x\}| \le \dfrac{x+1}{2}$. So by Lemma B.4 and Lemma B.5, we have

$$\binom{2n}{n} \le \prod_{p\le\sqrt{2n}} 2n \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p \le (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p \le (2n)^{\sqrt{2n}} 4^{\frac{2}{3}n}.$$

By Lemma B.4, we have $4^n < (2n+1)\dbinom{2n}{n}$. Therefore, $\dfrac{4^n}{2n+1} \le \dbinom{2n}{n}$.

Thus,

$$\frac{4^n}{2n+1} \le (2n)^{\sqrt{2n}} 4^{\frac{2n}{3}n}.$$

Since $2n+1 < 4n^2$ for positive n, we have by $(\dagger)$,

$$4^n \le (2n)^{\sqrt{2n}+2} 4^{\frac{2n}{3}} \qquad i.e., \qquad 4^{\frac{n}{3}} \le (2n)^{\sqrt{2n}+2}.$$

Taking log's then gives $\dfrac{n\log 4}{3} \le (\sqrt{2n}+2)\log 2n$. As $f(x) = \sqrt{x}\log 2 - \sqrt{2}\log 2x$ is an increasing function and positive for $x = 512$, this cannot hold. $\square$

There have been increasing generalizations of Bertrand's Hypothesis. For example, Schoefield showed that for $n \ge 2010760$, there always exists a prime p satisfying $n < p \le (1 + \frac{1}{16597})n$. Dusart, in a series of papers, has shown this can be improved to for all $x > 468991632$ in $\mathbb{R}$, there exists a prime p satisfying $x < p \le (1 + \frac{1}{5000\log^2 x})x$ while in 2001, Baker, Harman, and Pitz showed that for sufficiently large $x \in \mathbb{R}$, there exists a prime p satisfying $x - x^{0.525} < p \le x$. Dudek showed if the Riemann Hypothesis is true, then there always exists a prime p satisfying $x - \frac{4}{\pi}\sqrt{x}\log x < p < x$ for $x > 2$ in $\mathbb{R}$.

We turn to Chebyshev's Theorem approximating $\pi(x) := |\{p \mid p \text{ a prime and } p \le x\}|$. We use Corollary B.7 to get lower upper bounds and on $\pi(2n)$

Lemma B.8. *Let* $n > 1$. *Then*

(1) $\pi(2n) - \pi(n) \le \dfrac{2n\log 2}{\log n}$.

(2) $\pi(2n) \ge \dfrac{2\log 2}{\log 2n}$.

PROOF. It follows from Corollary B.7 that

$$n^{\pi(2n)-\pi(n)} \leq \prod_{n<p\leq 2n} n \leq \prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq \prod_{p\leq 2n} 2n = (2n)^{\pi(2n)}. \tag{*}$$

(1): Since $\binom{2n}{n}$ is just one term of the binomial expansion of $(1+1)^{2n}$, we have

$$n^{\pi(2n)-\pi(n)} \leq \binom{2n}{n} \leq (1+1)^{2n} = 2^{2n}.$$

Taking log and dividing by $\log n$, we obtain (1).

(2): By Lemma B.3, we see that

$$2^n < \binom{2n}{n} \leq (2n)^{\pi(2n)}.$$

Taking log and dividing by $\log 2n$, we get $\pi(2n) \geq \dfrac{n\log 2}{\log 2n}$ as needed. □

Theorem B.9. (Chebyshev) *There exists positive real numbers a and b satisfying*

$$a\,\frac{x}{\log x} < \pi(x) < b\,\frac{x}{\log x} \qquad \textit{for all } x \geq 2 \textit{ in } \mathbb{R}.$$

PROOF. We prove each inequality separately.

Claim. There exists such an $a > 0$:

Choose the even integer $2n$ to be the largest integer that satisfies $2n + 2 > x \geq 2n$ with $n \geq 1$. Using Lemma B.8(2) and the choice of $2n$, we have

$$\begin{aligned}\pi(x) \geq \pi(2n) &\geq \frac{n\log 2}{\log 2n} \geq n\,\frac{\log 2}{\log x} \geq \frac{n+1}{2}\,\frac{\log 2}{\log x}\\ &= \frac{2n+2}{4}\,\frac{\log 2}{\log x} > \frac{x}{4}\,\frac{\log 2}{\log x} = \frac{\log 2}{4}\,\frac{x}{\log x}.\end{aligned}$$

So $a = \dfrac{\log 2}{4}$ works.

Claim. There exists such a $b > 0$:

We first show that

$$\frac{\pi(2^{2m})}{2^{2m}} < \frac{4}{m} \text{ if } m \geq 1: \tag{B.10}$$

Let $r \geq 3$. Therefore, we can apply Lemma B.8(1) with $n = 2^{r-1}$. In particular,

$$\pi(2^r) - \pi(2^{r-1}) \leq \frac{2^r \log 2}{\log 2^{r-1}} = \frac{2^r \log 2}{(r-1)\log 2} = \frac{2^r}{r-1}. \tag{i}$$

We look at the telescoping series $\sum_{r\geq 2}^{2m}\bigl(\pi(2^r)-\pi(2^{r-1})\bigr)$. As $\pi(2^2)<2^2$, using (i), we see for $m\geq 2$, that

$$\pi(2^{2m})=\pi(2^{2m})-\pi(2^2)+\pi(2^2)=\pi(2^{2m})-\pi(2^2)+4=\sum_{r=3}^{2^m}\bigl(\pi(2^r)-\pi(2^{r-1})\bigr)+4$$

$$\text{(ii)}\qquad \leq\sum_{r=3}^{2m}\frac{2^r}{r-1}+4=\sum_{r=3}^{2m}\frac{2^r}{r-1}+\frac{2^2}{2-1}=\sum_{r=2}^{2m}\frac{2^r}{r-1}$$

$$=\sum_{r=2}^{m}\frac{2^r}{r-1}+\sum_{r=m+1}^{2m}\frac{2^r}{r-1}.$$

We estimate each of these final two sums.
Since $\sum_{j=1}^{k}2^j=2^{k+1}-1<2^k$, we have

$$\sum_{r=2}^{m}\frac{2^r}{r-1}\leq\sum_{r=2}^{m}\frac{2^r}{1}<2^{m+1}$$

and

$$\sum_{r=m+1}^{2m}\frac{2^r}{r-1}\leq\sum_{r=m+1}^{2m}\frac{2^r}{m}=\frac{1}{m}\sum_{r=m+1}^{2m}2^r<\frac{2^{2m+1}}{m}.$$

Plugging these into (ii) yields

$$\pi(2^{2m})\leq\sum_{r=2}^{m}\frac{2^r}{r-1}+\sum_{r=m+1}^{2m}\frac{2^r}{r-1}<2^{m+1}+\frac{2^{2m+1}}{m}$$

if $m\geq 2$. Now for any $m\geq 1$, we have $m<2^m$, so $2^{m+1}m<2^{m+1}2^m=2^{2m+1}$, hence $2^m+1<\frac{2^{2m+1}}{m}$. Therefore,

$$\pi(2^{2m})<\frac{2^{m+1}}{m}+\frac{2^{2m+1}}{m}<\frac{2m+1}{m}+\frac{2^{2m+1}}{m}=\frac{2}{m}2^{2m+1}=\frac{4}{m}2^{2m}$$

if $m\geq 2$. Checking this is also true for $m=1$ establishes Equation B.10.

Now let $x\geq 2$ in $\mathbb{R}$. Choose the unique $m\geq 1$ satisfying $2^{2m-2}<x\leq 2^{2m}$ Then by Equation B.10, we have

$$\frac{\pi(x)}{x}\leq\frac{\pi(2^{2m})}{2^{2m-2}}=4\frac{\pi(2^{2m})}{2^{2m}}\leq 4\cdot\frac{4}{m}=16\cdot\frac{1}{m}.$$

Since $x\leq 2^{2m}$, we have $\log x\leq 2m\log 2$, i.e.,

$$\frac{\pi(x)}{x}<16\cdot\frac{1}{m}\leq\frac{32\log 2}{\log x}$$

and $\pi(x)<b\,\dfrac{x}{\log x}$ with $b=32\,\log 2$. □

Remark B.11. The proof shows that $a=\frac{\log 2}{4}\sim .1733$ and $b=32\log 2\sim 22.1807$. Chebyshev's original proof showed $a>.92$ and $b<1.105$.

Corollary B.12. *Let p_r denote the rth positive prime. Then there exist positive real constants c, d satisfying $c\,r\log 2 < p_r < d\,r\log 2$ for all $r \geq 2$.*

PROOF. Let $x = p_r$ in Chebyshev's Theorem B.9. So

$$a\,\frac{p_r}{\log p_r} < r < b\,\frac{p_r}{\log p_r}. \tag{*}$$

Again, we look at each inequality separately.

Claim. Such a $c > 0$ exists:

As $p_r > r$, we have from the second inequality in (*) that $p_r > \frac{1}{b} r\log p_r$. So $c = \frac{1}{b}$ works.

Claim. Such a $d > 0$ exists:

From the first inequality in (*), we have $\frac{r\log p_r}{p_r} > a$. As there exist infinitely many primes, $p_r \to \infty$ as $r \to \infty$. We also know that $\frac{\log x}{\sqrt{x}} \to 0$ as $x \to \infty$. Therefore, there exists an integer $N > 0$ satisfying $\frac{\log p_r}{\sqrt{p_r}} < a$ for all $r \geq N$. So by the second inequality in (*),

$$\frac{r\log p_r}{p_r} > a > \frac{\log p_r}{\sqrt{p_r}} \text{ for all } r \geq N. \tag{†}$$

It follows that $r > \sqrt{p_r}$ for all $r \geq N$. Taking the log, we have $2\log r > \log p_r$. So using this and (†), we see that $p_r < \frac{2}{a} r\log r$ if $r \geq N$. Then $p_r < d\,r\log r$ if $r \geq 2$ for $d = \max\{\frac{2}{a}, \frac{p_2}{2\log 2}, \ldots, \frac{p_{N-1}}{(N-1)\log p_{N-1}}\}$. □

Corollary B.13. *There exists a positive real number k satisfying $\sum_{2<p\leq x} \frac{1}{p} < k\log\log x$ if $x \geq 3$.*

PROOF. In the notation of the previous corollary, $c\,r\log r < p_r$ for all $r \geq 2$. Therefore,

$$\begin{aligned}
\sum_{2<p\leq x} \frac{1}{p} &< \sum_{r=2}^{[x]} \frac{1}{cr\log r} = \frac{1}{c}\Big(\frac{1}{2\log 2} + \sum_{r=3}^{[x]} \frac{1}{r\log r}\Big) \\
&= \frac{1}{c}\Big(\frac{1}{2\log 2} + \sum_{r=3}^{[x]} \int_{r-1}^{r} \frac{dt}{r\log r}\Big) \leq \frac{1}{c}\Big(\frac{1}{2\log 2} + \sum_{r=3}^{[x]} \int_{r-1}^{r} \frac{dt}{t\log t}\Big) < \frac{1}{2c\log 2} + \frac{1}{c}\int_3^x \frac{dt}{t\log t} \\
&= \frac{1}{2c\log 2} + \frac{1}{c}\log\log x - \frac{1}{c}\log\log 2.
\end{aligned}$$

As $\log\log x \to \infty$ as $x \to \infty$, there exists a $k > 0$ as needed. □

Remark B.14. In fact, one can show that

$$\sum_{p\leq x} \frac{1}{p} \sim \log\log x \quad \textit{i.e.,} \lim_{x\to\infty} \frac{\sum_{p\leq x}\frac{1}{p}}{\log\log x} = 1.$$

Exercises B.15. 1. Show every integer $n > 6$ is a sum of distinct primes.

2. Show if a finite set of consecutive positive integers contains a prime, then one of the integers is relatively prime to all the others.
3. Show if $n > 1$, there exists a prime p satisfying $p \,||\, n$.

APPENDIX C

Matrix Representations

We review the notion of **matrix representation** of a linear transformation relative to a pair of ordered bases and what happens when we change bases. We do this in somewhat more generality then in linear algebra. All proofs, however, are the same, hence omitted. Throughout R will be a commutative ring. [So if R is a field, and R-module is just a vector space over R and an R-homomorphism is just a linear transformation.]

Setup.

Let R be a commutative ring.
V and W are free R-modules of rank n and m respectively.

$\mathcal{B} := \{v_1, ..., v_n\}$ is an ordered basis for V.
$\mathcal{B}'$ is a second ordered basis of V.

$\mathcal{C} := \{w_1, ..., w_m\}$ is an ordered basis for W.
$\mathcal{C}'$ is a second ordered basis of W.

$T : V \to W$ is an R-homomorphism.

Let v be an element in the free R-module V. Then there exist unique scalars $\alpha_1, ..., \alpha_n$ in R such that

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

The scalars $\alpha_1, ..., \alpha_n$ are called the *coordinates* of v relative to the (ordered) basis $\mathcal{B}$.

Let

$$[v]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

the *coordinate matrix* of v relative to the basis $\mathcal{B}$ and

$$V_{\mathcal{B}} := \{\, [v]_{\mathcal{B}} \mid v \in V\} = R^{n\times 1}.$$

We have an R-isomorphism

$$V \to V_{\mathcal{B}} \text{ given by } v \mapsto [v]_{\mathcal{B}}.$$

Similarly, if w is a vector in the free R-module W, then there exist unique scalars, $\beta_1, ..., \beta_m$, in R such that

$$w = \beta_1 w_1 + \cdots + \beta_m w_m.$$

Let

$$[w]_{\mathcal{C}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

the *coordinate matrix* of w relative to the basis $\mathcal{C}$ and

$$W_{\mathcal{C}} := \{\, [w]_{\mathcal{C}} \mid w \in W\} = R^{m\times 1}.$$

Examples C.1. 1.

$$[v_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \dots\ , [v_n]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

2. Suppose that $n = 3$. Then

$$[2v_1 - 3v_3]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix} \qquad [7v_2 - v_3]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 7 \\ -1 \end{pmatrix}.$$

We turn to the R-homomorphism $T : V \to W$. By the Universal Property of Free Modules, there exists a unique matrix

$$[T]_{\mathcal{B},\mathcal{C}} \in R^{m\times n}$$

called the *matrix representation* of T relative to the ordered bases $\mathcal{B}, \mathcal{C}$ that satisfies

$$[T]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}} \quad \forall v \in V.$$

If $V = W$ and $\mathcal{B} = \mathcal{C}$, we let $[T]_{\mathcal{B}} = [T]_{\mathcal{B},\mathcal{C}}$.

The matrix $[T]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}}$ is computed as follows. Write Tv_j in the $\mathcal{C}$ basis. Then the coordinate matrix of Tv_j relative to the $\mathcal{C}$ basis is the jth column of the matrix $[T]_{\mathcal{B},\mathcal{C}}$, i.e., if

$$Tv_j = \beta_{1j}w_1 + \cdots + \beta_{mj}w_m \text{ then } [T(v_j)]_{\mathcal{C}} = \begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{mj} \end{pmatrix}$$

and this is the jth column of $[T]_{\mathcal{B},\mathcal{C}}$. It is convenient' to write

Definition C.2. A matrix of the form $[1_V]_{\mathcal{B},\mathcal{B}'}$ is called a *change of basis matrix*.

It arises by writing the elements in the $\mathcal{B}$ basis in terms of the elements in the $\mathcal{B}'$ basis.

The main results are:

Proposition C.3. *Let V be an R-free module of rank n. Let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases for V. Then $[1_V]_{\mathcal{B},\mathcal{B}'}$ is an invertible matrix and*

$$[1_V]^{-1}_{\mathcal{B},\mathcal{B}'} = [1_V]_{\mathcal{B}',\mathcal{B}}$$

Remarks C.4. It can be shown that

$$\mathrm{GL}_n R = \{[T]_{\mathcal{B},\mathcal{C}} \,|\, T \text{ an isomorphism, } \mathcal{B},\mathcal{C} \text{ bases for } V\}.$$

Since R is commutative, determinants exist. It can also be shown that

$$\mathrm{GL}_n R = \{A \in \mathbf{M}_n R \,|\, \det(A) \in R^\times\}$$

Theorem C.5. *Let V, W, and X be finitely generated free R-modules. Let $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ be ordered bases for V, W, and X, respectively. Let $T : V \to W$ and $S : W \to X$ be R-homomorphisms. Then*

$$[S \circ T]_{\mathcal{B},\mathcal{D}} = [S]_{\mathcal{C},\mathcal{D}}[T]_{\mathcal{B},\mathcal{C}}$$

where the right hand side is matrix multiplication.

Theorem C.6. *Let V be a finitely generated free R-module and $\mathcal{B}$ an ordered basis for V. Then*

$$\mathrm{End}_R V \to \mathbb{M}_n R \text{ via } T \mapsto [T]_{\mathcal{B}}$$

is a ring isomorphism and induces a group isomorphism

$$\mathrm{Aut}_R V \to \mathrm{GL}_n R.$$

Theorem C.7. (Change of Basis Theorem.) *Let V and W be finitely generated R-free modules. Let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases for V and $\mathcal{C}$ and $\mathcal{C}'$ be ordered basis for W. Let $T : V \to W$ be an R-homomorphism. Then*

$$[T]_{\mathcal{B}',\mathcal{C}'} = [1_V]_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B},\mathcal{C}}[1_V]_{\mathcal{B}',\mathcal{B}} = [1_V]_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B},\mathcal{C}}[1_V]^{-1}_{\mathcal{B},\mathcal{B}'}.$$

The Change of Basis Theorem states that the following diagram commutes

$$\begin{tikzcd}
V_{\mathcal{B}} \arrow[r, "{[T]_{\mathcal{B},\mathcal{C}}}"] \arrow[d, "{[1_V]_{\mathcal{B},\mathcal{B}'}}"'] & W_{\mathcal{C}} \arrow[d, "{[1_V]_{\mathcal{C},\mathcal{C}'}}"] \\
V_{\mathcal{B}'} \arrow[r, "{[T]_{\mathcal{B}',\mathcal{C}'}}"] & W_{\mathcal{C}'}
\end{tikzcd}$$

Note that the inverses of the change of bases matrices go in the reverse direction. One can fill in more of the diagram. For example, the maps from the diagonals can also be read off, e.g.,

$$[T]_{\mathcal{B},\mathcal{C}'} = [T]_{\mathcal{B}',\mathcal{C}'}[1_V]_{\mathcal{B},\mathcal{B}'}$$
$$[T]_{\mathcal{B}',\mathcal{C}} = [1_V]^{-1}_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B}',\mathcal{C}'} = [1_V]_{\mathcal{C}',\mathcal{C}}[T]_{\mathcal{B}',\mathcal{C}'}$$

Warning C.8. Usually T is not an isomorphism, so $[T]_{\mathcal{B},\mathcal{C}'}$ is not invertible. So you cannot reverse arrows having T in them. If T is an isomorphism, then the matrix representation of T^{-1} is the inverse of the corresponding matrix representation of T.

Historically, one defined equivalence relationships between matrices. We define these and tell how they are related to R-homomorphisms of finitely generated free R-modules.

Definition C.9. If $A, B \in \mathbb{M}_n R$, we say that they are *similar* and write $A \sim B$ if there is an invertible matrix $C \in \mathrm{GL}_n R$ such that $A = C^{-1}BC$.

Clearly, $\sim$ is an equivalence relation. An important problem is to find good representatives for the classes under this equivalence relation when R is a nice ring, e.g., a PID. This leads to the study of Rational Canonical Forms and Jordan Canonical Forms in linear algebra.

Theorem C.10. *Let $A, B \in \mathbb{M}_n R$. Then A is similar to B in $\mathbb{M}_n R$ if and only if there exist a free R-module V of rank n with bases $\mathcal{B}, \mathcal{B}'$ and $T \in \mathrm{End}_R V$ such that $A = [T]_{\mathcal{B}}$ and $B = [T]_{\mathcal{B}'}$.*

Definition C.11. If $A, B \in R^{m \times n}$, we say that they are *equivalent* write $A \simeq B$ if there is an invertible matrices $P \in \mathrm{GL}_m R$ and $Q \in \mathrm{GL}_n R$ such that $A = PBQ$.

Clearly, $\simeq$ is an equivalence relation. An important problem is to find good representatives for the classes under this equivalence relation when R is a nice ring, e.g., a PID. This leads to the study of Smith Normal Forms in linear algebra.

Theorem C.12. *Let $A, B \in R^{m \times n}$. Then $A \simeq B$ in $R^{m \times n}$ if and only if there exist free R-modules V, W of rank n, m, respectively, with bases $\mathcal{B}, \mathcal{B}'$ for V and bases $\mathcal{C}, \mathcal{C}'$ for W and an R-homomorphism $T : V \to W$ such that $A = [T]_{\mathcal{B},\mathcal{C}}$ and $B = [T]_{\mathcal{B}',\mathcal{C}'}$.*

APPENDIX D

Smith Normal Form over a Euclidean Ring

Let R be a domain and $A \in R^{m\times n}$. Recall that A is in *Smith Normal Form* if A is a matrix of the form

$$\begin{pmatrix} q_1 & 0 & \dots & & & & \\ 0 & q_2 & & & & & \\ \vdots & & \ddots & & & & \\ & & & q_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & \end{pmatrix}$$

with $q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r$ in R and $q_r \neq 0$, i.e., the diagonal entries of A are $q_1, \dots, q_r, 0, \dots, 0$ with $q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r$ in R and $q_r \neq 0$ and all entries off the diagonal are 0.

Let R be a domain and $A \in \mathbb{M}_n(R)$. We say that $A = (a_{ij})$ is an *elementary matrix* if

(i) *Type I:* if there exists $0 \neq \lambda \in R$ and $l \neq k$ such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ \lambda & \text{if } (i,j) = (k,l) \\ 0 & \text{otherwise} \end{cases}$$

(ii) *Type II:* if there exists $k \neq l$ such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \neq l \text{ and } i = j \neq k \\ 0 & \text{if } i = j = l \text{ or } i = j = k \\ 1 & \text{if } (k,l) = (i,j) \text{ or } (k,l) = (j,i) \\ 0 & \text{otherwise} \end{cases}$$

(iii) *Type III:* if there exists a unit u in R and l such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \neq l \\ u & \text{if } i = j = l \\ 0 & \text{otherwise} \end{cases}$$

Remarks D.1. 1. Let $A \in R^{m\times n}$. Multiplying A on the left (respectively right) by a suitable size (*i*) Type I is adding a multiple of a row (respectively column) of A to another row (respectively column) of A.

(*ii*) Type II is interchanging two rows (respectively columns) of A.

(*iii*) Type III is multiplying a row (respectively column) of A by a unit.

2. All elementary matrices are invertible.

Recall if R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$.

Theorem D.2. *Let R be a euclidean ring, $A = (a_{ij}) \in R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form. Moreover, there exist matrices $P \in GL_m(R)$ and $Q \in \mathrm{GL}_n(R)$, each a product of matrices of Type I, Type II, and Type III, such that PAQ is in Smith Normal Form.*

Proof. The proof will, in fact, be an algorithm to find a Smith Normal Form for A. Let δ be the euclidean function on R. If $A = 0$ there is nothing to do, so assume that $A \neq 0$.

Step 1. Choose $a = a_{ij} \neq 0$ such that $\delta(a)$ is minimal among all the $\delta(a_{lk})$, $a_{lk} \neq 0$. Put a in the $(1, 1)$ spot using matrices of Type II. In particular, we may assume that $a = a_{11}$. [If a is a unit in R, use a Type III matrix to make $a = 1$.]

Step 2. If $a \mid a_{ij}$ for all i and j, use Type I matrices to transform A into a matrix of the form

$$\begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

Note that a divides every entry of A_1 [Check]. If a non-zero entry of A_1 has smaller δ value, $\delta(a)$ is not minimal so go back to Step 1.

[As δa is a non-negative integer this cannot happen infinitely often.]

If δa is still minimal, take A_1 and go back to Step 1.

[Note. If this occurs, by induction there exist matrices Q_1, P_1 such that $P_1A_1Q_1$ is in Smith Normal Form. Let

$$P = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & P_1 & \\ 0 & & & \end{pmatrix} \text{ and } Q = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & Q_1 & \\ 0 & & & \end{pmatrix}$$

then

$$P \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix} Q$$

is in Smith Normal Form.]

Step 3. Step 2 does not apply and there exists an entry $b = a_{ij}$ in either the first row or first column such that $a \nmid b$:

Write $b = qa + r$ in R with $r \neq 0$ and $\delta(r) < \delta(a)$. Use Type I matrices to change A into a matrix with r in it. Since $\delta(a)$ is not longer minimal, go back to Step 1.
[Since $\delta(a)$ is a non-negative integer and $\delta(r) < \delta(a)$, this must eventually stop.]

Step 4. Neither Step 2 nor Step 3 apply. Thus $a \mid a_{ij}$ whenever $i = 1$ or $j = 1$:

Use Type I matrices to convert A to

$$\begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

Now one of the following occurs.

(1) There exists a non-zero entry b in A_1 such that $\delta(b) < \delta(a)$. So $\delta(a)$ is no longer minimal. Go back to Step 1.

(2) $a \mid b$ for all entries b in A_1. This is impossible — You should have been in Step 2.
[No matter, take A_1 and go to Step 1.]

(3) There exists an entry b in A_1 such that $a \nmid b$:
Write $b = qa + r$ in R with $r \neq 0$ and $\delta(r) < \delta(a)$. Use Type I matrices to get b into the first column. (This does not change the $(1, 1)$ entry a.) Now use Type I matrices to change b to r. Since $\delta(a)$ is no longer minimal, go back to Step 1.

Clearly this algorithm yields a Smith Normal Form of A.

□

Remarks D.3. 1. If R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$. Compare this to change of basis in linear algebra of matrix representations of linear transformations. So the theorem says if R is a euclidean ring then any $m \times n$ matrix over R is equivalent to a matrix in Smith Normal Form.

2. Note that we do not really need Type III matrices in the above. We only used them to transform those diagonal entries q_i that were units into one.

3. We did not really need Step 2 as it is incorporated into Step 4, but it is useful to isolate it to complete the induction step of the proof.

As stated in Theorem 43.8 and proved, a Smith Normal Form of a matrix over a PID is unique up to units. We shall assume it here. Its proof depends only on an elementary determinant argument and could easily be done here, but we shall not repeat it in this appendix.

In Section 45, Smith Normal Forms are used in understanding parts of linear algebra. In this appendix, we shall prove (most) of Theorem 45.11 without the development needed in Section 45 used to prove it there. Rather we shall only use the Smith Normal Form of a matrix over the euclidean domain $F[t]$ with F a field and a generalization of the division algorithm for $F[t]$.

Let V be a finite dimensional vector space over a field F and $\mathcal{B}$ an ordered basis for V. Classifying linear operators $T : V \to V$ is equivalent to determining $[T]_{\mathcal{B}}$ up to matrix similarity. So we turn to the matrix formulation.

Let A_0 be a matrix in $\mathbb{M}_n(F)$ and $A = tI - A_0$, a matrix in $\mathbb{M}_n\big(F[t]\big)$. This matrix is called the *characteristic matrix of* A_0. As $\det A$ is the characteristic polynomial of A_0, hence nonzero, there is a Smith Normal Form of A of the form $\operatorname{diag}(1, \dots, 1, q_1, \dots, q_r)$ for some non-constant polynomials $q_1, \dots, q_r$ in $F[t]$ and some unique r. As the Smith Normal Form of A is unique up to units (by Theorem 43.8), using Type III operations, we may assume that the polynomials $q_1, \dots, q_r$ are all monic. This is will be called *the* Smith Normal Form of A and the unique monic polynomials $q_1, \dots, q_r$ will be called the *invariant factors* of A. If B_0 is another matrix in $\mathbb{M}_n(F)$ and $B = tI - B_0$, we want to show that A is equivalent to B in $\mathbb{M}_n\big(F[t]\big)$ if and only if A_0 is similar to B_0. As mentioned above this is also shown in Section 45, but our proof is not only more elementary but interesting in its own right.

We need a special case of the division algebra applied to matrices in $\mathbb{M}_n\big(F[t]\big)$. We know that the division algorithm says if $g \in F[t] \setminus \{0\}$ and $a \in F$, then $g = (t - a)h + r$ for some h in $F[t]$ and $r \in F$ — indeed $r = g(a)$ by the Remainder Theorem. We want a similar result for matrices in $\mathbb{M}_n\big(F[t]\big)$. However, dividing on the right and dividing on the left makes a difference, i.e., we may have $G = AQ_1 + R_1 = Q_2 A + R_2$, but $Q_1 \neq Q_2$ and $R_1 \neq R_2$ in the analogue of the above. However, the same algorithm in the commutative case can be mimicked as we shall now show.

We use the following notation:

Notation D.4. Let F be a field, $A_i \in \mathbb{M}_r(F)$, $i = 0, \dots, n$. The polynomial matrix

$$A_n t^n + A_{n-1} t^{n-1} + \cdots + A_0$$

will denote the matrix

$$A_n(t^n I) + A_{n-1}(t^{n-1} I) + \cdots + A_0$$

in $\mathbb{M}_n\big(F[t]\big)$ with t commuting with all elements of F and I the identity matrix. So if $A = (\alpha_{ij})$, then $At^n = (\alpha_{ij} t^n)$. Two polynomials matrices are the same if and only if their coefficients are equal, i.e., we are identifying $\mathbb{M}_n\big(F[t]\big)$ and $\big(\mathbb{M}_n(F)\big)[t]$.

Lemma D.5. *Let F be a field and A_0 an element in $\mathbb{M}_n\big(F\big)$. Set $A = tI - A_0$ in $\mathbb{M}_n\big(F[t]\big)$. Suppose that $P = P(t)$ is an element in $\mathbb{M}_n\big(F[t]\big)$. Then there exist matrices M and N in $\mathbb{M}_n\big(F[t]\big)$ and matrices R and S in $\mathbb{M}_n(F)$ satisfying*

(1) $P = AM + R$.
(2) $P = NA + Q$.

Proof. We show (1), the second being similar. Let

$$m = \max_{l,k}(\deg P_{lk}) \quad \text{and} \quad P_{ij} = \alpha_{ij} t^m + \ \text{ lower terms in } F[t]$$

for all appropriate l, k, i, j. In particular, we have

$$\alpha_{ij} = \begin{cases} \operatorname{lead} P_{ij} & if\ \deg P_{ij} = m \\ 0 & if\ \deg P_{ij} < m. \end{cases}$$

Let $(\alpha_{ij}) \in \mathbb{M}_n(F)$ and define a polynomial matrix

$$P_{m-1} = (\alpha_{ij})t^{m-1} = (\alpha_{ij}t^{m-1}) = (t^{m-1}\alpha_{ij}).$$

Every entry in

$$AP_{m-1} = (tI - A_0)(\alpha_{ij})t^{m-1} = (\alpha_{ij})t^m - A_0(\alpha_{ij})t^{m-1}$$

has degree at most m and the t^m-coefficient of $(AP_{m-1})_{ij}$ is α_{ij}.
Therefore, $P - AP_{m-1}$ has polynomial entries of degree at most $m-1$. Applying the same argument to $P - AP_{m-1}$ produces a matrix P_{m-2} in $\mathbb{M}_n\big(F[t]\big)$ with all the polynomial entries of $(P - AP_{m-1}) - AP_{m-2}$ having degree at most $m-2$. Continuing in this way, we construct matrices $P_{m-3}, \dots, P_0$ such that if $M := P_{m-1} + P_{m-2} + \cdots + P_0$, then $R := P - AM$ has only constant entries, i.e., $R \in \mathbb{M}_n(F)$. Therefore, $P = AM + R$ as needed. $\square$

We use the lemma to prove our theorem:

Theorem D.6. *Let F be a field and A_0, B_0 be elements in $\mathbb{M}_n(F)$. Set $A = tI - A_0$ and $B = tI - B_0$. Then A and B are equivalent in $\mathbb{M}_n\big(F(t]\big)$ if and only if A_0 and B_0 are similar in $\mathbb{M}_n(F)$.*

Proof. $(\Leftarrow)$: If $A_0 = PB_0P^{-1}$ for some $P \in \mathrm{GL}_n(F)$, then

$$P(tI - A_0)P^{-1} = PtIP^{-1} - PA_0P^{-1} = tI - B_0 = B.$$

So $B = PAP^{-1}$ and B is equivalent to A.
$(\Rightarrow)$: Suppose that there exist matrices P_1 and Q_1 in $\mathrm{GL}_n\big(F[t]\big)$, (each is a product of elementary matrices by Theorem D.2) satisfying

$$B = tI - B_0 = P_1AQ_1 = P_1(tI - A_0)Q_1.$$

By the Lemma D.5, we can write

(*i*) $P_1 = BP_2 + R$ with $P_2 \in \mathbb{M}_n\big(F[t]\big)$ and $R \in \mathbb{M}_n(F)$.
(*ii*) $Q_1 = Q_2B + S$ with $Q_2 \in \mathbb{M}_n\big(F[t]\big)$ and $S \in \mathbb{M}_n(F)$.

As $B = P_1AQ_1$ with $P_1, Q_1 \in \mathrm{GL}_n\big(F[t]\big)$, we also have

(*iii*) $P_1A = BQ_1^{-1}$.
(*iv*) $AQ_1 = P_1^{-1}B$.

Therefore, we have

$$\begin{aligned} B = P_1AQ_1 &\underset{(i)}{=} (BP_2 + R)AQ_1 = BP_2AQ_1 + RAQ_1 \\ &\underset{(iv)}{=} BP_2P_1^{-1}B + RAQ_1 = BP_2P_1^{-1}B + RA(Q_2B + S) \\ &= BP_2P_1^{-1}B + RAQ_2B + RAS, \end{aligned}$$

i.e., we have

(v) $B = BP_2P_1^{-1}B + RAQ_2B + RAS.$

We work on the middle term RAQ_2B of this equation. By (i), we have

$$R = P_1 - BP_2,$$

so plugging this in for RAQ_2B, yields

$$\begin{aligned} RAQ_2B &\underset{(i)}{=} (P_1 - BP_2)AQ_2B = P_1AQ_2B - BP_2AQ_2B \\ &\underset{(iii)}{=} BQ_1^{-1}Q_2B - BP_2AQ_2B = B[Q_1^{-1}Q_2 - P_2AQ_2]B, \end{aligned}$$

i.e., we have

(vi) $RAQ_2B = B[Q_1^{-1}Q_2 - P_2AQ_2]B.$

Plugging (vi) into (v) yields

$$\begin{aligned} B &\underset{(v)}{=} BP_2P_1^{-1}B + RAQ_2B + RAS \\ &\underset{(vi)}{=} BP_2P_1^{-1}B + B[Q_1^{-1}Q_2 - P_2AQ_2]B + RAS \\ &= B[P_2P_1^{-1} + Q_1^{-1}Q_2 - P_2AQ_2]B + RAS. \end{aligned}$$

Let

$$T = P_2P_1^{-1} + Q_1^{-1}Q_2 - P_2AQ_2.$$

So we have

(vii) $B = BTB + RAS.$

Next we look at the degree of the polynomial entries in these three matrices in (vii). Certainly,

$(viii)$ Every entry of $B = tI - B_0$ has degree at most one and every entry of $RAS = R(tI - A_0)$ has degree at most one.

So we must look at BTB. Write

$$T = T_mt^m + T_{m-1}t^{m-1} + \cdots + T_0 \quad \text{with} \quad T_0, \ldots, T_m \in \mathbb{M}_n(F).$$

Then

$$\begin{aligned} BTB &= (tI - B_0)(T_mt^m + \cdots + T_0)(tI - B_0) \\ &= T_mt^{m+2} + \quad \text{lower terms in } t. \end{aligned}$$

Comparing coefficients of the matrix polynomials in (vii) and using $(viii)$ shows that $T_m = 0$. It follows that $T = 0$. Thus (vii) becomes

$$\begin{aligned} tI - B_0 = B = BTB + RAS = RAS &= R(tI - A_0)S \\ &= RSt + RA_0S. \end{aligned} \tag{D.7}$$

Comparing coefficients of the polynomial matrices in equation (D.7) now implies that $I = RS$, i.e., $S = R^{-1}$, and $B_0 = RA_0S$. In particular,

$$B_0 = RA_0S = RA_0R^{-1}.$$

Hence B_0 is similar to A_0 as needed. □

We therefore have, as Smith Normal Form is unique,

Theorem D.8. *Let F be a field and A_0, B_0 be elements in $\mathbb{M}_n(F)$. Set $A = tI - A_0$ and $B = tI - B_0$. Then the following are equivalent.*

(1) A and B are equivalent in $\mathbb{M}_n\big(F[t]\big)$.
(2) A_0 and B_0 are similar in $\mathbb{M}_n(F)$.
(3) A and B have the same invariant factors.

Corollary D.9. *Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Then A is similar to A^t*

Proof. They have the same Smith Normal Form. □

Exercise D.10. Let R be a commutative ring. Let $E_n(R)$ be the subgroup of $\mathrm{GL}_n(R)$ generated by all matrices of the form $I + \lambda$ where λ is a matrix with precisely one non zero entry and this entry does not occur on the diagonal. Suppose that R is a euclidean ring. Show that $\mathrm{SL}_n(R) = E_n(R)$.

APPENDIX E

Symmetric Bilinear Forms

In this appendix, we establish Sylvestor's theorem about the signature of a symmetric matrix over the real numbers. We shall need its interpretation over a subfield F of the reals using the usual ordering $<$ of $\mathbb{R}$ restricted to F, i.e., we have a disjoint union $F = F^+ \cup \{0\} \cup F^-$ with $F^+ := \{x \in F \mid x > 0\}$ and $F^- := \{x \in F \mid x < 0\}$. The proof of Sylvestor's Theorem will then hold in the more general situation needed in Section 76.

We begin with a short discussion of symmetric bilinear spaces. For convenience, we restrict our attention to finite dimensional symmetric bilinear forms over a field of characteristic different from two.

Definition E.1. Let V be a finite dimensional vector space over a field F of characteristic different from two and $B : V \times V \to F$ a symmetric bilinear form, i.e., linear in each variable and $B(x, y) = B(y, x)$ for all $x, y \in V$. We call (V, B) a *symmetric bilinear space.* If $W \subset V$ is a subspace of V, then $B|_{W\times W} : W \times W \to F$ is a symmetric bilinear form, and we call $(W, B|_{W\times W})$ a (bilinear) subspace of (V, B). We say x, y in V are *orthogonal* (rel B) if $B(x, y) = 0$. We say two subspaces W_1, W_2 of V are *orthogonal* if $B(x, y) = 0$ for all $x \in W_1$, $y \in W_2$ and, if this is the case, we write $W_1 \perp W_2$. Note if this is the case, then $W_1 + W_2 = W_1 \oplus W_2$ if and only if $W_1 \cap W_2 = \{0\}$ (by dimension count). If $V = W_1 \oplus W_2$ and $W_1 \perp W_2$, then we write $(V, B) = (W_1, B|_{W_1\times W_1}) \perp (W_2, B|_{W_2\times W_2})$, and call it an *orthogonal decomposition* of (V, B).

Lemma E.2. *Let F be a field of characteristic different from two and (V, B) a finite dimensional symmetric bilinear space. If W is a subspace of V, then*

$$W^\perp := \{v \in V \mid B(v, w) = 0 \textit{ for all } w \in W\}$$

is a subspace of V and $(W^\perp, B|_{W^\perp\times W^\perp})$ is a subspace of (V, B). In particular, $(V, B) = (W, B|_{W\times W}) \perp (W^\perp, B|_{W^\perp\times W^\perp})$ if and only if $W \oplus W^\perp = V$.

Proof. If $v_1, v_2 \in W^\perp$ and $\alpha \in F$, then

$$B(\alpha v_1 + v_2, w) = \alpha B(v_1, w) + B(v_2, w) = 0$$

for all $w \in W$, i.e., $W^\perp$ is a subspace. The second statement is clear. □

Remark E.3. Let F be a field of characteristic different from two and (V, B) a finite dimensional symmetric bilinear space. Then

$$B(x, y) = \frac{1}{2}\big[(B(x + y, x + y) - B(x, x) - B(y, y)\big].$$

Since $2 \neq 0$ under the characteristic assumption on F, it follows that B is determined by the composition $q_B := B \circ \Delta : V \to V$, where $\Delta : V \to V \times V$ given by $\Delta(v) = (v, v)$ for all $v \in V$ is the diagonal map. The map q_B is called the *associated quadratic form* of B. In particular, if $B(x,y) \neq 0$, then by Remark E.3, $B(z,z) \neq 0$ for z at least one of $x, y, x+y$ since $2 \neq 0$. This is the major reason that we are assuming that the characteristic of F is not two.

Definition E.4. Let (V,B) be finite dimensional symmetric bilinear space. Let $\operatorname{rad}(B) := \{v \mid B(v,w) = 0 \text{ for all } w \in V\}$, called the *radical* of (V,B). So $\operatorname{rad}(B) = V^\perp$ and $B|_{\operatorname{rad}(B)\times\operatorname{rad}(B)} = 0$. If $\operatorname{rad}(B) = 0$, we say that (V,B) is *regular*.

Let (V,B) be a finite dimensional symmetric bilinear space over F, a field of characteristic different from two. Then for every $v \in V$, the map $B_v : V \to F$ defined by $x \mapsto B(v,x)$ is a linear functional, so defines a map $\widehat{B} : V \to V^*$ by $v \mapsto B_v$ with V^* the dual space of V. This map is checked to be an isomorphism if and only if V is regular.

Definition E.5. Let F be a field of characteristic different from two and $(V,B), (V',B')$ finite dimensional symmetric bilinear spaces over F. A vector space isomorphism $\varphi : V \to V'$ is called an *isometry* if $B'\big(\varphi(x), \varphi(y)\big) = B(x,y)$ for all $x, y \in V$. If φ is an isometry we write $(V,B) \cong (V',B')$.

Let F be a field of characteristic different from two and (V,B) a finite dimensional symmetric bilinear space. Let ${}^- : V \to V/\operatorname{rad}(B)\,(= \overline{V})$ be the canonical epimorphism. Define $\overline{B}$ to be the bilinear form on $\overline{V}$ determined by $\overline{B}(\overline{v_1}, \overline{v_2}) := B(v_1, v_2)$ for all $v_1, v_2 \in V$. Then $(\overline{V}, \overline{B})$ is a regular symmetric bilinear space, called the *induced bilinear space.*

Lemma E.6. *Let F be a field of characteristic different from two and (V,B) a finite dimensional symmetric bilinear space. Let W be any subspace of V satisfying $V = \operatorname{rad}(B) \oplus W$. Then $(W, B|_{W\times W})$ is regular and*

$$(V,B) = \big(\operatorname{rad}(B), 0\big) \perp (W, B|_{W\times W})$$

with $(W, B|_{W\times W}) \cong (\overline{V}, \overline{B})$, the form induced on $V/\operatorname{rad}(B)$. Moreover, $(W, B|_{W\times W})$ is unique up to isometry.

Proof. The first statement follows, as ${}^- : V \to \overline{V}$ induces a vector space isomorphism $V/W \to \overline{V}$. Since any isometry of symmetric bilinear spaces takes radicals to radicals, the uniqueness statement follows. $\square$

We leave the following lemma as an exercise.

Lemma E.7. *Let F be a field of characteristic different from two, (V,B) a finite dimensional symmetric bilinear over F, and $W \subset V$ a subspace. Define $f_W : V \to W^*$ by $v \mapsto (B|_v)|_W$. Then $W^\perp = \ker f_W$, i.e., we have an an exact sequence of vector spaces*

$$0 \ \to W^\perp \xrightarrow{\text{inc}} V \xrightarrow{f_W} W^*,$$

where inc *is the inclusion map. In particular,* $\dim W^\perp \geq \dim V - \dim W$*.*

Using the lemma, it is easy to determine when this is an equality.

Proposition E.8. *Let F be a field of characteristic different from two and (V,B) be a finite dimensional symmetric bilinear form. Let W be any subspace of V. Then the following are equivalent:*

(1) $W \cap \operatorname{rad}(B) = 0$.

(2) $f_W : V \to W^*$ *is surjective.*
(3) $\dim W^\perp = \dim V - \dim W$.

Proof. By linear algebra if $T : V \to W$ is a linear transformation, there is a dual map of dual spaces $T^t : W^* \to V^*$ where T^t, called the transpose of T is the linear transformation defined by $T^t(g) = g \circ T$. Moreover, applied to compositions, it takes exact sequences to exact sequences (with the maps reversed upon taking the dual sequence). As V is finite dimensional, $V = V^{**}$. It follows that (1) holds if and only if the transpose map $(f_W)^t : W \to V^*$ is injective if and only if the map $f_W : V \to W^*$ is surjective (why?) if and only if (3) holds. □

A key observation is:

Proposition E.9. *Let F be a field of characteristic different from two and (V, B) a finite dimensional symmetric bilinear space. Suppose that W is a subspace of V with $(W, B|_{W\times W})$ regular. Then $(V, B) = (W, B)|_{W\times W}) \perp (W^\perp, B|_{W^\perp \times W^\perp})$. In particular, if (V, B) is also regular, then so is $(W^\perp, B|_{W^\perp\times W^\perp})$.*

Proof. By Proposition E.8, $\dim W^\perp = \dim V - \dim W$, hence $V = W \oplus W^\perp$. The result follows. □

Corollary E.10. *Let F be a field of characteristic different from two and (V, B) a finite dimensional symmetric bilinear space. If $v \in V$ satisfies $B(v, v) \neq 0$, then $(V, B) = (Fv, B|_{Fv\times Fv}) \perp (Fv^\perp, B|_{(Fv)^\perp \times (Fv)^\perp})$.*

Theorem E.11. *Let F be a field of characteristic different from two and (V, B) a finite dimensional symmetric bilinear space. Then V has an orthogonal basis.*

Proof. Let $\dim V = n$. If $V = \operatorname{rad} V$ any basis works, so we may assume this is not so. In particular, there exists a vector $v \in V$ such that $B(v, v) \neq 0$ by Remark E.3. The result follows by induction on $\dim V$ in view of Corollary E.10. □

If we had not put the condition that the field is of characteristic not two, then an orthogonal basis may not exist for a finite dimensional symmetric bilinear space.

As with linear operators on a finite dimensional vector space, we have matrix representations of finite dimensional symmetric bilinear spaces (V, B). Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be an ordered basis for V. Then the symmetric matrix $\big(B(v_i, v_j)\big) \in \mathbb{M}_n(F)$ is called the *matrix representation* of (V, B) relative to the basis V and we denote it by $[B]_\mathcal{B}$. We have $B(v, w) = v^t B w$ for all $v, w \in V$ (writing w as a column matrix in the matrix equation) where v^t is the transpose of v. In particular,

$\mathcal{B}$ an orthogonal basis if and only if $[B]_\mathcal{B}$ is a diagonal matrix.

If $\mathcal{C}$ is another ordered basis for V, computation shows that

$$[1_V]^t_{\mathcal{B},\mathcal{C}}[B]_\mathcal{B}[1_V]_{\mathcal{B},\mathcal{C}} = [B]_\mathcal{C}.$$

Let F be a field of characteristic different from two and (V, B) and (V', B') two finite dimensional symmetric bilinear spaces with $\mathcal{B}$, and $\mathcal{B}'$ ordered bases for V and V', respectively. Then it is easy to see that $(V, B) \cong (V', B')$ if and only if $A^t[B]_\mathcal{B} A = [B']_{\mathcal{B}'}$ where A is the matrix representation of the isomorphism $T : V \to V'$ taking the basis $\mathcal{B}$ to $\mathcal{B}'$.

We turn to the special case of interest in this book. Let F be a subfield of $\mathbb{R}$, V a finite dimensional vector space of dimension n and $B : V \times V \to F$ a symmetric bilinear form. Let $<$ be the ordering on F induced by $<$ on $\mathbb{R}$, so if $P := \{x \in F \mid x > 0\}$ and $-P := \{x \in F \mid x < 0\}$, we have So $F = P \cup \{0\} \cup -P$ is a disjoint union. For example, if $F = \mathbb{R}$, then $P = \{x^2 \in F^\times \mid x \in F\}$.

Proposition E.12. *Let F be a subfield of $\mathbb{R}$, V a finite dimensional vector space of dimension n and $B : V \times V \to F$ a symmetric bilinear form. Let $<$ be the ordering on F induced by $<$ on $\mathbb{R}$ Then there exists an orthogonal decomposition $V = V_0 \oplus V_+ \oplus V_-$. with*

$$\begin{array}{lll} B(v,v) = 0 & \textit{for all} & v \in V_0 \\ B(v,v) > 0 & \textit{for all} & 0 \neq v \in V_+ \\ Bv,v) < 0 & \textit{for all} & 0 \neq v \in V_-. \end{array} \tag{E.13}$$

Moreover, the dimensions $\dim V_0$, $\dim V_+$, and $\dim V_-$ are independent of the orthogonal decomposition satisfying (E.13).

PROOF. We first show that there exists an orthogonal decomposition $V = V_0 \oplus V_+ \oplus V_-$ satisfying (E.13).

By Theorem E.11, we know that V has an orthogonal basis $\mathcal{B} = \{v_1, \dots, v_n\}$. Reordering this basis if necessary, we may assume that

$$\begin{aligned} B(v_i, v_i) &= 0, \quad i = 1, \dots, r \\ B(v_i, v_i) &> 0, \quad i = r+1, \dots, p \\ B(v_i, v_i) &< 0, \quad i = p, \dots, q. \end{aligned}$$

So $V_0 = \perp_{i=1}^{r} Fv_i$, $V_+ = \perp_{i=r+1}^{p} Fv_i$, $V_- = \perp_{i=p+1}^{q} Fv_i$ work.

Let $V = V_0 \oplus V_+ \oplus V_-$ satisfy (E.13) and $W \subset V$ be a subspace satisfying $B(v,v) > 0$ for all $0 \neq v \in W$. Then we establish the following:

Claim. $W + V_0 + V_- = W \oplus V_0 \oplus V_-$.

Let $w \in W$, $v_0 \in V_0$, $v_- \in V_-$ satisfy $w + v_0 + v_- = 0$. Then

$$\begin{aligned} 0 = B(w, w + v_0 + v_-) &= B(w,w) + B(w, v_0) + B(w, v_-) \\ &= B(w,w) + B(w, v_-) \\ 0 = B(v_-, w + v_0 + v_-) &= B(v_-, w) + B(v_-, v_0) + B(v_-, v_-) \\ &= B(v_-, w) + B(v_-, v_-). \end{aligned}$$

As B is symmetric, we have $B(w,w) = B(v_-, v_-)$. But $B(w,w) = 0$ if and only if $w = 0$ and $B(v_-, v_-) = 0$ if and only if $v_- = 0$. So $w = 0 = v_-$. Hence also $v_0 = 0$ and the claim is established.

The claim implies that $\dim W \leq \dim V_+$. In particular, if we have another orthogonal decomposition $V = V_0' \oplus V_+' \oplus V_-'$ satisfying (E.13), then $\dim V_+' \leq \dim V_+$. Applying the same argument shows that $\dim V_+ \leq \dim V_+'$, so $\dim V_+ = \dim V_+'$. An analogous argument shows that $\dim V_- = \dim V_-'$, hence also that $\dim V_0 = \dim V_0'$. $\square$

Remark E.14. In the above the subspace $V_0 = \operatorname{rad} V$, so is unique by Lemma E.6, but V_+ and V_- are not, although their dimensions are.

Definition E.15. Let (V, B) be a bilinear space over F, a subfield of $\mathbb{R}$ and $<$ an ordering on F arising from $\mathbb{R}$. Let $V = V_0 \oplus V_+ \oplus V_-$ be an orthogonal decomposition satisfying (E.13). Then $\dim V_+ - \dim V_-$ is called the *signature* of B. We denote the signature by $\operatorname{sgn}_P(B)$, where $P = \{x \in F \mid x > 0\}$.

In the above, if $V = \operatorname{rad}(B) \perp W$, then $(W, B|W \times W)$ is regular by Lemma E.6, unique up to isometry, so satisfies $\operatorname{sgn}_P(V) = \operatorname{sgn}_P(W)$.

Proposition E.12 implies

Corollary E.16. *Let (V, B) be a bilinear space over F, a subfield of $\mathbb{R}$ and $<$ an ordering on F arising from $\mathbb{R}$. Then the signature of B, $\operatorname{sgn}_P(B)$ is independent of orthogonal decomposition satisfying (E.13).*

If $F = \mathbb{R}$, then we classify isometry classes of finite dimensional real symmetric bilinear spaces.

Theorem E.17. *(*Sylvestor's Law of Inertia*) Two regular finite dimensional symmetric bilinear spaces over $\mathbb{R}$ are isometric if and only if they have the same signature.*

Proof. Let $V = \mathbb{R}v$ be a one dimensional real vector space with symmetric bilinear form B given by $B(v, v) > 0$. Then $B(v, v) = x^2$, some $x \in \mathbb{R}^\times$. The map $V \to \mathbb{R}$ given by $v \mapsto x$ determines an isometry $(V, B) \to (\mathbb{R}, B')$ with $B'(1, 1) = 1$. Similarly, if $B(v, v) = -x^2$, we get an isometry $(V, B) \to (\mathbb{R}, B')$ with $B'(1, 1) = -1$. $\square$

Exercise E.18. Prove Lemma E.7.

APPENDIX F

Primitive Roots

Let G be a finite abelian group. In this section, we compute the automorphism group of G. This is useful in the study of cyclotomic extensions of a field. We know that $G \cong \mathbb{Z}/n\mathbb{Z}$ for some positive integer n and $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$ (as generators must go to generators), so it is abelian of order $\varphi(n)$. By the Chinese Remainder Theorem, we are reduced to the case of of computing $(\mathbb{Z}/p^n\mathbb{Z})^\times$ where p is a (positive) prime. We also know that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic, so can assume that $n > 1$. We shall show that $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is also cyclic unless $p = 2$. A generator for this automorphism group is called a *primitive root*. The proofs rely on the binomial theorem and its special case, the *Children's Binomial Theorem*: that $(a+b)^p \equiv a^p + b^p \mod p$ for any integers a and b. In particular, it implies

Lemma F.1. *Let p be a (positive) prime and $n > 1$. If a, b, and k are integers, then $b^p \equiv (b + kp^n)^p \mod p^{n+1}$. In particular, if $a \equiv b \mod p^n$, then $a^p \equiv b^p \mod p^{n+1}$.*

Lemma F.2. *Let p be a (positive) odd prime and $n \geq 2$. Then for every integer a, we have $(1+ap)^{p^{n-2}} \equiv 1 + ap^{n-1} \mod p^n$.*

PROOF. We induct on n. The case $n = 2$ is trivial, so we may assume that $n > 2$. In particular, $2(n-1) > n+1$, so $p(n-1) > n+1$. By induction, we may assume the result for $n-2$ and show the result for $n-1$. By the lemma, induction, and the Binomial Theorem, we have

$$\begin{aligned}(1+ap)^{p^{n-1}} &\equiv \left((1+ap)^{p^{n-2}}\right)^p \equiv (1+ap^{n-1})^p \\ &\equiv 1 + \binom{p}{1} ap^{n-1} + \binom{p}{2} a^2 p^{2(n-1)} + \cdots + (ap^{n-1})^p \\ &\equiv 1 + \binom{p}{1} ap^{n-1} \equiv 1 + ap^n \mod p^{n+1}.\end{aligned}$$

□

Lemma F.3. *Let p be a (positive) odd prime and a an integer not divisible by p. Then for every integer $n \geq 2$, the congruence class of $1+ap$ in $\mathbb{Z}/p^n\mathbb{Z}$ has order p^{n-1}.*

PROOF. Let $^- : \mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ the canonical epimorphism. By the previous lemma, we know that $(1+ap)^{p^{n-1}} \equiv 1+ap^n \mod p^{n+1}$, hence $(1+ap)^{p^{n-1}} \equiv 1 \mod p^n$ and the order of $\overline{(1+ap)}$ divides p^{n-1}. The previous lemma also implies that $(1+ap)^{p^{n-1}} \equiv 1+ap^{n-1} \not\equiv 1 \mod p^n$, so the order of $\overline{(1+ap)}$ must be p^{n-1}. □

Proposition F.4. *Let p be a (positive) odd prime. Then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic for all n.*

PROOF. Let a be an integer relatively prime to p. We know that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic of order $p-1$, so generated by the residue class of a.

Claim 1. We may assume that $a^{p-1} \not\equiv 1 \mod p^2$:
Suppose that $a^{p-1} \equiv 1 \mod p^2$. Let $b = a + p$. Then

$$\begin{aligned} b^{p-1} = (a+p)^{p-1} &\equiv a^{p-1} + (p-1)a^{p-2}p \\ &\equiv 1 + (p-1)a^{p-2}p \not\equiv 1 \mod p^2 \end{aligned}$$

by the Binomial Theorem, so replacing a by b works.
Claim 2. Let $^{-} : \mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ the canonical epimorphism. If a is as in the first claim, then $\langle \overline{a} \rangle = (\mathbb{Z}/p^n\mathbb{Z})^\times$:
We know that $\varphi(p^n) = p^{n-1}(p-1)$ is the order of $(\mathbb{Z}/p^n\mathbb{Z})^\times$, so it suffices to prove that if $\overline{a}^N = 1$, then $p^{n-1} \mid N$ and $p-1 \mid N$ in $\mathbb{Z}$, as p and $p-1$ are relatively prime. By the first claim, we can write $a = 1 + xp$ for some integer x not divisible by p. By the last lemma, $\overline{1+xp}$ has order p^{n-1} in $(\mathbb{Z}/p^n\mathbb{Z})^\times$. As $\overline{a}^N = 1$, we must have $p^{n-1} \mid N$. Write $N = p^{n-1}M$ for some integer M. By Fermat's Little Theorem, $1 \equiv (a^N) = (a^M)^{p^{n-1}} \equiv a^M \mod p$, so $p-1 \mid M$ as a is relatively prime to p. Hence $\overline{a}^M$ has order $p-1$. This establishes Claim 2 and the theorem. □

We know that $(\mathbb{Z}/2\mathbb{Z})^\times \cong 1$ and $(\mathbb{Z}/2^2\mathbb{Z})^\times \cong \mathbb{Z}/2\mathbb{Z}$ are cyclic, however, this is not the case for $(\mathbb{Z}/2^n\mathbb{Z})^\times$ when $n > 2$.

Proposition F.5. *Let $n \geq 3$ be an integer. Then $(\mathbb{Z}/2^n\mathbb{Z})^\times \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{n-2}\mathbb{Z}$,*

PROOF. Let $n \geq 3$. We begin with the following:
Claim. $5^{2^{n-3}} \equiv 1 + 2^{n-1} \mod 2^n$:
This is true for $n = 3$, so we proceed by induction. As $n \geq 3$, we have $2n - 2 \geq n + 1$. Assuming the result for $n - 3$ we show it holds for $n - 2$. By Lemma F.1

$$5^{2^{n-2}} = (5^{2^{n-3}})^2 = (1 + 2^{n-1})^2 = 1 + 2^n + 2^{2n-2} \equiv 1 + 2^n \mod 2^{n+1},$$

proving the claim. The claim implies that $5^{2^{n-3}} \equiv 1+2^{n-1} \not\equiv 1 \mod 2^n$ and $5^{2^{n-2}} \equiv 1+2^n \mod 2^{n+1}$. It follows that $5^{2^{n-2}} \equiv 1 \mod 2^n$, and the order of $\overline{5}$ in $(\mathbb{Z}/2^n\mathbb{Z})^\times$ is 2^{n-2}, where $^{-} : \mathbb{Z} \to \mathbb{Z}/2^n\mathbb{Z}$ is the canonical epimorphism. To finish, it suffices to show that the subgroup $\langle \overline{-1}, \overline{5} \rangle$ in $\mathbb{Z}/2^n\mathbb{Z}$ has order 2^{n-1}, as this would imply that $(\mathbb{Z}/2^n\mathbb{Z})^\times \cong \langle \overline{-1} \rangle \times \langle \overline{5} \rangle$. So suppose that $(-1)^a 5^b \equiv (-1)^{a'} 5^{b'} \mod 2^n$ for some positive integers a, a', b, b'. As $n \geq 2$ and $5 \equiv 1 \mod 4$, we have $(-1)^a \equiv (-1)^{a'} \mod 4$, hence $a \equiv a' \mod 2$. It follows that $5^b \equiv 5^{b'} \mod 2^n$. We may assume that $b \geq b'$, hence $5^{b-b'} \equiv 1 \mod 2^n$. Since $\overline{5}$ has order 2^{n-2}, we have $2^{n-2} \mid b - b'$, i.e., $b \equiv b' \mod 2^n$. The result follows. □

The two propositions and the Chinese Remainder Theorem yield the desired result.

Theorem F.6. *Let n be a positive integer. Then $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic if and only if $n = 2, 4, p^r$, or $2p^r$ where p is an odd prime.*

APPENDIX G

The Sign of the Gauss Sum

In Section 59, we used Proposition 59.14 to determined the square of the Gauss sum defined in that proposition. This proposition was used to prove the Law of Quadratic Reciprocity 59.19 and Theorem 59.18 that every quadratic extension of $\mathbb{Q}$ was a subfield of a cyclotomic extension of $\mathbb{Q}$. We indicated what the sign of the Gauss sum in Proposition 59.14 was in Remark 59.15 without proof. It was this problem that was the most difficult for Gauss to solve and took him several years after proving what the square was. In this appendix, we shall prove Remark 59.15.

Let $\chi = \big(\frac{-}{p}\big) : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ denote the Legendre symbol with p an odd prime and ζ a primitive pth root of unity in $\mathbb{C}$. Set

$$\tau(\chi) := \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^a,$$

the Gauss sum S, defined in Proposition 59.14. Proposition 59.14 stated that $S^2 = \big(\frac{-1}{p}\big)p$. It follows that

$$\tau(\chi) = \pm\sqrt{p} \text{ or } \pm\sqrt{-p}, \tag{G.1}$$

but only remarked in Remark 59.15 what the correct signs were which depended on the congruence of p modulo 4. We now state again and prove the rest of the full theorem.

Theorem G.2. *Let p be an odd prime. Then*

$$\tau(\chi) = \begin{cases} \sqrt{p}, & \text{if } p \equiv 1 \mod 4 \\ \sqrt{-p}, & \text{if } p \equiv 3 \mod 4. \end{cases}$$

Proof. It follows by equation (G.1) that

$$\tau(\chi) = \begin{cases} \pm\sqrt{p}, & \text{if } p \equiv 1 \mod 4 \\ \pm\sqrt{-1}\sqrt{p}, & \text{if } p \equiv 3 \mod 4. \end{cases} \tag{G.3}$$

We must show that the sign is $+1$ in both cases.

Let $R = \{1, \dots, p-1\}$ (a system of representatives in $\mathbb{Z}$ of $(\mathbb{Z}/p\mathbb{Z})^\times$), $Q :=\{a \in R \mid \big(\frac{a}{p}\big) = 1\}$ and $N := \{a \in R \mid \big(\frac{a}{p}\big) = -1\}$. Therefore, we have

$$\tau(\chi) = \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^a = \sum_{q\in Q} \zeta^q - \sum_{n\in N} \zeta^n.$$

Since $0 = 1+\zeta+\cdots+\zeta^{p-1} = 1+\sum_{q\in Q}\zeta^q+\sum_{n\in N}\zeta^n$, we have $\tau(\chi) = 1+2\sum_Q \zeta^q$. As x runs through $0, 1, \dots, p-1$, then modulo p, we have x^2 takes the value 0 once and each square modulo p twice. As $\zeta^p = 1$, follows that

$$\tau(\chi) = \sum_{a=0}^{p-1} \zeta^{a^2}. \tag{G.4}$$

Using this, we shall turn the proof of the theorem into a linear algebra problem. Let $A \in \mathbb{M}_p(\mathbb{C})$ be the Vandermonde matrix

$$A = (\zeta^{kl})_{0\le k,l\le p-1} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \zeta & \zeta^2 & \cdots & \zeta^{p-1} \\ 1 & \zeta^2 & \zeta^4 & \cdots & \zeta^{2(p-1)} \\ \vdots & & & & \vdots \\ 1 & \zeta^{p-1} & \zeta^{2(p-1)} & \dots & \zeta^{(p-1)^2}. \end{pmatrix}$$

By equation (G.4), we know that $\tau(\chi) = \operatorname{trace} A$. This indicates why the matrix A will be involved in the proof, as will the matrix A^2. In particular, we shall use the Vandermonde determinant

$$\det A = \prod_{0\le l<k\le p-1} (\zeta^l - \zeta^k)$$

as well as $\det A^2$. We want to compute the trace of A. This turns our problem into an eigenvalue problem.

As A is triangularizable over $\mathbb{C}$, we that know that if $\lambda_1, \dots, \lambda_p$ (not necessarily distinct) in $\mathbb{C}$ are the eigenvalues of A that

$$\tau(\chi) = \operatorname{trace} A = \lambda_1 + \cdots + \lambda_p$$

and the eigenvalues of A^2 are $\lambda_1^2, \dots, \lambda_p^2$. We shall compute the λ_i^2's. The problem, of course, is that knowing λ_i^2 only determines its square root up to sign.

We first compute the characteristic polynomial f_{A^2} of A^2. As ζ^j, $j = 1, \dots, p-1$, is a primitive pth root of unity (with p an odd prime), the (k,l)th entry of A^2 is given by

$$(A^2)_{kl} = \sum_{j=0}^{p-1} \zeta^{kj}\zeta^{jl} = \sum_{j=0}^{p-1} \zeta^{j(k+l)} = \begin{cases} p & \text{if } k+l \equiv 0 \mod p \\ 0 & \text{if } k+l \not\equiv 0 \mod p. \end{cases}$$

Therefore, we have

$$A^2 = \begin{pmatrix} p & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & p \\ 1 & 0 & 0 & \cdots & p & 0 \\ \vdots & & & & & \vdots \\ 0 & p & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

as p is odd. Computation $\big[$using the properties of the determinant, e.g, if $B \in \mathbb{M}_{p-2}(\mathbb{C})$

$$\det\begin{pmatrix} t & 0\cdots 0 & -p \\ \vdots & B & \vdots \\ -p & 0\cdots 0 & t \end{pmatrix} = \det\begin{pmatrix} t-p & 0\cdots 0 & t-p \\ \vdots & B & \vdots \\ -p & 0\cdots 0 & t \end{pmatrix}$$
$$= (t-p)\det\begin{pmatrix} 1 & 0\cdots 0 & 1 \\ \vdots & B & \vdots \\ -p & 0\cdots 0 & t \end{pmatrix} = (t-p)\det\begin{pmatrix} 1 & 0\cdots 0 & 1 \\ \vdots & B & \vdots \\ 0 & 0\cdots 0 & t+p \end{pmatrix}$$
$$= (t-p)(t+p)\det B \quad \Big]$$

shows that the characteristic polynomial f_{A^2} of A^2 is

$$f_{A^2} = (t-p)^{\frac{p+1}{2}}(t+p)^{\frac{p-1}{2}}. \tag{G.5}$$

In particular,

$$\det A^2 = p^{\frac{p+1}{2}}(-p)^{\frac{p-1}{2}} = p^p(-1)^{\frac{p-1}{2}} = p^p(-1)^{\frac{p(p-1)}{2}}$$

as p is odd and

$$\det A = \pm p^{\frac{p}{2}}(\sqrt{-1})^{\frac{p(p-1)}{2}}. \tag{G.6}$$

It also follows that $\lambda_1^2, \dots, \lambda_p^2$ contain $\frac{p+1}{2}$ of the λ_i^2 having eigenvalue p and $\frac{p-1}{2}$ of the λ_i^2 having eigenvalue $-p$. Therefore, every λ_i lies in $\{+\sqrt{p}, -\sqrt{p}, \sqrt{-p}, -\sqrt{-p}\,\}$. We must compute the multiplicities of these eigenvalues. Let

$$r_+ = |\{i \mid \lambda_i = +\sqrt{p}\}| \qquad\qquad r_- = |\{i \mid \lambda_i = -\sqrt{p}\}|$$
$$s_+ = |\{i \mid \lambda_i = +\sqrt{-p}\}| \qquad\qquad s_- = |\{i \mid \lambda_i = -\sqrt{-p}\}|.$$

By equation (G.5), we have

$$r_+ + r_- = \frac{p+1}{2} \quad\text{and}\quad s_+ + s_- = \frac{p-1}{2}. \tag{G.7}$$

Hence

$$\tau(\chi) = \lambda_1 + \cdots + \lambda_p = \big(r_+ - r_- + (s_+ - s_-)\sqrt{-1}\big)\sqrt{p}. \tag{G.8}$$

Using equations (G.3) and (G.8), we see that

$$\begin{aligned} &r_+ - r_- = \pm 1 && \text{and} \quad s_+ = s_- && \text{if } p \equiv 1 \mod 4 \\ &r_+ = r_- && \text{and} \quad s_+ - s_- = \pm 1 && \text{if } p \equiv 3 \mod 4. \end{aligned} \tag{G.9}$$

We want to determine the sign in equations (G.9). Equation (G.6) implies that

$$\begin{aligned} \det A &= \prod_{i=1}^{p} \lambda_i = p^{\frac{p}{2}}(-1)^{r_-}(\sqrt{-1})^{s_+}(-\sqrt{-1})^{s_-} \\ &= p^{\frac{p}{2}}(\sqrt{-1})^{2r_- + s_+ - s_-}. \end{aligned} \tag{G.10}$$

We can now determine the sign of $\det A$ as computed in equation (G.6) using that $\det A$ is a Vandermonde determinant. Let ζ_{2p} be a primitive $2p$th root of unity. So we have

$$\zeta^a = \zeta_{2p}^{2a} = \cos\frac{2\pi a}{2p} + \sqrt{-1}\sin\frac{2\pi a}{2p} = \cos\frac{\pi a}{p} + \sqrt{-1}\sin\frac{\pi a}{p}.$$

Then for integers $0 \le l < k \le p-1$, we have

$$\begin{aligned}\det A &= \prod_{0\le l<k\le p-1} (\zeta^k - \zeta^l) = \prod_{0\le l<k\le p-1} (\,\zeta_{2p}^{2k} - \zeta_{2p}^{2l}\,)\\ &= \prod_{0\le l<k\le p-1} \zeta_{2p}^{k+l}(\,\zeta_{2p}^{k-l} - \zeta_{2p}^{-(k-l)}\,)\\ &= \prod_{0\le l<k\le p-1} \zeta_{2p}^{k+l} \prod_{0\le l<k\le p-1} \big(2\sqrt{-1}\,\sin\frac{(k-l)\pi}{p}\big).\end{aligned}$$

Now

$$\sum_{0\le l<k\le p-1}(k+l) = \sum_{k=1}^{p-1}\sum_{l=0}^{k-1}(k+l) = \sum_{r=1}^{p-1}\big(k^2 + \frac{k(k-1)}{2}\big) = 2p(\frac{p-1}{2})^2$$

is divisible by $2p$ and $\zeta_{2p}^{2pn} = 1$ for all integers n. Therefore, by equation (G.6), we have

$$\pm\, p^{\frac{p}{2}}(\sqrt{-1})^{\frac{p(p-1)}{2}} = \det A = (\sqrt{-1})^{\frac{p(p-1)}{2}} 2^{\frac{p(p-1)}{2}} \sin\frac{(k-l)\pi}{p}. \tag{G.11}$$

Since $\sin(\dfrac{(k-l)\pi}{p}) > 0$ for $0 \le l < k \le p-1$, we must have the plus sign in equation (G.11). Hence, by equations (G.6) and (G.10),

$$p^{\frac{p}{2}}(\sqrt{-1})^{\frac{p(p-1)}{2}} = \det A = \prod_{k=1}^{p}\lambda_k = p^{\frac{p}{2}}(\sqrt{-1})^{2r_-+s_+-s_-}.$$

Comparing exponents, implies that

$$2r_- + s_+ - s_- \equiv p(\frac{p-1}{2}) \mod 4. \tag{G.12}$$

Consequently, by equations (G.7) and (G.9) we deduce that

$$\begin{aligned}&\text{If } p\equiv 1 \mod 4, \text{ then } s_+ = s_- \text{ and}\\ &\qquad r_+ - r_- = \frac{p+1}{2} - 2r_- = \frac{p+1}{2} - \frac{p-1}{2} \equiv 1 \mod 4\\ &\text{If } p\equiv 3 \mod 4, \text{ then } r_+ = r_- \text{ and}\\ &\qquad s_+ - s_- = -\frac{p-1}{2} + 2r_- = -\frac{p-1}{2} + \frac{p+1}{2} \equiv 1 \mod 4.\end{aligned}$$

Therefore, $r_+ - r_-$ and $s_+ - s_-$ are both are $+1$, and the result follows by equation (G.9). □

APPENDIX H

Pell's Equation

In this section, we show that the well-known diophantine equation

$$(*) \qquad x^2 - dy^2 = 1 \text{ with } d \text{ a positive integer,}$$

is solvable, i.e., there exist integers x, y satisfying (*). In elementary number theory, it is shown that the solutions may be found using continued fractions, a method based on repeated use of the division algorithm, although it may not so easy to calculate. We need two lemmas.

Lemma H.1. *Let α be an irrational number. Then there exist infinitely many relatively prime integers x, y satisfying*

$$|\frac{x}{y} - \alpha| < \frac{1}{y^2}.$$

Proof. Let n be a positive integer. Partition the half-open interval $[0, 1)$ into n half-open subintervals,

$$(\dagger) \qquad [0,1] = \bigvee_{j=0}^{n-1} [\frac{j}{n}, \frac{j+1}{n}).$$

Recall if β is a real number then $[\beta]$ is defined to be the largest integer $\leq \beta$, and $\beta - [\beta]$ is then called the *fractional part* of β. Consider the $n + 1$ fractional parts of the elements $0, \alpha, 2\alpha, \ldots, n\alpha$. By the Dirichlet Pigeon Hole Principle, two of these fractional parts must lie in the same subinterval in (†). Hence there exist integers j and k satisfying $0 \leq j < k \leq n$ and

$$|(k\alpha - [k\alpha]) - (j\alpha - [j\alpha])| < \frac{1}{n}.$$

Set $y = k - j$ and $x = [j\alpha] - [k\alpha]$. We have $0 < y < n$ and

$$0 < |\frac{x}{y} - \alpha| < \frac{1}{yn} < \frac{1}{y^2}.$$

Therefore, x/y is one solution. We may assume that x and y are relatively prime, since dividing by the gcd of x and y would also lead to a solution. Now choose an integer $n_1 > 1/|(x/y) - \alpha|$. Repeating the construction would yield a new solution x_1, y_1 in relatively prime integers. If follows that we can construct infinitely many such solutions. $\square$

Lemma H.2. *Let d be a positive square-free integer. Then there exists a constant c satisfying*

$$|x^2 - dy^2| < c$$

has infinitely many solutions in integers x, y.

PROOF. We have $x^2 - dy^2 = (x - y\sqrt{d})(x + y\sqrt{d})$ in $\mathbb{Z}[d]$. By the lemma, there exist infinitely many relatively prime integers x, y with $y > 0$ and satisfying $|x - y\sqrt{d}| < 1/y$. Therefore,

$$|x + y\sqrt{d}| < |x - y\sqrt{d}| + 2y\sqrt{d} < \frac{1}{y} + 2y\sqrt{d}$$

and

$$|x^2 - dy^2| < \frac{1}{y^2} + 2\sqrt{d} < 1 + 2\sqrt{d},$$

so $c = 1 + 2\sqrt{d}$ works. □

Proposition H.3. *Let d be a square-free positive integer. Then Pell's equation $x^2 - dy^2 = 1$ has infinitely many solutions in integers. Further, there exists an integral solution (x_1, y_1) such that every solution is of the form (x_n, y_n) with $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$ for some integer n, In particular, the solution set in integers to Pell's equation forms an infinite cyclic group.*

PROOF. We begin with some notation. If $\gamma = x + y\sqrt{d}$, x, y in $\mathbb{Q}$, let $\overline{\gamma} = x - y\sqrt{d}$ and $\mathrm{N}(\gamma) = \gamma\overline{\gamma} = x^2 - dy^2$. By the last lemma, we know that there exists a nonzero integer m such that $x^2 - dy^2 = m$ has infinitely many solutions x, y in positive integers. We may also assume that we have reduced this set of solutions to those having different x's. In particular, among these infinitely many solutions in positive integers, there exist two such pairs x_1, y_1 and x_2, y_2 to this equation with $x_1 \neq x_2$, $x_1 \equiv x_2 \mod |m|$, and $y_1 \equiv y_2 \mod |m|$. Let $\alpha = x_1 + y_1\sqrt{d}$ and $\beta = x_2 + y_2\sqrt{d}$. By choice, $\mathrm{N}(\alpha) = m = \mathrm{N}(\beta)$. Write $\alpha\overline{\beta} = r + s\sqrt{d}$ with r and s integers satisfying $r \equiv 0 \equiv s \mod |m|$. It follows that

$$\alpha\overline{\beta} = m(a + b\sqrt{d})$$

for some integers a and b. Taking N of this equation yields $a^2 - db^2 = 1$. Suppose that $b = 0$, then $a = \pm 1$ and $\alpha m = \alpha\beta\overline{\beta} = \pm m\beta$, i.e., $\alpha = \pm\beta$. This implies $x_1 = x_2$, a contradiction. It follows that Pell's equation has a solution a, b to $a^2 - db^2 = 1$ with $ab \neq 0$.

Among all the solutions a, b to Pell's equation choose one in positive integers a, b with the real number $\alpha = a + b\sqrt{d} > 0$ minimal. Suppose that x, y is another solution in positive integers. Set $\beta = x + y\sqrt{d}$. We show that $\beta = \alpha^n$ for some positive integer n. Suppose not. Then there exists a positive integer n such that $\alpha^n < \beta < \alpha^{n+1}$. Since $\alpha\overline{\alpha} = 1$, i.e., $\alpha^{-1} = \overline{\alpha}$, we have

$$1 < \beta\alpha^{-n} < \alpha.$$

We can write $1 < \beta\alpha^{-n} = \beta\overline{\alpha}^n = r + s\sqrt{d}$ with r and s integers. As $\mathrm{N}(\beta\overline{\alpha}^n) = \mathrm{N}(\beta)(\overline{\alpha}) = 1$, we have r, s is also a solution of Pell's equation. Since $r + s\sqrt{d} > 0$, we have $0 < r - s\sqrt{d} = (r + s\sqrt{d})^{-1} < 1$. In particular, r is positive and $s\sqrt{d} > r - 1 \geq 0$. Therefore, s is also positive, contradicting the minimality of a, b. If x, y is a solution in integers to Pell's equation with $x > 0$ and $y < 0$, then $\beta = x + y\sqrt{d}$ satisfies $\mathrm{N}(\beta) = 1$, so $\overline{\beta} = \alpha^n$ for some n, hence $\beta = \alpha^{-n}$. The cases $x < 0$, $y > 0$ and $x < 0, y < 0$ lead to solutions $-\alpha^n$ for some n in $\mathbb{Z}$. □

Bibliography

Texts on Abstract Algebra

[1] M. Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991.
[2] Cohn, P. M., *Algebra. Vol. 1*, Second Edition, John Wiley & Sons Ltd., Chichester, 1982.
[3] Cohn, P. M., *Algebra. Vol. 2*, Second Edition, John Wiley & Sons Ltd., Chichester, 1989.
[4] Cohn, P. M., *Algebra. Vol. 3*, Second Edition, John Wiley & Sons Ltd., Chichester, 1991.
[5] Dummit, David S. and Foote, Richard M., *Abstract algebra*, Third Edition, John Wiley & Sons Inc., Hoboken, NJ, 2004.
[6] Herstein, I. N., *Topics in algebra*, Second Edition, Xerox College Publishing, Lexington, Mass., 1975.
[7] Jacobson, Nathan, *Basic algebra. I*, Second Edition, W. H. Freeman and Company, New York, 1985.
[8] Jacobson, Nathan, *Basic algebra. II*, Second Edition, W. H. Freeman and Company, New York, 1989.
[9] Lang, Serge, *Algebra*, Graduate Texts in Mathematics, Vol. 211, Third Edition, Springer-Verlag, New York, 2002.
[10] Rotman, Joseph J., *Advanced modern algebra*, Graduate Studies in Mathematics, Vol. 114, Second Edition, American Mathematical Society, Providence, RI, 2010.

Books on Group Theory

[11] Robinson, Derek J. S., *A course in the theory of groups*, Graduate Texts in Mathematics, Vol 80, Second Edition, Springer-Verlag, New York, 1996.
[12] Rose, John S., *A course on group theory*, Reprint of the 1978 original [Dover, New York]], Dover Publications Inc., New York, 1994.
[13] Rotman, Joseph J., *An introduction to the theory of groups*, Graduate Texts in Mathematics, Vol. 148, Fourth Edition, Springer-Verlag, New York, 1995.

Books on Field Theory

[14] Artin, Emil, *Galois theory*, Second Edition, Edited and with a supplemental chapter by Arthur N. Milgram, Dover Publications Inc., 1998.
[15] Cox, David A., *Galois theory*, Pure and Applied Mathematics (Hoboken), Second Edition, John Wiley & Sons Inc., Hoboken, NJ, 2012.
[16] Hadlock, Charles Robert, *Field theory and its classical problems*, Carus Mathematical Monographs, Vol. 19, Mathematical Association of America, Washington, D.C., 1978.
[17] Kaplansky, Irving, *Fields and rings*, Chicago Lectures in Mathematics, Reprint of the second (1972) edition, University of Chicago Press, Chicago, IL, 1995.
[18] McCarthy, Paul J., *Algebraic extensions of fields*, Second Edition, Dover Publications Inc., New York, 1991.

Books on Number Theory

[19] Apostol, Tom M., *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976.

[20] Ireland, Kenneth and Rosen, Michael, *A classical introduction to modern number theory*, Graduate Texts in Mathematics, Vol. 84, Second Edition, Springer-Verlag, New York, 1990.

[21] Lang, Serge, *Algebraic number theory*, Graduate Texts in Mathematics, Vol 110, Second Edition, Springer-Verlag, New York, 1994.

[22] LeVeque, William Judson, *Topics in number theory. Vol. I, II*, Reprint of the 1956 original [Addison-Wesley Publishing Co., Inc., Reading, Mass.; MR0080682 (18,283d)], with separate errata list for this edition by the author, Dover Publications Inc., Mineola, NY, 2002.

[23] LeVeque, William J., *Fundamentals of number theory*, Reprint of the 1977 original, Dover Publications Inc., Mineola, NY, 1996.

[24] Marcus, Daniel A., *Number fields*, Universitext, Springer-Verlag, New York, 1977.

[25] Neukirch, Jürgen, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 322, Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder, Springer-Verlag, Berlin, 1999.

Books on Transcendental Number Theory

[26] Lipman, Joseph, *Transcendetal Numbers*, Queen's Papers in Pure and Appllied Mathematics – No. 7, Kingston, Ontario, New York, 1966

[27] Murty, M.Ram and Rath, Purusottam, *Number fields*, Springer-Verlag, New York, 2014

Books on Commutative Algebra

[28] Atiyah, M. F. and Macdonald, I. G., *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969.

[29] Eisenbud, David, *Commutative algebra – With a view toward algebraic geometry*, Graduate Texts in Mathematics, Vol. 150, Springer-Verlag, New York, 1995.

[30] Kaplansky, Irving, *Commutative rings*, Revised, The University of Chicago Press, Chicago, Ill.-London, 1974.

Books on Non-Commutative Algebra

[31] Curtis, Charles W. and Reiner, Irving, *Methods of representation theory. Vol. I With applications to finite groups and orders*, Wiley-Interscience Publication, John Wiley & Sons, 1981, Wiley Classics Library, Reprint of the 1981 edition, 1990.

[32] Lam, T. Y., *A first course in noncommutative rings*, Graduate Texts in Mathematics, Vol. 131, Second Edition, Springer-Verlag, 2001.

[33] Pierce, Richard S., *Associative algebras*, Graduate Texts in Mathematics, Vol. 88, Studies in the History of Modern Science, 9, Springer-Verlag, New York, 1982.

Books on Formally Real Fields

[34] Scharlau, Winfried, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 270, Springer-Verlag, Berlin, 1985.

[35] Lam, T. Y., *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics, Vol. 67, American Mathematical Society, Providence, RI, 2005.

Books on Homological Algebra

[36] Weibel, Charles, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics, Vol. 38, Cambridge University Press, Cambridge, 1994.

[37] Rotman, Joseph, *An introduction to homological algebra, 2nd edition*, Universitext, Springer-Verlag, New York, 2009.

Books on Category Theory

[38] Riehl, Emily, *Category Theory in Context*, Aurara Dover Modern Math Originals, Dover, New York, 2016.

[39] Mac Lane, Saunders, *Categories for the Working Mathematician*, Graduate Texts in Mathematics, Vol. 5, Springer-Verlag, New York, 1971.

Notation

$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$, representation of a permutation, 57
(A^*, d^*), cochain complex, 686, 801
(A_*, d_*), chain complex, 685
$(L/F, f)$, crossed product algebra, 607
$(R, \mathfrak{m})$, local ring with maximal ideal $\mathfrak{m}$, 514
$(\mathbb{Z}/m\mathbb{Z})^\times$, units in $\mathbb{Z}/m\mathbb{Z}$, 32
$(\mathfrak{A} : \mathfrak{B})$, 526
(a), principal ideal generated by a, 142
$(a_1 \cdots a_r)$, r-cycle, 125
$(a_1, \dots, a_n)$, ideal generated by $a_1, \dots, a_n$, 142
$(g_i)_I$, elements in $\bigtimes_I G_i$, 45
$* : G \times S \to S$, G-action, 97
1_R, identity (unity) of R, 137
1_S, the identity map on S, 41
$A < B$, same as $A \subset B$, $A \neq B$, 53
$A \setminus B$, element in A not B, 3
$A \xrightarrow{f} B$, map from A to B, 6
A^{op}, opposite algebra, 594
A^K, 594
A_K, integral closure of A in K, 447
A_n, alternating group group on n letters, 57
A_n, alternating group on n-letters, 127
$C(P_1, \dots, P_n)$, constructible points from $P_1, \dots, P_n$, 298
$C(a)$, conjugacy class of a, 103
$C(z_1, \dots, z_n)$, constructible points from $z_1, \dots, z_n$, 299
$D(R)$, basic open set, 509
D_n, dihedral group of order $2n$, 43
$D_n(R)$, diagonal group, 44
$F(X)$, smallest field containing F and S, 277
F^S, fixed field of F under the set S of automorphisms, 313
F^p, $\{x^p \mid x \in F\}$, 309
$F_G(S)$, G-fixed point set, 99
$G \cong G'$, isomorphic, 48
$G(K/F)$, Galois group of K/F, 314
G/H, factor group of G by H, 62
G/H, set of cosets, 51
$G^{(n)}$, $(G^{(n-1)})'$, 59, 82
G^{ab}, abelianization of G, 68
$G_{\mathbb{A}}(R)$, 560
G_s, isotropy subgroup (stabilizer) of s, 98
$H \rtimes_\varphi G$, semidirect product of H and G via φ, 59
$H \triangleleft G$, normal subgroup, 56
$H \triangleleft\triangleleft G$, H a characteristic subgroup of G, 58
$H_n(A)$, homology of (A_*, d_*), 685, 801
K/F, field extension, 275
$L_A(M)$; ring of left multiplications by A on M, 630
$M_n(R)$, matrices coefficients in R, 30
$N_G(H)$, normalizer of H in G, 46, 105
Q, the rational numbers, 3
R, ring, 137
R, the set of real numbers, 6
$R/\mathfrak{A}$, quotient (factor) ring of R by $\mathfrak{A}$, 147
$R[[t]]$, formal power series over R, 139
$R^\times$, units in R, 41
$R^\times$, units in ring R, 32
$R^\times$, units of R, 138
$R_\mathfrak{p}$, localization at $\mathfrak{p}$, 512
R_a, localization at set of the powers of a, 512
$S \vee T$, disjoint union of S and T, 10
$S^{-1}R$, localization of R by S, 163
S_n, symmetric group on n letters, 41
$V_R(T)$, variety of T, 233, 507
$Z(G)$, center of G, 57, 103
$Z(R)$, center of R, 141
$Z_F(\mathfrak{A})$, zero set in F^n of $\mathfrak{A}$, 199
$Z_G(a)$, centralizer of a, 103
$[G, G] = G'$, commutator subgroup, 59, 82
$[G : H]$, index of H in G, 52
$[K : F]$, degree of K over F, 276
$[T]_{\mathcal{B},\mathcal{C}}$, matrix representation relative to bases $\mathcal{B}, \mathcal{C}$, 57
$[x, y]$, commutator of x and y, 82
$[x]_\sim$, equivalence class of x, 27
$\mathrm{Ass}_R\, V(\mathfrak{A})$, associated primes of $\mathfrak{A}$ (or $R/\mathfrak{A}$), 528
$\mathrm{Aut}(G)$, automorphism group of G, 56

$\mathrm{Aut}_F(K)$, F-(algebra) automorphisms of K, 290
$\mathrm{Aut}_F(V)$, automorphism group of F-vector space V, 44
$\mathrm{Aut}_F(V)$, automorphism group of vector space V, 108
$\mathrm{Br}(F)$, Brauer group of F, 604
$\mathbb{C}$, set of complex numbers, 3
$\mathrm{char}(R)$, characteristic of R, 149
Δ_K, prime subfield of K, 153
$\mathrm{End}(A)$, endomorphism ring of A, 146
$\mathrm{End}_F(V)$, endomorphism ring of vector space V, 108
$\mathrm{GL}_n(R)$, general linear group over R, 44
$\Longrightarrow$, implies, 6
$\mathbb{M}_n(R)$, $n \times n$ matrix ring over R, 26
$\mathrm{Max}(R)$, set of maximal ideals, 234
$\mathrm{Min}(R)$, set of minimal primes in R, 509
$\mathbb{Q}[t]$, polynomials with rational coefficients, 3
$\mathbb{R}^+$, the set of positive reals, 41
$\Sigma(S)$, permutation group on S, 41
$\mathrm{Spec}(R)$, Spectrum of R, 233
$\mathrm{Syl}_p(G)$, set of Sylow p-subgroups of G, 107, 111
$\mathbb{Z}$, the set of integers, 3
$\mathbb{Z}/m\mathbb{Z}$, integers modulo m, 29
$\mathbb{Z}^+$, non-negative integers, 6
$\mathbb{Z}^+$, positive integers, 6
$\mathbb{Z}^+$, the set of positive integers, 8
$\approx$, usually an equivalence relation, 25
$\bigcup_I A_i$, $\bigcup_{i\in I} A_i$, the union of the sets A_i, 27
$\bigvee_I A_i$, disjoint union of the sets i in I, 27
$\mathcal{C}^*$, is $= \mathcal{C} \setminus Z(G)$, 103
$\mathcal{F}(K/F)$, intermediate fields of K/F, 330
$\mathcal{G}$, set of subgroups of G, 106
$\mathcal{G}(K/F)$, subgroups of $G(K/F)$, 330
$\mathcal{O}$, system of representatives for a G-action, 98
$\mathcal{O}^*$, is $\mathcal{O} \setminus F_G(S)$, 99
$\mathcal{P}(S)$, power set of S, 106
$\cong$, isomorphism, 140
$\coprod_I M_i$, coproduct of the M_i, 210
$\deg A$, index of the algebra A, 598
det, determinant, 49
$\emptyset$, the empty set, 9
$\equiv \mod m$, congruence modulo m, 29
$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}$, ideals, 141
$\mathfrak{A}\mathfrak{B}$, ideal generated by $\mathfrak{A}$, and $\mathfrak{B}$, 142
$\mathfrak{B} \mid \mathfrak{A}$, ideal division, 455
$\mathfrak{m}$, maximal ideal, 145
$\mathfrak{p}, \mathfrak{P}$, prime ideals, 145
$\mathrm{im}\,\varphi$, image of φ, 48, 140
$\mathrm{ind}_H^G(W)$, induced module, 649
$\ker\varphi$, kernel of φ, 48, 140
$\langle W\rangle$, group (etc.) generated by W, 42
$\langle a\rangle$, group (etc) generated by a, 42
λ_a, left translation by a, 52
λ_x, left multiplication by x, 62
$\mid A \mid$, cardinality of A, 6
μ_n, nth roots of unity (in $\mathbb{C}$), 42
$\mathrm{nil}(R)$, nilradical of R, 160
$\mathrm{SL}_n(R)$, special linear group over R, 44
$\overline{A}$, set of equivalence class of A, 27
$^- : G \to G/H$, canonical group epimorphism, 51
$\overline{x}$, equivalence class of x, 27
$^a\varphi$, 510
$\pi(x)$, 3
$\prod R_i$, $\bigtimes R_i$, product of rings, 153
$\prod_I M_i$, product of the M_i, 210
$\mathrm{rad}(R)$, Jacobson radical of R, 518
$\mathrm{res}_H^G(V)$, restriction module, 649
sgn, signum (sign) map, 127
$\sim$, usually an equivalence relation, 25
$\sim_G$, G=equivalence, 98
$\sqrt{\mathfrak{A}}$, radical of $\mathfrak{A}$, 162, 508
θ_x, conjugation by x, 55
$\tilde{\varphi} : R[t] \to S[t]$, ring extension of $\varphi : R \to S$, 288
$\mathrm{tr}\deg_F A$, $\mathrm{tr}\deg_F qf(A)$, 545
$\mathrm{tr}\deg_F K$, transcendence degree of K/F, 285
$\varphi : G \widetilde{\to} G'$, isomorphism, 48
$\mathrm{zd}(R)$, zero divisors of R, 159
$a \approx b$, a is an associate of b, 144
$a \mid b$, a divides b, 3, 144
$a \nmid b$, a does not divide b, 3
a^{-1}, inverse of a, 40
$e = e_G$, unity in the group G, 39
$f : A \to B$, map from A to B, 6
$f^{-1}(D)$, preimage of D, 35, 64
$f^{-1}(b)$, fiber of f at b, 35
$l(M)$, length of a module M having a composition series., 538
$m_F(\alpha)$, minimal polynomial of α over F, 279
$qf(R)$, quotient field of domain R, 150
t, represents a variable, 3
$v_{\mathfrak{p}}(\mathfrak{A})$, power of $\mathfrak{p}$ dividing $\mathfrak{A}$, 455
xHx^{-1}, conjugate of H, 55
$\mathrm{Sp}_n(R)$, symplectic group over R, 44
$\mathrm{O}_n(R)$, orthogonal group over R, 44
$\mathrm{O}_{3,1}(\mathbb{R})$, Lorenz group, 44
$\mathrm{SO}_n(R)$, special orthogonal group over R, 44
$\mathrm{ST}_n(R)$, strictly upper triangular group over R, 44
$\mathrm{SU}_n(R)$, special unitary group, 44
$\mathrm{T}_n(R)$, upper triangular group over R, 44
$\mathrm{U}_n(\mathbb{C})$, unitary group, 44

$\mathrm{Core}_G(H)$, core of H in G, 45
$\mathrm{Inn}(G)$, 56
$\mathrm{PSL}_n(F)$, projective special linear group, 131
$\mathrm{Perm}_n(\mathbb{R})$, permutation matrices over $\mathbb{R}$, 57
:=, defined by, 3
[H,K], the group generated by $[h,k]$, 86

module
induced, 649

Index

C_n field, 591
G-action, 101
G-module
  fixed points of, 645
  trivial, 645
G-set, 101
R-sequence, 578, 778
  on a module, 578, 778
S-units, 520
$\mathrm{End}_R(M, N)$, 726
Hom
  left exact, 716
$\mathfrak{A}$-adic topology, 586
$\mathfrak{p}$-primary ideal, 545
$\mathrm{sgn}_P(B)$, signature at P, 861
$\underline{\mathrm{End}}^*_R(M, N)$, 747
n-(multi)linear map, 686
p-adic integers, 587
p-group, 107
(category
  object, 785

Abel's Theorem, 60, 133
Abel-Ruffini Theorem, 373
abelian group, 42
  basis, 74
  free, 92
  rank, 74
absolutely irreducible module, 667
ACC, 180
action
  compatible, 223, 262
action of a group on a set, 101
action of a ring, 219
adjoint
  reflector, 816
Adjoint Associativity Theorem, 758
adjoint functors, 802
adjoints, 758
adjunction, 803
affine (plane) curve, 249
affine algebra, 243, 563
affine group scheme, 801
affine scheme, 815
affine variety defined by $\mathfrak{A}$, 211
Akizuki's Theorem, 559
algebra, 151, 603, 692
  F-split, 650
  K-split, 618
  affine, 243, 563
  center of, 603
  central, 614
  central simple, 603
    similarity of, 603
  crossed product, 612, 627
  cyclic of degree d, 609
  derivation, 441
  exterior, 696
  finite dimensional, 603
  finitely generated commutative, 243
  free commutative, 692
  graded, 692
    homogeneous element of, 692
  opposite, 614
  semisimple
    basic set for, 607
    simple components, 607
    Wedderburn decomposition, 607
  split, 618
  symmetric, 694
  tensor, 693
algebra homomorphism
  graded, 692
algebraic closure of a field, 296
algebraic element, 292
algebraic integer, 429
algebraic integers
  ring of, 366
algebraic number, 429
algebraic number field, 433, 468
algebraically closed
  right (respectively, left), 639
algebraically closed field, 296
algebraically independent elements, 298

almost all, 166, 222
almost all zero, 205
alphabet
  letters of, 94
  of a free group, 94
alternating group, 59, 131
  3-cycles and, 132
  A_5, 134
  simplicity for $n \geq 5$, 60, 133
alternating group on n letters, 59
analytic at infinity, 407
analytic function, 403, 433
  zero, 433
annihilator, 224
  of a module, 478
anti-derivation, 700
antiautomorphism, 192
antihomomorphism, 192
arithmetic function, 22
  completely multiplicative, 22
  division functions, 22
  Euler φ-function, 22
  identity arithmetic function, 22
  Liouville function, 22
  Möbius μ-function, 22
  multiplicative, 22
  von Mangoldt function, 22
Arithmetic Lemma, 664
arithmetic progressions, 190
Artin's Lemma, 327
  consequence of, 328
Artin's Theorem, 329
Artin-Tate Lemma, 244
Artin-Weddeerburn Theorem, 607
Artin-Wedderburn Theorem, 605
Artinian module, 242
Artinian ring, 183
ascending chain condition, 180, 239
associated map, 530
associated prime ideal, 548
associated primes
  isolated subset, 550
Auslander's Theorem, 754
Auslander-Buchsbaum Theorem, 783
automorphism, 58
  field, 291
    Frobenius, 336
  fixing, 186
  Frobenius, 485
  inner, 58
  moves element, 291
  ring, 150
automorphism group, 113, 223
  of a group, 58
  of vector space, 46
Axiom of Choice, 835

Bézout domain, 497, 499
Baer Criterion, 230, 719
balanced biadditive map, 755
base extension, 690
base units, 519
basic open set, 529
basis
  complementary, 469
  for a free abelian group, 92
  for a free group, 92
  integral, 433, 471
  minimal, 487
  projective, 736
Bertrand's Hypothesis, 12
Bertrand's Theorem, 837
Bezout's Lemma, 250
bilinear form, 686
bilinear space
  discriminant, 514
  positive definite, 514
bimodule, 757
binary operation, 41
Binet-Cauchy Equations, 703
binomial coefficients, 12
binomial theorem, 13, 156
blinear space, 514
boundaries, 697
Brauer equivalence, 624
Brauer Group, 624
Brauer group, 603
Brill's Theorem, 490
Burnside Conjecture, 677
Burnside Counting Theorem, 388
Burnside's $p^a q^b$-Theorem, 667
Burnside's Lemma, 665
Burnside's Theorem, 135, 650
Butterfly Lemma (Zassenhaus), 86

cancellation law, 42, 148
canonical homomorphism, 689
Cantor's Theorem, 6
Casimir element, 691
categorical definitions, 681
category, 785
  abelian, 820
  additive, 817
  arrows, 785

comma, 790
concrete, 793
congruence relation, 787
contravariant functor, 789
discrete, 809
enough injectives, 826
enough projective objects, 828
epimorphism, 788
split, 788
equivalent, 794
full subcategory, 793
functor, 789
functor category, 795
hom-set, 796
initial object, 788
isomorphism, 789
locally small, 786
monomorphism, 788
split, 788
morphism, 785
codomain, 785
compatible, 785
domain, 785
natural transformation, 794
of cocones, 808
of cones, 807
opposite, 788
poset, 786
preadditive, 817
product, 791
projective object, 828
quotient, 787
slice, 787, 788
small, 786
subcategory
reflective, 816
terminal object, 788
zero object, 788
Cauchy's Theorem, 70, 113
abelian case, 70
Cayley's Theorem, 65, 800
general form, 64
Cayley-Hamilton Characteristic Sequence, 284
Cayley-Hamilton Theorem, 272, 544
central F-algebra, 614
central simple algebra
exponent, 633
index of, 618
period, 633
split, 603, 611
centralizer
of a subalgebra, 616
chain, 165, 833
stablized, 239
chain complex, 225, 707, 821, 823
n-chain, 707, 824
acyclic, 225
augmentation, 746
augmented, 823
boundaries of, 707, 824
cycles, 697
cycles of, 707, 824
differential of degree -1, 697
homology, 697
homology of, 707, 823
long exact sequence in homology, 710, 824
chain complexes
differentials, 707, 823
chain homotopy, 712
chain map
of degree i, 712
change of rings theorems
first, 773
general, 772
second, 775
third, 777
character
degree of, 655
irreducible, 655
krnel of, 657
real, 662
regular, 663
trivial, 655
character (linear), 325
character of a representation, 655
character table, 661
characteristic
of a domain, 159
characteristic matrix, 275, 852
characteristic polynomial, 269
reduced, 621
characteristic sequence, 274, 284
characterization of completely reducible modules, 599
characters
algebraically independent, 385
complex, 662
dependent, 325
independent, 325
inequivalent, 659
Chebysev's Theorem, 841
Chebyshev, 16
Bertrand's Hypothesis, 16, 837
binomial coefficient, 16

primenumber theorem approximation, 16
Chevalley-Warning Theorem, 397
Children's Binomial Theorem, 35, 165, 307
Chinese Remainder Theorem (for $\mathbb{Z}$), 33
homomorphism interpretation, 35
Chinese Remainder Theorem (for commutative rings, 163
circle group, 44
class equation, 107
class function, 661
class group, 476
class number, 522
class sum, 646
classical adjoint, 704
Classification of Cyclic Groups, 51, 130
classification of similarity classes of matrices over a field, 273
closed points, 248
cochain comples
n-cochain, 708, 824
cochain complex, 708, 824
acyclic, 708
augmentation, 723
coboundaries of, 708, 824
cocycles of, 708, 824
cohomology of, 708, 824
positive, 723
cocone
seecone
over a diagram to a functor with nadir, 808
codimension, 568
cofactor, 704
Cohen's Theorem, 186
Cohen-Macaulay ring, 579
Cohen-Seidenberg Theorem, 539
cohomological δ functor, 826
cohomological *delta*-functor
universal, 826
cohomology, 821
cohomology group
1st, 380
2nd, 628
coimage, 683
cokernel, 226, 682
colon of ideals, 546
combinatorial dimension
of a topological space, 534
commutative diagram, 37
commutative ring, 32, 42, 148
(integral domain), 148
PID, 153
(integral) domain, 48
local, 175
Artinian, 183
ascending chain condition (or ACC), 180
chain of ideals stabilizing, 180, 183
coherent, 241
completion relative to an ideal, 586
descending chain condition (or DCC), 183
discrete valuation ring, 472
division, 154
properties of, 154
Frobenius homomorphism, 164
ideal
finitely generated (or fg), 180
irreducible, 184, 550
nilradical, 170
primary, 172
prime ideal, 154
radical of, 172
idempotent, 172
Jacobson radical of, 172
Japanese, 589
Krull dimension of, 472, 534
local, 164, 212, 472
equicharacteristic, 581
localization
construction of, 172
localization of, 173
maximal ideal, 154
Maximal Principle, 180
minimal principle, 184
multiplicative set, 168
exclusion and prime ideals, 169
exclusion by (of), 168
saturated, 185, 186
Nagata ring, 590
nilpotent element, 169
nilradical, 170
Noetherian, *see also* Noetherian ring
polynomial ring
evaluation, 200
reduced, 170
relation of division and ideals, 154
semi-local, 498
Universally Japanese, 589
valuation ring, 496
zero divisor, 169
commutative ring homomorphism
evaluation, 191
Comparison Theorem
injective resolutions, 723
projective resolutions, 747
complementary basis, 469

completion of a ring relative to an ideal, 586
complex
acyclic resolution, 823
deleted, 823
complex number
irrational, 3
algebraic, 3
transcendental, 3
composition series
of modules, 557
compositum of fields, 351, 383
cone
over a diagram to a functor with apex, 807
product
limit, 809
under a diagram to a functor with nadir, 808
congruence modulo m, 31
conjugation, 57
Constructibility Criterion, 315
refined form, 369
Constructible points, 312
Construction problems
circles, 312
Constuction problems
lines, 312
convex set, 513
coproduct, 684
of modules, 222
Correspondence Principle, 66
alternate form, 67
modules, 224
rings, 158
coset, 53
counterexample
minimal, 10
counting norm, 507
Cramer's Rule, 704
crossed homomorphism, 380
principal, 380
crossed product algebra, 627
cube, 103
cycle, 129
length of, 129
cyclic algebra
characterization of split of prime degree, 612
maximal subfield of, 611
cyclic group, 44
cyclic groups
classification, 51
classification of subgroups, 52

DCC, 183
Dedekind domain, 366
class group, 476
conductor, 481
discrete valuation ring, 472
discriminant, 487
division of ideals, 475
extension
unramified, 481
factorization of ideals, 475
fractional ideal in, 475
greatest common division of ideals, 475
invertible ideal, 475
prime ideal
decomposition field of, 484
decomposition group of, 484
inertia field of, 485
inertia group of, 485
inertia index of, 479
lying over, 479
ramification index of, 479
ramified, 481
splits completely, 481
splitting behavior of, 479
totally ramified, 481
unramified, 481
relatively prime ideals, 475
Dedekind's Lemma, 325
Dedekind's Modular Law, 86
Dedekind's Theorem on Ramification, 492
degree
homogeneous, 398
of a field extension, 290
of a finite dimensional central simple algebra, 616
of a purely inseparable field extension, 541
of a separable field extension, 541
degree of a field extension, 293
dense action, 649
depth
of a ring, 578
derivation, 623, 699
inner, 623
derivative
of a polynomial, 321
derivative (formal), 205
descending chain condition, 183, 242
determinant, 256, 702
Vandermonde, 390
determinant map, 51
diagonal map, 686
diagram chasing, 227
diagram of shape $\mathcal{J}$, 806

dihedral group, 45
dimension
 flat, 771
 injective, 733
 left global injective, 733
 left global projective, 752
 of an affine variety, 534
 projective, 752
 right weak, 771
Dimension Shifting, 829
dimension shifting, 733
diophantine equation, 19
 linear, 25
direct limit, 636, 759
direct product
 of abelian groups, 74
 of groups, 47, 69
 of modules, 222
 universal property, 228
direct sum
 external, 74, 222
 internal, 74, 222
 of abelian groups, 74
 universal property, 227
Dirichlet product, 48, 357
Dirichlet Unit Theorem, 517
Dirichlet's Pigeonhole Principle, 49, 52
Dirichlet's Theorem on Primes in an Arithmetic Progression, *see also* primes in an arithmetic progression, 360
 special case, 360
Dirichlet-Chevalley-Hasse Unit Theorem, 520
discrete subgroup, 510
discrete valuation ring, 472
discriminant
 field extension, 384
 of a polynomial, 390
distributative laws, 147
divisible group, 79
division algebra
 algebraic, 401
 algebraic over a finite field, 401
 classification over the reals, 608
 finite dimensional, 608
division algorithm, 17, 181
division algorithm (general form)
 polynomials, 184
division ring, 43, 148, 608
 center of, 398
dodecahedron, 103
domain
 Bésout, 183
 Bézout, 497, 499
 Dedekind, *see also* Dedekind domain, *see also* Dedekind domain
 division algorithm, 181
 Euclid's Argument, 179
 euclidean, *see also* euclidean domain
 Gaussian integers, 186
 GCD-domain, 183
 greatest common divisor (gcd), 178
 integrally closed, 183
 irreducible element, 177
 non-commutative, 609
 normal, 183, 467, 537
 prime element, 177
 principal ideal domain (PID), 153
 quotient field
  construction of, 160
 quotient field of, 160
 reducible element, 177
 relatively prime elements, 178
 unique factorization domain or (UFD), *see also* UFD
 valuation, 183
domain (integral domain), 48, 148
 associate, 154
dominant map, 530
domination of local rings, 500
Double Centralizer Theorem, 616
dual basis, 233
duality, 682

eigenspace, 281
eigenvalue, 269
eigenvector, 269
Eisenstein's Criterion, 208
Eisenstein-Hermite Theorem, 515
element
 idempotent, 599
 integral, 465
 norm of, 375
 primitive nth root of unity, 333
 trace of, 375
elementary divisors, 277
embed, 36
embedded prime, 548
 of a module, 554
endomorphism, 221, 223
 of a vector space, 112
endomorphism ring, 112, 221, 223
entire function, 433
epic, 50
epimorphism, 50, 150, 222

split, 228
equivalence class, 29
representative of, 30
equivalence relation, 27
equivalence class under, 29
equivalence relations
examples, 28, 29
euclicean function
normed, 182
Euclid's Argument, 179
Euclid's Lemma, 20
for a PID, 178
general form, 13, 20
Euclidean Algorithm, 20
Euclidean Constructibility Problems
construction of a regular n-gon, 317
doubling of the cube, 317
squaring of the circle, 317
trisection of an angle, 316
Euclidean Construction Problems, 312
euclidean domain, 181
euclidean function, 181
examples, 182
Gaussian integers, 187
euclidean domains, 177
euclidean function, 181
strong, 181
euler ϕ-function, 36
Euler φ-function, 35
Euler Formulas, 23
Euler products, 23
Euler's Criterion, 361
Euler's Equation about four squares, 193
Euler's Formula, 189
Euler's Theorem, 55
evaluation, 200
evaluation map, 113, 221
exact
left, 714, 716
right, 757
exact sequence, 225
short exact, 226
extension of scalars, 453

factor group, 64
factor ring (by an ideal), 157
factor set, 627
normalized, 628
principal, 628
Feit-Thompson Theorem, 136
Fermat descent, 194
Fermat number, 318
Fermat prime, 318
Fermat's Little Theorem, 35, 55, 360
Fermat's Theorem on sums of two squares, 188
FFR, 782
fiber, 37
Fiber Dimension Theorem, 587
field, 25, 28, 43, 148
C_n-field, 591
F-automorphism, 291
algebraic closure of, 204, 296
existence of, 310
uniqueness of, 311
algebraic number field, 468
algebraically closed, 204, 296, 309
subfields of finite codimension, 456
algebraically closed field, 112
base field, 289
composite, 698
degree of extension, 290
element
algebraic, 292
algebraic - characterization of, 294
transcendental, 292
transcendental - characterization of, 294
element of
purely inseparable, 323
separable, 322
elements
algebraically independent over, 298
elments of
algebraically dependent, 299
euclidean, 448
extension, 290
finite extension of, 290
finite multiplicative subgroups in, 203
fixed field, 291
fixed field of, 327
formally real, 447
global, 473
intermediate field, 289
nontrivial extension, 290
of complex constructible numbers, 316
ordered, 449
existence of a real closure, 451
extension of, 449
real closure with respect to an ordering of, 449
uniqueness of real closures, 454
perfect, 322, 323, 342
prime field, 163
prime subfield, 163
pythagorean, 448

quadratically closed, 448
quasi-algebraically closed, 591
real closed, 447
real closed with respect to an ordering, 449
root field of a polynomial, 301
separable closure of, 383
separable element over
characterization of, 341
splitting field, 300
existence, 301
uniqueness of, 303
splitting field over
existence and uniqueness (general case), 343
tower, 289
field extension
abelian, 333, 359, 369
algebraic, 295
splitting field, 311
tower of, 295
algebraic closure of, 309, 336
and solvable Galois group, 370
characterization of normal (general case), 343
conjugate ones, 338
cyclic, 359, 369, 376
cyclotomic, 333
degree of, 293
finite
characterization of being Galois, 342
characterization of being normal and separable, 342
characterization when Galois, 331
characterization when normal, 337
degree vs order of Galois group, 328
finite Galois, 328
finitely generated, 299
Galois, 328
characterization of (general case), 352
exponent of, 381
Galois (general), 328, 351
Galois and solvable, 377
Galois group of, 304
Kummer, 381
norm of, 373
normal (finite), 336
characterization of normal subextensions, 338
normal (general), 336
normal and separable
characterization of (general case), 352
normal closure (finite), 338
normal closure of (finite case)
existence and uniqueness of, 338
purely inseparable, 323
purely transcendental, 393
radical, 368
separable, 322
separable (finite)
characterization of, 339
simple (or primitive), 291
splitting field, 300
splitting field (of a set of polynomials), 336
splitting field of a seat of polynomials, 311
square root tower, 315
trace of, 373
transcendece degree of, 299
transcendence basis of, 298
transcendental, 295
field homomorphism, 150
field homomorphisms
fixed field of, 327
field of fractions
seequotient field, 147
field of quotients, 160
finite extension of fields, 290
First Isomorphism of Sets
alternate version, 38
First Isomorphism Theorem
of Sets, 37
First Isomorphism Theorem of Sets, 52
Five Lemma, 227
fixed field, 291
formally real ring, 461
Frattini Argument, 122
free abelian group, 92
free group, 92
word in, 94
free module, 230
basis, 230
coordinate, 232
dual basis, 233
linear independent set, 230
linearly dependent set, 230
rank of, 234
spanning set, 230
standard basis, 231, 233
free modules
homomorphism of
invariants factors of, 258
Smith Normal Form of, 258
free presentation, 259
free resolution, 259
Frobenius automorphism, 336, 485
Frobenius homomorphism, 164

Frobenius Reciprocity, 671
Frobenius' Arithmetic Lemma, 664
Frobenius' Theorem, 608
function, 27
  n-linear, 256
  alternating, 256
  arithmentic
    examples of, 367
  arithmetic, 357
  bijective, 6
  fiber of, 37
  graph of, 38
  image of, 6
  injective, 6
  Möbius, 358
  multilinear, 256
  multiplicative, 357
  polynomial, 202, 420
  preimage, 37
  restriction of, 9
  surjective, 6
functor, 789
  acyclic object of, 827
  additive, 818
  adjunction, 803
  bifunctor, 791
  coeffaceable, 826
  constant, 807
  contravariant, 789
  diagonal, 807
  diagonal of shape, 806
  direct image, 806
  effaceable, 826
  embedding
    full, 793
  embedding , 793
  essentially surjective on objects, 793
  exact, 821
  faithul, 793
  forgetful, 789
  full, 793
  fully faithful, 793
  isomorphism, 789
  left exact, 821
  presheaf, 790
  projection, 791
  quotient, 789
  represenable, 796
  representable, 796
  right exact, 821
  set functor
    univeral element of, 801
  Yoneda functor, 797, 798
functors
  left derived, 829
  right derived, 827
    cohomology groups of, 827
  set valued, 796
fundamental domain, 509
Fundamental Theorem of Algebra, 112, 204, 309, 349, 451
Fundamental Theorem of Arithmetic, 5, 21, 89
Fundamental Theorem of fg Modules over a PID, Form I, 260
Fundamental Theorem of fg Modules over a PID, Form II, 267
Fundamental Theorem of Finite Abelian Groups, 73
Fundamental Theorem of Finite Abelian Groups (Alternate Form), 73
Fundamental Theorem of Finitely Generated Abelian Groups, 77
Fundamental Theorem of Free Modules, 232
Fundamental Theorem of Galois Theory, 344
  general case, 356
Fundamental Theorem of Symmetric Polynomials, 423
fundamental unit, 496
funtor
  adjoint, 802

Galois Correspondence, 345
Galois group, 304, 328
  absolute, 383
  element moved, 328
  of a polynomial, 304
  of irreducible polynomial
    transitivity of, 305
Galois' Theorem, 388
Galois' viewpoint, 331
gamma function, 514
Gauss sum, 362
gauss sum
  sign of, 865
Gauss' Lemma, 207
Gaussian integers, 149, 186
gcd, 18, 178, 205
  properties of, 20
gdimension
  weak, 771
Gelfond-Schneider Theorem, 444
General Division Algorithm, 201
general linear group, 41, 46
general quaternion group

properties, 651
germ, 175
global field, 473
Going Down Theorem, 543
graded ring, 580
associate to an ideal, 580
homogeneous element of, 580
homomorphism of, 580
isomorphism of, 580
Great Trick, 57, 58
greatest common divisor, 18, 178
properties of, 20
greatest common divisor (or gcd), 205
greatest integer function, 13
properties of, 13
Green-Tao Theorem, 190
group, 41, 59
p-complement, 127
p-group, 72, 107
p-primary, 78
(set) second isomorphism counting, 68
abelian, 42
abelian of order p^2q, 122
abelianization, 70
acting on a set, *see also* group action
additive, 42
automorphism, 223
automorphism group of, 58
bounded period, 675
center of, 59, 107
properties, 60
centerless, 138
cohomology group, 380, 628
conjugacy class, 107
conjugation, 107
conjugation by an element, 57
cyclic group, 44
divisible, 79
element
commutator, 61, 70
of finite order, 49
elementary p-group, 123, 386, 647
example of (see also specific group)
alternating group, 59
automorphism group, 113
automorphism group of a vector space, 46
circle group, 44
diagonal group, 46
dihedral, 45
general linear group, 46
group of all permutations, 43
Klein 4-group, 47
Lorenz group, 46
of nth roots of unity, 44
orthogonal group, 46
projective special linear group, 135
quaternion, 45
set of units in a ring, 43
special orthogonal group, 46
special unitary group, 46
strictly upper triangular group, 46
symmetric group, 43
symplectic group, 46
unitary group, 46
upper triangular group, 46
exponent of, 381
factor group, 64
finite, 49
finitely generated, 44
finitely presented, 100
free, 92
free abelian group, 74
Galois group, 304, 328
general linear group, 41
general quaternion group, 651
generator, 45
generators, 45
generators for, 44
infinite, 49
inner automorphism group of, 58
left transversal, 54
linear, 66
minimal normal subgroup, 386
nilpotent, 90
class of, 90
finite, 122
normalizer of a subset, 110
of affine transformations, 387
of order $2n$, n odd, 137
of order p^2q, 122
of order p^2q^2, 122
of order p^aq, 121
of order pqr, 122
of principle fractional ideals, 476
order of, 49
periodic, 677
permutation group, 65
polycyclic, 85, 369
presentation of, 95
projective general linear group, 141
projective linear group, 394
projective special linear group, 141
pullback action, 113
quotient group, 64

relation, 45
relations, 45
relations in, 45
representation of, *see also* representation
series
abelian, 85
ascending central series of, 90
central series, 90
characteristic, 86
composition, 85
cyclic, 85
descending central series of, 90
equivalent, 88
length, 89
links, 89
normal, 85
proper refinement, 88
refinement, 88
subnormal, 85
simple, 58, 134, 135
simple group, 133
solvable, 85, 369
properties, 85
stabilizer subgroup, 43
subgroup, 43, 44, 48
nth derived, 86
centralizer of an element, 107
characteristic, 60, 61, 118
commutator, 61, 70
derived, 61, 85
index of, 54
isotropy subgroup, 102
left coset of, 53
minimal normal, 123
normal, 58
normalizer, 109
of smallest prime index, 65
proper, 55
right cosets of, 54
Sylow subgroup, 115
subset closed under $\cdot$, 48
subset generating, 44
torsion, 75, 657
torsion element, 75
torsion-free, 75
transitive, 43
trivial group, 43, 54
group action, 101
centralizer, 107
class equation, 107
conjugacy class, 110
conjugacy class of an element, 107
conjugation, 110
conjugation action, 107
doubly transitive action, 104
fixed point, 103
fixed point set, 103, 115
isotropy subgroup, 102
moved element, 133
orbit
one point orbit, 103
orbit under, 102
restriction to a subgroup, 101
right, 111
stabilizer of a point, 102
system of representatives, 102
transitive action, 104
translation, 111
group of all permutations, 43
group ring, 643
groups
external direct product of, 47
free product of, 97
free product with amalgamation of, 98
internal direct product of, 69
isomorphic, 50
semidirect product of groups, 61
semidirect product of subgroups, 70
with cyclic Sylow 2-group, 137

Hall's Theorem, 125
Hamiltonian quaternions, 193
height
of a prime ideal, 533
Hilbert Basis Theorem, 180, 211, 242
Hilbert class field, 492, 523
Hilbert Irreducibility Theorem, 412
Hilbert Nullstellensatz, 211
Algebraic Form, 567
strong, 212
weak, 211
Hilbert Nullstellensatz (Strong Form), 247, 566
Hilbert Nullstellensatz (Weak Form), 246, 566
Hilbert ring, 567
Hilbert Syzygy Theorem, 775
Hilbert Theorem 90, 375
Hilbert Theorem 90, 380
Hilbert's 17th Problem, 463
Hilbert's Seventh Problem, 5, 432
homogeneous element, 580
homological ∂-functor, 826
homological *partial*-functor
universal, 826
homology

long exact sequence, 710, 729, 730, 750
naturality, 711
homomorphism
F-algebra, 301
automorphism, 301
algebra, 151, 301, 692
automorphism, 301
epimorphism, 301
involution, 302
isomorphism, 301
monomorphism, 301
automorphism, 58
bijective, 50
canonical, 689
coimage, 683
cokernel, 226, 682
epimorphism, 33, 50
field, 150
Frobenius, 164
group, 50
bijective, 52
properties of, 50
image, 683
image of, 50, 150
isomorphism, 50
kernel, 682
kernel of, 50, 150
linear, 222
linear transformation, 301
module, 222
epimorphism, 222
extend linearly, 232
isomorphism, 222
monomorphism, 222
monomorphism, 50
natural, 756
ring, 33, 149
epimorphism, 150
extension of , 301
finite, 469, 537
integral, 468, 537
isomorphism, 35
lift of, 301
monomorphism, 150
of finite type, 469, 537
rng, 150
trivial, 51
homorphism
algebra
epimorphism, 151
isomorphism, 151
monomorphism, 151
linear, 301
homotopy, 98
of loops, 98
Horseshoe Lemma, 727, 748
Hurwitz Theorem, 653
Hurwitz-Radon Theorem, 654
hypersurface, 211

icosahedron, 103
ideal, 147
2-sided, 151
associated prime ideal of, 548
discriminant, 487
embedded prime ideals of, 548
finitely generated (fg), 180
fractional, 475
generated by, 152
graded, 694
homogeneous, 694
invertible, 475
irreducible, 184, 550
isolated prime ideal of, 548
left, 151
left radical, 401
maximal ideal, 154
maximal left (right), 155
minimal left, 597
minimal prime containing, 249
minimal prime ideal of, 548
primary, 172, 545
prime, 147
minimal, 529
prime ideal, 154
characterization of, 155
principal, 153
principal ideal generated by, 152
radical, 528
radical of, 172, 212, 247
right, 151
right radical, 401
trivial, 151
unit ideal, 151
ideal colon, 546
ideal norm, 505
idempotent, 172, 599
central, 401, 606
orthogonal, 599
trivial, 599
Idempotent Theorem, 663
IDP, 234
image, 50, 150, 683
incomparable elements, 165

induction
First Principle of Finite Induction, 10
Second Principle of Finite Induction, 10
induction hypothesis, 10
induction step, 10
inertial index, 479
Infinite descent, 177
infinite extension of fields, 290
injective equivalence, 731
inner automorphism group of a group, 58
integer
algebraic, 429
composite, 12
division of, 3
Fermat number, 318
Fermat prime, 318
perfect, 4, 26
Euclid/Euler Theorem, 4
prime, 3
standard factorization, 5
standard representation, 5, 21
integers
congruence modulo m, 31
division algorithm
modified, 25
division of
properties, 17
relatively prime, 12, 18
ring mod m, 31
integers mod m
unit, 34
unity or 1, 32
zero, 32
integers modulo m
characterization of units in, 34
integral basis, 433, 471
integral element, 465
integrally closed ring, 467
internal direct product
groups, 121
invariant dimension property, 234
invariant factors, 258, 260
invariant subspace, 221
Inverse Galois Problem, 348
inverse limit, 636, 764
involution, 186, 192, 302
irreducible component, 249, 532
irreducible decomposition, 249
irreducible decomposition of an ideal, 550
irreducible element, 147, 177
irredundant primary decomposition, 547
isometry, 514
isomorphic
rings, 150
isomorphism, 50, 222
of rings, 150
ring, 35
isomorphism theorem
(set) second isomorphism theorem counting version, 68
isomorphism theorems
first isomorphism theorem for groups, 63
first isomorphism theorem for modules, 224
first isomorphism theorem for rings, 157
first isomorphism theorem for sets, 37
second isomorphism theorem for groups, 68
second isomorphism theorem for modules, 224
second isomorphism theorem for rings, 164
third isomorphism theorem for groups, 67
third isomorphism theorem for modules, 224
third isomorphism theorem for rings, 158

Jacobian matrix, 585
Jacobson Density Theorem, 649
Jacobson radical, 172, 538
Jacobson ring, 567
Jacobson's Theorem, 401
Japanese ring, 589
Jordan canonical form, 112, 277
Jordan Canonical Form Theorem, 277
Jordan Decomposition
additive form, 282
multiplicative form for invertible matrices, 282
Jordan-Hölder Theorem, 89
for modules, 558

Kaplansky's Theorem, 185
Kaplansky's Theorem on projective modules, 737
kernel, 50, 150, 682
Key Observation, 21
Key Trick, 21
koszul complex, 697
homology, 698
with coefficients, 697
Kronecker's Criterion, 411
Kronecker's Theorem, 203, 300
Kronecker-Weber Theorem, 333, 361
special case of a quadratic extension, 363
Krull dimension, 472, 534
Krull Intersection Theorem, 242
Krull's Conjecture, 781
Krull's Intersection Theorem, 576

Krull's Theorem, 169
Krull's Theorem on Algebraic Galois Extensions, 354
Krull-Akizuki Theorem, 473
Krull-Akuzuki Theorem, 561
Kummer-Dedekind Theorem, 481

Lüroth's Theorem, 394
Lagrange Interpolation, 406
Lagrange's Theorem, 54
Lagrange's theorem on roots of a polynomial, 202
Lagrange's Theorem on sums of four squares, 194
Lambek's Theorem, 766
Lang Homomorphism Theorem, 459
Lang's Theorem, 442
Lang's Theorem on C_n-fields, 593
Laplace expansion, 704
Lasker-Noether Theorem, 551
lattice
  full, 509
  lattice basis, 509
latttice, 509
Law of Quadratic Reciprocity, 190, 363
lcm, 26
least common multiple, 26
left translation, 54
Legendre symbol, 189, 361
  properties of, 361
lexicographic order, 165, 422
  on a finite dimensional real vector space, 166, 170
limit
  cone, 808
  of cones, 808
Lindemann's Theorem, 292, 424
Lindemann-WeierstraßTheorem, 429
linear operator
  diagonalizable, 278
  eigenspace of, 281
  nilpotent, 282
  semisimple, 281, 283
  triangularizable, 278
  unipotent, 282
Liouville number, 292, 418
Liouville's Theorem, 417
local ring, 164, 175, 212, 472, 483, 534
  regular, 539, 575
localization, 173
  at a prime ideal, 532
  at powers of an element, 532
  of a module, 535, 686
locally ringed space, 175
logarithmic map, 516
loop, 98

Möbius function, 358
Möbius Inversion Formula, 358
Möbius Inversion Theorem
  general form, 367
Mantra for cosets, 54
Mantra of G-actions, 102
Mantra of Equivalence Relations, 30, 54
map, 6
  n-(multi)linear, 686
  of G-sets, 104
  alternating, 695
  associated, 530
  balanced, 755
    biadditive, 755
  bilinear, 686
  canonical surjection, 29
  diagonal map, 686
  dominant, 530
  equivariant, 104
  evaluation, 191
  left translation, 54
  multiplicative, 187
  sgn, 132
  signum, 132
  surjective
    summary about, 37
  symmetric, 694
  translation, 387
map, induced by, 37
Maschke's Theorem, 646
matrices
  equivalent, 28, 253
  Jordan canonical form, 112
  rational canonical form, 112
  similarity of, 28
matrix
  classical adjoint, 704
  cofactor, 704
  companion, 139, 270
  diagonalizable, 277, 281
  elementary
    Type I, 255
    Type II, 255
  elementary divisors of, 277
  invariant factors of, 272
  Jordan block, 277
  Jordan canonical form, 253

Jordan canonical form of, 277
minor of, 257
permutation, 59
rational canonical form, 253
rational canonical form of, 272
Smith Normal Form, 253, 254
triangularizable, 278
matrix representation of a linear map, 59
matrix ring, 28
maximal counterexample, 181
maximal ideal, 154
Maximal Principle, 180, 239
and Zorn's Lemma, 180
Maximum Modulus Principle, 434
maxmial subfield
self centralizing, 626
meromorphic function, 433
transcendental, 434
pole, 433
meromorphic functions
linear independence of, 434
Mersenne number, 4
Mersenne prime, 4
minimal basis, 487
minimal counterexample, 10
minimal element, 9
minimal left ideal, 597
minimal polynomial, 270
minimal prime
of a module, 554
minimal prime ideal, 170, 529
Minimal Principle, 184
Minimum Principle, 242
Minkowski bound, 521
Minkowski Lattice Point Theorem, 513
minor, 257
Mitchel Embedding Theorem, 822
module, 219
$\mathfrak{p}$-primary, 744
p-primary, 266
absolutely irreducible, 667
acyclic resolution, 723
acyclic resolution of, 746
Artinian, 242
ascending chain condition, 239
associated prime, 552
completely reducible, 597
coproduct, 684
cyclic, 222
depth of, 578
descending chain condition, 242
direct product
universal property, 228
direct sum
universal property, 227
direct summand of, 232
divisible, 722
dual, 233, 739
embedded prime of, 554
factor module, 221
finite length of, 557
finitely generated, 222
finitely presented, 241, 739
flat, 692, 765
free, 230
free resolution, 259
generators, 222
indecomposable, 605
induced, 669
injective, 229, 717
Baer Criterion, 230, 719
injective resolution of, 723
invariant factors of, 260
irreducible, 557, 597
length of, 558
localization, 686
Maximal Principle, 239
minimal (or isolated prime) of, 554
Minimum Principle, 242
Noetherian, 239
presentation, 739
product, 683
projective, 235, 734
local rank, 746
projective resolution of, 746
quotient module, 221
restricted, 669
simple, 227, 557, 597
stably free, 782
submodule of, 220
support of, 555
torsion, 253, 265
torsion element, 265
torsion-free, 242, 265
unitary, 220
modules
direct limit of, 636, 759
direct product, 222
direct sum (external) of, 222
direct sum (internal) of, 222
inverse limit of, 636, 764
isomorphic, 222
product, 222
tensor product of, 686

monic, 50
monoid, 42
monomorphism, 50, 150, 222
split, 228
Morita Equivalence, 794
morphism
coequalizer, 819
coimage of, 819
cokernel of, 819
equalizer, 809
kernel, 810
fiber product, 810
identity, 786
image of, 819
inverse (or projective) limit, 811
kernel of, 819
pullback, 810
variable element parametrized by, 796
morphisms
parallel pair, 809
multinomial coefficients, 15
multiplicative map, 187
multiplicative set, 168
examples, 168
saturated, 185, 186

Nagata ring, 590
Nagata's Lemma, 781
Nakayama's Lemma, 483, 539, 776
natural homomorphism, 756
natural transformation, 794
components, 794
cone over with apex, 807
constant natural, 807
counit, 804
induced, 797
isomorphism, 794
unit, 804
universal mapping property, 804
negative element, 449
nilpotent element, 169
nilradical, 170
characterization of, 170
nodal cubic, 565
Noetherian domain
and products of irreducible elements, 181
Noetherian induction, 181
Noetherian ring, 180, 240
properties, 180
Noetherian space, 533
norm, 645
counting, 507
ideal, 505
of a finite extension, 541
of a finite Galois extension, 373
of a finite separable extension, 375
of an element, 375
reduced, 621
norm map on $\mathbb{C}$, 186
norm map on the quaternions, 192
normal basis, 385
normal closure of a subgroup, 95
normalbasisthm, 386
normic form, 591
number
algebraic, 429
Liouville, 418

object
quotient, 819
suboject, 818
octahedron, 103
octonians, 193
opposite algebra, 614
Orbit Decomposition Theorem, 104
ordering, 447
orderings
existence of, 449
orthogonal group, 46
orthogonality relations, 659

pairing of groups, 381
parallelopiped, 509
partially ordered set, 165, 833
directed, 760
inductive
strongly, 833
upper bound
least, 833
partition, 30
Pell's equation, 495, 870
perfect field, 322
permutation, 43
3-cycle, 132
cycle type of, 138
even, 60, 131
odd, 60, 131
regular, 138
transposition, 131
permutation group, 65
fixed points of, 43, 130
full cycle decomposition, 130
transposition, 131
permutation matrices, 59

permutation representation, 115
PID, 153
  Euclid's Lemma, 178
  examples of, 153
  gcd in, 178
  left, 609
  PID is a UFD, 181
  prime ideals in, 155
Poincaré's Lemma, 57
point
  constructible
    complex, 316
  fixed point, 326
points
  constructible, 312
    complex, 313
  not constructible, 312
Pointwise Adjointness Theorem, 805
polynomial, 3
  characteristic polynomial for a linear polynomial, 269
  companion matrix of, 139, 270
  constant, 199
  content of, 206
  criterion for solvable by radicals, 378
  cyclotomic, 306, 358
    if arbitrary degree, 308
    of degree 6, 306
    of prime power order, 306
  degree of, 199
  derivative (formal), 205
  derivative of, 321
  discriminant of, 390
  elementary symmetric, 347, 423
  Galois group of, 304, 372
  inseparable, 322
  irreducible
    separable, 322
    splitting field and Galois group of, 332
    transitivity of Galois group on roots, 305
  irreducible or minimal, 293
  leading coefficient, 199, 422
  leading term, 422
  linear, 204
  minimal for a linear operator, 270
  minimal of an algebraic element, 293
  minimal or irreducible, 293
  monic, 187, 199
  monomial, 422
  multiple root
    criterion for, 321
  multiple root of, 205, 250, 277, 305
  multiplicity of a root, 202
  of homogeneous degree, 398
  primitive, 206
  quadratic, 204, 398
  reducible, 317
  root
    multiple, 321
    multiplicity of, 321
    simple, 321
  root field of, 301
  root of, 201
  separable, 322
  solvable by radicals, 372
    and solvable Galois group, 372
  split, 277
  splits over a field extension, 300
  splitting field of, 300
    existence, 301
    uniqueness, 303
  symmetric, 422
  total degree, 250, 397, 422
polynomial function, 202
polynomial ring, 25, 199
  in n variables, 199
polynomials
  division algorithm, 25
  and content, 206
  general division algorithm, 184, 201
  irreducible over the reals, 204
poset, 165, 833
  chain in, 165, 833
  comparable elements, 165
  incomparable elements, 165
  inductive, 166, 833
    strongly, 833
  maximal element, 166, 833
  upper bound, 166, 833
  well-ordered, 834
positive element, 449
positive semi-definite function, 463
power series
  regular element of, 212
power set, 110
preimage, 66
  properties of, 66
preordering, 447
presentation
  minimal, 779
presheaf, 536
primary decomposition, 547
  irredundant, 547
  of a module, 557

Primary Decomposition Theorem, 266
primary decompostiom
  irredundant
    of a module, 557
primary ideal, 172, 545
  $\mathfrak{p}$-primary, 545
prime
  associated
    of a module, 552
  embedded) of a module, 554
  inert, 366
  minimal (or isolated) of a module, 554
  ramified, 366, 479
  splits completely, 366
Prime Avoidance Lemma, 157, 541
prime element, 177
prime field, 163
prime ideal, 154
  embedded of an ideal, 548
  height of, 185, 533
  isolated prime of an ideal, 548
  minimal, 170
  minimal prime ideal of an ideal, 548
  unramified, 481
prime ideals
  in the integers, 155
  saturated chain, 568
Prime Number Theorem, 4
primes
  infinite, 520
primes in an arithmetic progression, 190
  1 mod 4, 189
  3 mod 4, 35
Primitive Element Theorem, 348
primitive root, 863
principal ideal domain, 153
Principal Ideal Theorem, 571
product, 683
  cofiber, 731
  fiber, 750
  of modules, 222
projection, 38
projective general linear group, 141
projective special linear group, 141
pseudo-polynomial, 214
pullback, 221, 750
purely inseparable degree, 541
pushout, 731
pythagorean closure, 449

quadratic form, 398, 514
  discriminant of, 514
  positive definite, 514
Quadratic Reciprocity, 190
quadratic reciprocity, 363
  supplements of, 364
quasi-algebraically closed field, 591
quaternion algebra
  generalized, 594
quaternion group, 45, 192
quaternions, 192, 193, 202
  conjugation, 192
  pure, 196
quaternions (general), 196
  conjugate of an element, 196
quotient field, 147, 160
quotient group, 64
quotient ring, 157

Rabinowitch Trick, 247
radical
  Jacobson, 172, 538, 776
  nilradical, 170
  of an ideal, 172
radical ideal, 528
ramification index, 479
ramified prime, 479
rational canonical form, 112
Rational Canonical Form Theorem, 271
real closed field, 447
  relative to an ordering, 449
reduced norm, 621
reduced polynomial, 621
reduced trace, 621
reducible element, 177
regular local ring, 539
regular representation, 65, 136, 644, 655
regular value (of an affine plane curve), 250
regulator, 519
relation, 27
  reflexitivity, 27
  symmetry, 27
  transitivity, 27
Remainder Theorem, 201
representation, 644
  characterization over algebraically closed field, 646, 651
  degree of, 644
  equivalence of, 644
  faithful, 644
  general quaternion group, 652
  group ring, 644
  irreducible, 644
  linear, 647

regular, 644, 655
structure afforded by, 644
trivial, 645
residue class modulo m, 31
resolution
acyclic, 723, 746
finite free, 782
flat, 766
injective, 723
projective, 746
Riemann surface, 251
ring, 25, 32, 42, 147
1, 32
catenary, 570
center of, 151, 192, 399
characteristic of, 159
Cohen-Macaulay, 579
coherent, 241
commutative, 32, 42, 148
depth of, 578
division ring, 43, 148
domain, 48, 148
dual numbers, 576
endomorphism, 223
endomorphism ring, 221
example
endomorphism ring, 112
existence of maximal ideals, 168
factor ring of, 157
field, 148
formal power series over, 149
group of units, 43, 148
group ring, 205, 643
ideal, 151
maximal ideal, 154
maximal left (right), 154
ideal (2-sided), 151
idempotent, 172
integrally closed, 467
integrally closed in, 467
left hereditary, 737
left Noetherian, 720
left semi-hereditary, 746
linear combination of elements, 152
local, 483, 534
monoid ring, 205, 643
multiplicative inverse in, 34
nilpotent element, 169
Noetherian, 240
normal, 467
of algebraic integers, 366
of algebraic integers (in), 468
polynomial ring, 199
in n variables, 199
properties of, 148
quotient ring, 157
regular, 781
right Noetherian, 720
ring endomorphism, 221
semi-local, 542
semisimple, 597
simple, 151
matrix ring over a field, 156
subring of, 150
subrng of, 150
total quotient ring, 175
total ring of fractions, 175
trivial ring, 43, 148
twisted Laurent series ring, 613
twisted polynomial ring, 609
twisted power series ring, 613
units
in $\mathbb{Z}/m\mathbb{Z}$, 34
units in, 34
universally catenary, 570
zero, 32
zero ring, 43
ring antiautomorphism, 192
ring antihomomorphism, 192
ring epimorphism, 33
ring extension, 465
integral closure in, 467
ring homomorphism, 33
quasi-finite, 542
ring of algebraic integers, 468
ringed space, 815
local, 816
rings
examples, 32
rng, 148
example without maximal ideals, 171
subrng of, 150
root function, 403
root of unity
primitive, 308
primitive pth root of unity, 296
Roth's Theorem, 419
RSA code, 55
Russell's Paradox, 7

scalar multiplication, 219
Schanuel's Lemma, 731
scheme
affine, 536, 815

Schneider's Theorem, 436
Schreier Refinement Theorem, 88
Schroeder-Bernstein Theorem, 7, 836
Schur's Lemma, 601
Schur-Zassenhaus Theorem, 128
semi-local, 498
semi-local ring, 542
semi-real ring, 461
separable closure of F in K, 490
separable degree, 541
sequence
  characteristic, 284
  exact, 225, 821
  split exact, 228
  zero, 225
Serre's Theorem on Global Dimension, 779
Serre's Theorem on global dimension, 779
set
  cardinality, 836
  cardinality of, 6
  countable, 6
    facts about, 6
  finite, 6
  indexing, 29
  infinite, 6
  partially ordered, 165, 833
    chain, 165, 833
  partially ordered set
    comparable elements, 165
    incomparable elements, 165
    inductive, 166, 833
    maximal element, 166, 833
    minimal element, 166
    upper bound, 166, 833
  partion of, 30
  power set of, 110, 165
  totally ordered, 165
sets
  cartesian product of, 8, 47
  disjoint union of, 10, 29
  intersection of, 29
  union of, 29
sheaf, 175
  rings, 536
  stalk, 175, 816
signature, 453, 861
signum map, 132
simple point, 585
Six Exponentials Theorem, 445
Skolem-Noether Theorem, 619
Smith Normal Form, 253, 254
Snake Lemma, 708
special linear group, 46
Spectrum of a ring, 247, 527
split exact sequence, 228
splitting type of a prime, 366
square root tower, 315
Square Root Tower Theorem, 315, 348
squares
  sums of four squares, 194
  sums of three squares, 195
  sums of two squares, 188
standard basis, 231, 233
Stickelberger's Criterion, 490
strong euclidean function, *see also* euclidean domain
Strong Hilbert Nullstellensatz, 212
subcategory, 787
subgroup, 43, 48
  p'-subgroup, 127
  centralizer, 66
  characteristic, 60, 83, 118
  commutator, 61, 70
  core of, 47, 60
  derived, 61
  discrete, 510
  Hall, 123
  index, 54
  isotropy subgroup, 102
  left coset of, 53
  maximal, 121
  minimal normal, 123, 386
  normal, 58
  normal closure of, 95
  normalizer, 109
  normalizer of, 48, 60
  pure, 75
  right coset of, 54
  stabilizer, 102
  Sylow, 115
submodule, 220
sums of four squares, 194
sums of two squares, 188
  characterization of, 189
support of a module, 555
Sylow Theorems, 117
  First, 65, 115, 117
  First (general form), 115
  Fourth, 117
  Second, 117
  Third, 117
Sylvester's Lemma on real roots, 454
symmetric
  1-cycle, 130

symmetric group, 43
(full) cycle decomposition, 130
cycle, 129
cycle decomposition, 130
cycles
properties of, 129
disjoint cycles, 130
transpositions and, 131
symmetric group on n letters, 43
symmetric set, 513
system of parameters, 575
regular, 575
system of representative (under a group action), 102
system of representatives (for an equivalence relation), 30

tensor algebra, 693
tensor product, 236, 686
tetrahedron, 103
toplogy
clopen set, 354
topological group
closure of subgroup, 355
topological space, 353
continuous function, 354
irreducible, 529
irreducible component of, 532
irreducible subset of, 529
Noetherian, 533
topology, 353
base, 353
closed set, 353
compact set, 354
connected, 354
connected component, 354
Hausdorff space, 353
open sets, 353
point
base for, 353
fundamental neighborhood system of, 353
neighborhood of, 353
product space, 354
profinite, 354
subbase, 353
subspace, 354
totally disconnected space, 354
Zariski, 248
torsion element, 265
torsion group, 657
totally negative element, 449
totally ordered set, 165
totally positive element, 449
tower of fields, 289
separable
characterization of (finite extension case) , 342
separablity
characterization of (general case), 343
trace, 645
of a finite Galois extension, 373
of a finite separable extension, 375
of a linear transformation, 691
of an element, 375
reduced, 621
trace form, 452
transcendence
of π, 424
of e, 420
transcendence basis of a field extension, 298
transcendental element, 292
transposition, 131
transvection, 140
Tsen's Theorem, 603, 623
Twin Prime Conjecture, 190
Tychonoff's Theorem, 354

UFD, 179
characterization of, 185
Gaussian integers, 187
irreducibles and primes, 179
length of an element, 254
PID, 181
polynomial ring over, 208
polynomials and content over, 206
unique factorization domain, *see also* UFD
universal mapping property, 804
universal property
of quotient fields, 161
of cokernels, 683
of coproducts, 684
of exterior algebras, 696
of free abelian groups, 92, 100
of free groups, 92
of free modules, 232, 681
of free products of groups, 97
of kernels, 682
of localization, 175
of products, 684
of symmetric algebras, 695
of tensor algebras, 694
of tensor products, 687
of vector spaces, 90
universally Japanese ring, 589

Useful Counting Result for groups, 65

valuation ring, 496
Van Dyck's Theorem, 96
van Kampen's Theorem, 99
Vandermonde determinant, 390
Vandermonde matrix, 390, 866
variety, 245, 247, 566
  irreducible, 249
  irreducible component, 249
  irreducible decomposition of, 249
vector space
  basis
    characterization of, 167
    linear independent subset, 166
    spanning sets, 167
  dimension, 89
  extension to a basis, 166
  invariant subspace, 221

Waring Problem, 196
Weak Hilbert Nullstellensatz, 211
Weak Real Nullstellensatz, 463
Wedderburn's Theorem for simple rings, 601
Wedderburn's Theorem on finite division rings, 398
Wedderburn's Theorem on polynomials over a division ring, 639
Weierstraß Preparation Theorem, 213
well-defined, 31, 32
Well-ordering Axiom, 834
Well-Ordering Principle, 9
  modified, 9
Wilson's Theorem, 57
word
  empty word, 94
  reduced word, 94

Yoneda Embedding Theorem, 797
Yoneda functor, 797, 798
Yoneda Lemma, 799

Zariski topology, 248
  basic open set, 529
  geometric, 248
Zariski's Lemma, 245, 566
Zariski's lemma, 503
zero divisor, 169, 553
zero sequence, 225
zeta function, 4, 367
Zorn's Lemma, 166, 833
zyzygy, 751