Math 117: Algebra with Applications

Michael Andrews
UCLA Mathematics Department

June 4, 2016

Contents

1 Rings and fields

1.1 The definition

Consider the following list of properties that some collection of numbers might satisfy.

- (A1) $a + (b + c) = (a + b) + c$ for all a, b, c (associativity of addition)
- (A2) $a + b = b + a$ for all a, b (commutativity of addition)
- (A3) there is an element 0 with the property that $a + 0 = a = 0 + a$ for all a (identity for addition)
- (A4) for each a there is an element $(-a)$ with the property that $a + (-a) = 0 = (-a) + a$ (inverses for addition)
- (M1) $a(bc) = (ab)c$ for all a, b, c (associativity of multiplication)
- (M2) $ab = ba$ for all a, b (commutativity of multiplication)
- (M3) there is an element 1 with the property that $a \cdot 1 = a = 1 \cdot a$ for all a (identity for multiplication)
- (M4) for each $a \neq 0$ there is an element a^{-1} with the property that $a \cdot a^{-1} = a^{-1} \cdot a = 1$ (inverses for multiplication)
- (D) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ for all a, b, c (distributivity)

Definition 1.1.1. A set R equipped with an addition

$$+ : R \times R \longrightarrow R,\ (a, b) \longmapsto a + b$$

and multiplication

$$\cdot : R \times R \longrightarrow R,\ (a, b) \longmapsto a \cdot b$$

satisfying (A1)-(A4), (M1), (M3), and (D) is called a *ring*.

If, in addition, the multiplication satisfies (M2), then R is called a *commutative ring*.

A commutative ring for which (M4) is also satisfied is called a *field*.

1.2 Lots of examples

Example 1.2.1. The integers $\mathbb{Z}$ are a commutative ring but they are not a field since (M4) fails: there is not an *integer* n with the property that

$$2 \cdot n = n \cdot 2 = 1.$$

Example 1.2.2. The rationals $\mathbb{Q}$ are a field. You have known this for a long time.

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 n_2 + m_2 n_1}{n_1 n_2},$$
$$\frac{m_1}{n_1} \cdot \frac{m_2}{n_2} = \frac{m_1 m_2}{n_1 n_2}.$$

Example 1.2.3. The real numbers $\mathbb{R}$ are a field. You have known this for a long time. The reals are actually more confusing though. To think coherently about what a real number is you have to do math 131A.

Example 1.2.4. Fix an $n \in \mathbb{N}$. The $n \times n$ matrices with real entries $M_n(\mathbb{R})$ form a ring. If $n \geq 2$ they are not a commutative ring since

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

(M4) also fails since not all matrices have non-zero determinant.

Example 1.2.5. Fix an $n \in \mathbb{N}$. More generally we can look at $n \times n$ matrices with coefficients in a commutatative ring R: $M_n(R)$. Is the following matrix invertible in $M_2(\mathbb{Z})$?

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Example 1.2.6. If V is a vector space over $\mathbb{R}$. Then the set

$$\mathrm{End}_{\mathbb{R}}(V) = \{T : V \longrightarrow V \ : \ T \text{ is linear}\}$$

can be made into a ring using pointwise addition and composition:

$$(S+T)(v) = Sv + Tv, \ (ST)(v) = S(Tv).$$

If V is finite dimensional then, by choosing a basis, this is basically the same as $M_{\dim V}(\mathbb{R})$. If V is not finite dimensional it is something you may not have encountered before. Kevin will talk about this in discussion.

Example 1.2.7. Suppose $a, b \in \mathbb{R}$ with $a < b$. The continuous real-valued functions on the open interval (a, b) form a commutative ring $C^0(a, b)$ under pointwise addition and multiplication:

$$(f+g)(x) = f(x) + g(x), \ (fg)(x) = f(x)g(x).$$

Is $C^0(a, b)$ a field? (Homework.)

Example 1.2.8. Suppose $a, b \in \mathbb{R}$ with $a < b$. The differentiable real-valued functions on the open interval (a, b) form a commutative ring $C^1(a, b)$ under pointwise addition and multiplication. Is $C^1(a, b)$ a field? (Homework.)

2 The rationals (a discussion from lectures elaborated)

Now just relax. Settle back in your chair. Take a deep breath. Relax your arms. Relax your legs. Relax your nerves. Relax all over. Look at the center of this disk I am holding in my hand. Do not look off. Do not say anything. Keep your mind on my words. Think of nothing else. Gaze right at the center of this disk. Soon your eyes will get heavy, and you wish to close them and go to sleep. Your eyes are getting heavy, very heavy and tired. You are getting very sleepy. Soon you will be sound asleep... Now focus carefully on what I am about to say.

YOU NO LONGER KNOW WHAT A RATIONAL NUMBER IS.

Awake.

I think we've all been really frustrated recently... You have a cookie. You and a friend both want a piece, but there's just no way for you both to enjoy it: only *whole* cookies exist. Thankfully, I just came up with a new number system that is going to change the world and cookie enjoyment forever! Here goes...

Okay. So we're happy with the integers $\mathbb{Z}$. Wow, though... do you remember when only positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$ existed and keeping track of debt was really arduous? Thank goodness that's over...

Consider the set

$$\mathbb{Z} \times \mathbb{N} = \{(m, n) : m \in \mathbb{Z},\ n \in \mathbb{N}\}.$$

I'm going to construct my new numbers by viewing some of these elements as the same.

Definition 2.1. Suppose (m_1, n_1) and (m_2, n_2) are in $\mathbb{Z} \times \mathbb{N}$. We'll say (m_1, n_1) and (m_2, n_2) are *rationally congruent* and write

$$(m_1, n_1) \equiv_{\mathbb{Q}} (m_2, n_2) \quad \text{or} \quad [(m_1, n_1)]_{\mathbb{Q}} = [(m_2, n_2)]_{\mathbb{Q}}$$

if $m_1 n_2 = m_2 n_1$.

Example 2.2. $(1, 2) \equiv_{\mathbb{Q}} (2, 4) \equiv_{\mathbb{Q}} (3, 6) \equiv_{\mathbb{Q}} (4, 8) \equiv_{\mathbb{Q}} \ldots$ or saying the same thing

$$[(1, 2)]_{\mathbb{Q}} = [(2, 4)]_{\mathbb{Q}} = [(3, 6)]_{\mathbb{Q}} = [(4, 8)]_{\mathbb{Q}} = \ldots$$

This is because $1 \cdot 4 = 2 \cdot 2$, $2 \cdot 6 = 3 \cdot 4$, $3 \cdot 8 = 4 \cdot 6$.

Definition 2.3. My new numbers, which I call *the rational numbers* are the set

$$\mathbb{Q} = \{[(m, n)]_{\mathbb{Q}} : (m, n) \in \mathbb{Z} \times \mathbb{N}\}.$$

Remark 2.4. Notice that my new numbers contain the integers. If $[(m_1, 1)]_{\mathbb{Q}} = [(m_2, 1)]_{\mathbb{Q}}$. Then $m_1 = m_1 \cdot 1 = m_2 \cdot 1 = m_2$. So the subset

$$\{[(m, 1)]_{\mathbb{Q}} : m \in \mathbb{Z}\} \subset \mathbb{Q}$$

is a copy of the integers.

I even know how to add and multply my new numbers.

Definition 2.5. I add and multiply rationals using the following formulae.

$$[(m_1, n_1)]_{\mathbb{Q}} + [(m_2, n_2)]_{\mathbb{Q}} = [(m_1 n_2 + m_2 n_1, n_1 n_2)]_{\mathbb{Q}},\ [(m_1, n_1)]_{\mathbb{Q}} \cdot [(m_2, n_2)]_{\mathbb{Q}} = [(m_1 m_2, n_1 n_2)]_{\mathbb{Q}}$$

Let me convince you that makes sense.

Theorem 2.6. *The addition and multiplication just defined make sense.*

Proof. What do we need to show? Suppose that

$$[(m_1, n_1)]_{\mathbb{Q}} = [(\overline{m}_1, \overline{n}_1)]_{\mathbb{Q}},\ [(m_2, n_2)]_{\mathbb{Q}} = [(\overline{m}_2, \overline{n}_2)]_{\mathbb{Q}}. \tag{2.7}$$

We need to show that

$$[(m_1 n_2 + m_2 n_1, n_1 n_2)]_{\mathbb{Q}} = [(\overline{m}_1 \overline{n}_2 + \overline{m}_2 \overline{n}_1, \overline{n}_1 \overline{n}_2)]_{\mathbb{Q}},\ [(m_1 m_2, n_1 n_2)]_{\mathbb{Q}} = [(\overline{m}_1 \overline{m}_2, \overline{n}_1 \overline{n}_2)]_{\mathbb{Q}}.$$

What we're checking is that renaming our rationals does not change the result of adding or multiplying them, since our definition depends very explicitly on the name we're using.

By definition of $[-]_{\mathbb{Q}}$ this means we must show

$$(m_1 n_2 + m_2 n_1) \cdot \overline{n}_1 \overline{n}_2 = (\overline{m}_1 \overline{n}_2 + \overline{m}_2 \overline{n}_1) \cdot n_1 n_2,\ m_1 m_2 \overline{n}_1 \overline{n}_2 = \overline{m}_1 \overline{m}_2 n_1 n_2.$$

Multiplying the first one out, we see that we want the following.

$$m_1 n_2 \overline{n}_1 \overline{n}_2 + m_2 n_1 \overline{n}_1 \overline{n}_2 = \overline{m}_1 \overline{n}_2 n_1 n_2 + \overline{m}_2 \overline{n}_1 n_1 n_2$$

The right hand side is obtained from the left hand side by shifting the bar from the n_1 to the m_1 in the first term and shifting the bar from the n_2 to the m_2 in the second expression. We need $m_1 \overline{n}_1 = \overline{m}_1 n_1$ and $m_2 \overline{n}_2 = \overline{m}_2 n_2$ and this is exactly what 2.7 says, by definition.

The second equality follows by mulitplying together these last two equations. □

I claim that these numbers give what we have been looking for all this time: an example of a field containing the integers; we can finally divide our cookies!

Theorem 2.8. *The rational numbers $\mathbb{Q}$ that I've just constructed are a field.*

Proof. I'll leave you to check (A1).

(A2)

$$\begin{aligned}[(m_1, n_1)]_{\mathbb{Q}} + [(m_2, n_2)]_{\mathbb{Q}} &= [(m_1 n_2 + m_2 n_1, n_1 n_2)]_{\mathbb{Q}} \\ &= [(m_2 n_1 + m_1 n_2, n_2 n_1)]_{\mathbb{Q}} = [(m_2, n_2)]_{\mathbb{Q}} + [(m_1, n_1)]_{\mathbb{Q}}\end{aligned}$$

where the first and last equality follows from the definition of addition and the middle equality follows from (A2) and (M2) for $\mathbb{Z}$.

(A3) My 0 element is $[(0, 1)]_{\mathbb{Q}}$ since

$$[(m, n)]_{\mathbb{Q}} + [(0, 1)]_{\mathbb{Q}} = [(m \cdot 1 + 0 \cdot n, n \cdot 1)]_{\mathbb{Q}} = [(m, n)]_{\mathbb{Q}}$$

where the first equality follows from the definition of addition and the second from (M3) for $\mathbb{Z}$, the fact that $0 \cdot n = 0$ (your homework shows this follows from axioms), and (A3) for $\mathbb{Z}$.

(A4) We have $-[(m,n)]_{\mathbb{Q}} = [(-m,n)]_{\mathbb{Q}}$ since

$$[(m,n)]_{\mathbb{Q}} + [(-m,n)]_{\mathbb{Q}} = [(mn + (-m)n, n^2)]_{\mathbb{Q}} = [(0,n^2)]_{\mathbb{Q}} = [(0,1)]_{\mathbb{Q}}$$

where the first equality follows from definition of addition, the last follows from definition of $[-]_{\mathbb{Q}}$ since $0 \cdot 1 = 0 \cdot n^2$ and the middle inequality follows from addition laws in $\mathbb{Z}$.

(M1) and (M2) follow quickly from (M1) and (M2) for $\mathbb{Z}$. My 1 element is $[(1,1)]_{\mathbb{Q}}$, which can be verified using (M3) for $\mathbb{Z}$. I'll leave it to you to check (D).

We're just left with checking (M4). Suppose $[(m,n)]_{\mathbb{Q}} \neq 0$. This means $[(m,n)]_{\mathbb{Q}} \neq [(0,1)]_{\mathbb{Q}}$ so that $m \cdot 1 \neq 0 \cdot n$, i.e. $m \neq 0$. We claim

$$[(m,n)]_{\mathbb{Q}}^{-1} = [(n \cdot \operatorname{sgn}(m), |m|)],$$

where $\operatorname{sgn}(m) = \frac{m}{|m|}$. This is because

$$[(m,n)]_{\mathbb{Q}}[(n \cdot \operatorname{sgn}(m), |m|)]_{\mathbb{Q}} = [(mn \cdot \operatorname{sgn}(m), n \cdot |m|)]_{\mathbb{Q}} = [(1,1)]_{\mathbb{Q}}$$

where the first equality follows from definition of multiplication and the second follows from the definition of $[-]_{\mathbb{Q}}$ since $m \cdot \operatorname{sgn}(m) = |m|$.

[If we had constructed $\mathbb{Q}$ using $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ we could have taken $[(m,n)]^{-1} = [(n,m)]_{\mathbb{Q}}$ instead, which would have been easier, probably better even, but I preferred typing $\mathbb{N}$ over $\mathbb{Z} \setminus \{0\}$.] $\square$

Notation 2.9. Write $\frac{m}{n}$ for $[(m,n)]_{\mathbb{Q}}$ so that

$$\mathbb{Q} = \left\{ \frac{m}{n} : \ m \in \mathbb{Z}, \ n \in \mathbb{N} \right\}.$$

We're now back to where we were before I hypnotized you.

Remark 2.10. What was the point in all of this? In the next section, we'll check addition on $\mathbb{Z}/n$ is well-defined and it will actually be much easier. If you had have been taught modular arithmetic when you were younger as opposed to adding and dividing fractions, you would be awesome at it by now.

When you were younger, you believed some older authority that the addition and multiplication of fractions makes sense; now you have seen why it works. How could it have failed? Maybe, when you first learned to add fractions, you tried the "rule"

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 + m_2}{n_1 + n_2}.$$

Your teacher will have told you this is wrong, but maybe they didn't give you a good reason as to why. Suppose we try to define addition this way. Then

$$\frac{0}{1} + \frac{1}{1} = \frac{1}{2}.$$

However, $\frac{1}{1}$ also has the name $\frac{2}{2}$. Using this name, with the definition we get

$$\frac{0}{1} + \frac{2}{2} = \frac{2}{3}.$$

Uh oh... $\frac{1}{2} \neq \frac{2}{3}$. This definition of addition leads to nonsense since renaming elements changes the answer we get.

Thank you to Chris Jeon for inspiring this!

3 Modular arithmetic

Modular arithmetic is

"usual integer arithmetic" + "ignoring multiples of some integer."

If it's 10am and someone asks you what time it will be in 5 hours, you answer, "3pm" because

$$10 + 5 \equiv 3 \pmod{12}.$$

Upon taking the remainder of 15 upon division by 12, you get 3.

Definition 3.1. If $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}$ we say that *a is congruent to b modulo n* and write

$$a \equiv b \pmod{n}, \quad a \equiv b \ (n) \quad \text{or} \quad [a]_n = [b]_n$$

if there is a $c \in \mathbb{Z}$ such that $a = b + cn$.

Definition 3.2. If $n \in \mathbb{N}$, we write $\mathbb{Z}/n$ for the set of integers modulo n, that is

$$\mathbb{Z}/n = \{[x]_n : x \in \mathbb{Z}\}.$$

Example 3.3. What is $\mathbb{Z}/7$? Well,

$$\begin{aligned}
\ldots &\equiv -21 \equiv -14 \equiv -7 \equiv 0 \equiv 7 \equiv 14 \equiv 21 \equiv \ldots \ (7)\\
\ldots &\equiv -20 \equiv -13 \equiv -6 \equiv 1 \equiv 8 \equiv 15 \equiv 22 \equiv \ldots \ (7)\\
\ldots &\equiv -19 \equiv -12 \equiv -5 \equiv 2 \equiv 9 \equiv 16 \equiv 23 \equiv \ldots \ (7)\\
\ldots &\equiv -18 \equiv -11 \equiv -4 \equiv 3 \equiv 10 \equiv 17 \equiv 24 \equiv \ldots \ (7)\\
\ldots &\equiv -17 \equiv -10 \equiv -3 \equiv 4 \equiv 11 \equiv 18 \equiv 25 \equiv \ldots \ (7)\\
\ldots &\equiv -16 \equiv -9 \equiv -2 \equiv 5 \equiv 12 \equiv 19 \equiv 26 \equiv \ldots \ (7)\\
\ldots &\equiv -15 \equiv -8 \equiv -1 \equiv 6 \equiv 13 \equiv 20 \equiv 27 \equiv \ldots \ (7)
\end{aligned}$$

i.e.

$$\begin{aligned}
\ldots &= [-21]_7 = [-14]_7 = [-7]_7 = [0]_7 = [7]_7 = [14]_7 = [21]_7 = \ldots\\
\ldots &= [-20]_7 = [-13]_7 = [-6]_7 = [1]_7 = [8]_7 = [15]_7 = [22]_7 = \ldots\\
\ldots &= [-19]_7 = [-12]_7 = [-5]_7 = [2]_7 = [9]_7 = [16]_7 = [23]_7 = \ldots\\
\ldots &= [-18]_7 = [-11]_7 = [-4]_7 = [3]_7 = [10]_7 = [17]_7 = [24]_7 = \ldots\\
\ldots &= [-17]_7 = [-10]_7 = [-3]_7 = [4]_7 = [11]_7 = [18]_7 = [25]_7 = \ldots\\
\ldots &= [-16]_7 = [-9]_7 = [-2]_7 = [5]_7 = [12]_7 = [19]_7 = [26]_7 = \ldots\\
\ldots &= [-15]_7 = [-8]_7 = [-1]_7 = [6]_7 = [13]_7 = [20]_7 = [27]_7 = \ldots
\end{aligned}$$

so

$$\mathbb{Z}/7 = \{[x]_7 : x \in \mathbb{Z}\} = \{[0]_7, [1]_7, [2]_7, [3]_7, [4]_7, [5]_7, [6]_7\}.$$

We have found that we have lots of names for the same element. Actually, this is not new. You have known for a long time that in $\mathbb{Q}$

$$\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \frac{4}{8} = \frac{5}{10} = \ldots,$$

that is, we have many names for $\frac{1}{2}$.

We can add and multiply elements modulo n. We just do the normal addition and multiplication in the integers, add or substract multiples of n if we wish, and then say "modulo n." For instance,

$$2+3 \equiv 5 \ (7), \quad 4+5 \equiv 2 \ (7), \quad 2 \cdot 4 \equiv 1 \ (7), \quad 4 \cdot 5 \equiv -1 \ (7).$$

We have to check that this makes sense though.

Theorem 3.4. *Addition and multiplication on the integers $\mathbb{Z}$ "descends to" a well-defined addition and multiplication on the integers modulo n, $\mathbb{Z}/n$.*

Proof. The phrase "descend to" means we are trying to define an addition by

$$[x]_n + [y_n] = [x+y]_n$$

and a multiplication by

$$[x]_n \cdot [y]_n = [xy]_n.$$

We worry that something could go wrong because we have different names for the same thing and our definition depends on the name used. For instance, $[2]_7 = [9]_7$ and $[3]_7 = [10]_7$. On the one hand our definition says

$$[2]_7 \cdot [3]_7 = [2 \cdot 3]_7 = [6]_7.$$

On the other hand our definition says

$$[9]_7 \cdot [10]_7 = [9 \cdot 10]_7 = [90]_7.$$

Thankfully $[6]_7 = [90]_7$ because $6 = 90 - 12 \cdot 7$ (we're taking $c = -12$ in the definition) and so all is right with the world.

Again, one can remark that, back when we were little boys and girls being taught to add and multiply fractions, we should have checked something similar. Does the formula in $\mathbb{Q}$

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 n_2 + m_2 n_1}{n_1 n_2}$$

depend on the way we named our fractions? Thankfully not! This is theorem 2.6.

Okay, we've chatted for a long time; better actually prove something...

Suppose $[x]_n = [x']_n$ and $[y]_n = [y']_n$. This tells us, by definition, that there are $c, d \in \mathbb{Z}$ such that

$$x = x' + cn, \ y = y' + dn.$$

Thus,

$$x + y = (x' + y') + (c+d)n \quad \text{and} \quad xy = x'y' + (x'd + cy' + cdn)n.$$

By definition, this tells us that

$$[x+y]_n = [x'+y']_n \quad \text{and} \quad [xy]_n = [x'y']_n.$$

So whatever name we pick to do the calculation, we get the same answer. □

Theorem 3.5. *$\mathbb{Z}/n$ is a commutative ring.*

Proof. (A1):

$$\begin{aligned}[x]_n + ([y]_n + [z]_n) &= [x]_n + [y+z]_n = [x + (y+z)]_n \\ &= [(x+y)+z]_n = [x+y]_n + [z]_n = ([x]_n + [y]_n) + [z]_n,\end{aligned}$$

where the first, second, fourth and fifth equality follow from the definition and the third follows from (A1) for $\mathbb{Z}$.

(A2):

$$[x]_n + [y]_n = [x+y]_n = [y+x]_n = [y]_n + [x]_n,$$

where the first and last equality follow from the definition and the middle inequality follows from (A2) for $\mathbb{Z}$.

(A3): $0 = [0]_n$ since

$$[x_n] + [0]_n = [x+0]_n = [x]_n = [0+x]_n = [0]_n + [x]_n,$$

where the first and last equality follow from the definition and the middle two follow from (A3) for $\mathbb{Z}$.

(A4): $-[x]_n = [-x]_n$ since

$$[x]_n + [-x]_n = [x + (-x)]_n = [0]_n = [(-x) + x]_n = [-x]_n + [x]_n$$

where the first and last equality follow from the definition and the middle two follow from (A4) for $\mathbb{Z}$.

(M1)-(M3) and (D) are left for the homework. □

Example 3.6. $\mathbb{Z}/2 = \{[0]_2, [1]_2\}$. The addition is described by

$$\begin{aligned}[0]_2 + [0]_2 = [1]_2 + [1]_2 = [0]_2 \\ [0]_2 + [1]_2 = [1]_2 + [0]_2 = [1]_2\end{aligned}$$

and the multiplication is described by

$$[0]_2[0]_2 = [0]_2[1]_2 = [1]_2[0]_2 = [0]_2,\ [1]_2[1]_2 = [1]_2.$$

Since $[1]_2$ is the only non-zero element and $[1]_2[1]_2 = [1]_2$, $\mathbb{Z}/2$ is a field.

A good way to think of $\mathbb{Z}/2$ is as the set $\{\text{even}, \text{odd}\}$; the addition and multiplication rules are exactly the rules for adding and multiplying even and odd integers.

Often we write 0 and 1 for $[0]_2$ and $[1]_2$, respectively, to avoid cumbersome notation. After all, these elements are the 0 and 1 that appear in the axioms for a ring. If we are feeling lazy in other settings we may miss out the $[-]_n$, too.

Example 3.7. $\mathbb{Z}/6 = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6, \}$ and in class I'll write out the multiplication table. It is not a field since $[2]_6 \cdot x$ is never equal to $[1]_6$.

4 Polynomials

You have encountered polynomials throughout your math education. Up until now, they probably will have had real coefficients, or maybe complex coefficients. We can actually allow coefficients in any commutative ring.

4.1 Polynomial rings

Definition 4.1.1. Let R be a commutative ring. We write $R[x]$ for the set of *polynomials*

$$\{f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \ : \ n \in \mathbb{N} \cup \{0\},\ a_0, \ldots, a_n \in R\}.$$

Remark 4.1.2. As before we can name things in more than one way by omitting or inserting expressions like $0x^k$. For example, in $\mathbb{Z}[x]$,

$$7 - x + 3x^2 + 2x^5 = 7 + (-1)x + 3x^2 + 0x^3 + 0x^4 + 2x^5 + 0x^6 + 0x^7.$$

We can also change the order in which we write monomials so that, in $\mathbb{Z}[x]$, $1 + x = x + 1$.

Two polynomials are the same if the coefficients of each of their x^k terms are equal.

Definition 4.1.3. If $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$ is a non-zero element of $R[x]$ then the largest k such that $a_k \neq 0$ is called the *degree* of $f(x)$. If $a_{\deg f(x)} = 1$, then $f(x)$ is called *monic*.

Example 4.1.4. Consider the polynomials $\mathbb{Z}/2[x]$. Since $\mathbb{Z}/2$ only has two elements we can write down all the polynomials of degree less than or equal to 2 quickly.

$$0,\ 1,\ x,\ x+1,\ x^2,\ x^2+1,\ x^2+x,\ x^2+x+1.$$

Some good lessons can be learned from this example.

1. When one encounters polynomials in lower division courses they are often real-valued polynomials and we think of them as the same as the real-valued function they define, i.e. we think of the polynomial $x^2 - 2x + 5 \in \mathbb{R}[x]$ as the same as the function $f : \mathbb{R} \longrightarrow \mathbb{R}$, defined by $f(x) = x^2 - 2x + 5$.

   A polynomial in $\mathbb{Z}/2[x]$ does define a function $\mathbb{Z}/2 \longrightarrow \mathbb{Z}/2$, but different polynomials can give rise to the same function. For instance, the polynomials x and x^2 both give the identity function $\mathbb{Z}/2 \longrightarrow \mathbb{Z}/2$. However, they are *different* polynomials since the coefficients of x and x^2 differ.

2. When one wishes to factor a polynomial in $\mathbb{R}[x]$ one looks for roots to the polynomial. The same trick works here. 1 is a root of $x^2 + 1$ and indeed, because $[2]_2 = 0$, $x^2 + 1$ factors as $(x+1)^2$. This "trick" is corollary 5.3.3. We can try plugging in 0 and 1 to see that x^2+x+1 has no roots. Thus, $x^2 + x + 1$ does not factor.

Example 4.1.5. In $\mathbb{Z}/2[x]$ we have

$$x^4 - x = x(x^3 - 1) = x(x - 1)(x^2 + x + 1)$$

and cannot factor any further.

Example 4.1.6. In $\mathbb{Z}/2[x]$ we have

$$\begin{aligned} x^8 - x = x(x^7 - 1) &= x(x-1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1) \\ &= x(x-1)(x^3 + x + 1)(x^3 + x^2 + 1) \end{aligned}$$

and cannot factor any further.

Example 4.1.7. In $\mathbb{Z}/3[x]$ we have

$$\begin{aligned} x^9 - x = x(x^8 - 1) = x(x^4 - 1)(x^4 + 1) &= x(x^2 - 1)(x^2 + 1)(x^4 + 1) \\ &= x(x-1)(x+1)(x^2 + 1)(x^4 + 1) \\ &= x(x-1)(x+1)(x^2 + 1)(x^2 - x - 1)(x^2 + x - 1) \end{aligned}$$

and cannot factor any further.

Example 4.1.8. In $\mathbb{Z}/5[x]$ we have

$$\begin{aligned} x^{25} - x &= x(x^{24} - 1) \\ &= x(x^{12} - 1)(x^{12} + 1) \\ &= x(x^6 - 1)(x^6 + 1)(x^{12} + 1) \\ &= x(x^3 - 1)(x^3 + 1)(x^6 + 1)(x^{12} + 1) \\ &= x(x-1)(x^2 + x + 1)(x+1)(x^2 - x + 1)(x^2 + 1)(x^4 - x^2 + 1)(x^4 + 1)(x^8 - x^4 + 1) \\ &= x(x-1)(x^2 + x + 1)(x+1)(x^2 - x + 1)(x - [2]_5)(x + [2]_5) \\ &\quad (x^4 - x^2 + 1)(x^2 - [2]_5)(x^2 + [2]_5)(x^8 - x^4 + 1) \\ &= x(x-1)(x+1)(x - [2]_5)(x + [2]_5)(x^2 - x + 1)(x^2 + x + 1)(x^2 - [2]_5)(x^2 + [2]_5) \\ &\quad (x^4 - x^2 + 1)(x^8 - x^4 + 1) \\ &= x(x-1)(x+1)(x - [2]_5)(x + [2]_5)(x^2 - x + 1)(x^2 + x + 1)(x^2 - [2]_5)(x^2 + [2]_5) \\ &\quad (x^2 - [2]_5 x - 1)(x^2 + [2]_5 x - 1)(x^4 - [2]_5 x^2 - 1)(x^4 + [2]_5 x^2 - 1) \end{aligned}$$

You can actually factor into degree 2 polynomials. Good luck factoring the final two quartics!

Example 4.1.9. In $\mathbb{Z}/4[x]$, $(x + [2]_4)^2 = x^2 + ([2]_4 + [2]_4)x + [2]_4 \cdot [2]_4 = x^2$. Because $[2]_4[2]_4 = 0$ we get some strange looking formulae. The formula says that either we should regard 0 as a repeated root, or $[2]_4$ as a repeated root, but not 0 and $[2]_4$ as both being roots since $x^2 \neq x(x + [2]_4)$. Our usual terminology does not make sense, the reason being that unique factorization does not occur in $\mathbb{Z}/4[x]$.

Theorem 4.1.10. *If R is a commutative ring, then so is $R[x]$.*

Proof. Suppose $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$ and $g(x) = b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m$. By inserting terms like $0x^k$ we can make sure $n = m$. Then we define

$$f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \ldots + (a_n + b_n)x^n$$

and $f(x)g(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_{2n} x^{2n}$ where $c_k = \sum_{i+j=k} a_i b_j$ ((A1) for R tells us we don't need to use brackets when adding, and (A2) says we don't care about the order so this summation is well-defined).

One then checks the axioms. I'll get you to do some of this in the homework. □

Definition 4.1.11. If R is a commutative ring, $R[x]$ is called *the polynomial ring over R.*

4.2 Modular arithmetic with polynomials

Modular arithmetic with polynomials is

"usual polynomial arithmetic" + "ignoring multiples of some polynomial."

Definition 4.2.1. If $f(x), g(x), q(x) \in R[x]$ we say that $f(x)$ *is congruent to* $g(x)$ *modulo* $q(x)$ and write

$$f(x) \equiv g(x) \pmod{q(x)}, \quad \text{or} \quad [f(x)]_{q(x)} = [g(x)]_{q(x)}$$

if there is a $h(x) \in R[x]$ such that $f(x) = g(x) + h(x)q(x)$.

Definition 4.2.2. If $q(x) \in R[x]$, we write $R[x]/(q(x))$ for the set of polynomials with coefficients in R modulo $q(x)$, that is

$$R[x]/(q(x)) = \{[f(x)]_{q(x)} : f(x) \in R[x]\}.$$

Theorem 4.2.3. *Addition and multiplication on the polynomial ring $R[x]$ "descends to" a well-defined addition and multiplication on the polynomial ring modulo $q(x)$, $R[x]/(q(x))$. These operations make $R[x]/(q(x))$ into a ring.*

Proof. "Descends to" means we define

$$[f(x)]_{q(x)} + [g(x)]_{q(x)} = [f(x) + g(x)]_{q(x)} \quad \text{and} \quad [f(x)]_{q(x)} \cdot [g(x)]_{q(x)} = [f(x)g(x)]_{q(x)}.$$

The rest of the proof is the same as the proof of theorem 3.4 and 3.5. In particular,

$$0 = [0]_{q(x)} \quad \text{and} \quad 1 = [1]_{q(x)},$$

where the 0 and 1 inside the brackets denote the 0 and 1 of $R[x]$. We showed in the first homework that these are the zero polynomial and the constant polynomial with value 1, respectively. □

Example 4.2.4. What is $\mathbb{R}[x]/(x^2+1)$? We start with $\mathbb{R}$, adjoin an element x, and then make some stuff equal by using the brackets $[-]_{x^2+1}$. Everything is determined by the fact that

$$[x^2 + 1]_{x^2+1} = [0]_{x^2+1} = 0.$$

This relation holds by the definition of $[-]_{x^2+1}$ and the fact that $x^2 + 1 = 0 + 1 \cdot (x^2 + 1)$. It forces

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

In general, we have

$$[x]^{4n}_{x^2+1} = 1, \ [x]^{4n+1}_{x^2+1} = [x]_{x^2+1}, \ [x]^{4n+2}_{x^2+1} = -1, \ [x]^{4n+3}_{x^2+1} = [-x]_{x^2+1}.$$

This means we can always eliminate x^2, x^3, ... from inside $[-]_{x^2+1}$ and so

$$\mathbb{R}[x]/(x^2 + 1) = \{[a + bx]_{x^2+1} : a, b \in \mathbb{R}\}.$$

What did we do? We started with $\mathbb{R}$, and adjoined an element $i = [x]_{x^2+1}$ with the property that $i^2 = [x]^2_{x^2+1} = -1$. This is $\mathbb{C}$.

Example 4.2.5. $\mathbb{Z}/2[x]/(x^2+x+1)$. The thing to notice is that

$$x^2 \equiv x+1 \pmod{x^2+x+1}.$$

This is because $x^2 = (x+1) + 1\cdot(x^2+x+1)$ since, in $\mathbb{Z}/2[x]$, $x+x = 1+1 = 0$.

Thus, for any $n \in \mathbb{N}\cup\{0\}$ we have

$$x^{n+2} \equiv x^{n+1} + x^n \pmod{x^2+x+1}$$

and this allows us to write monomials x^k with $k \geq 2$ in terms of polynomials with lower degree. Thus,

$$\mathbb{Z}/2[x]/(x^2+x+1) = \{0, 1, [x], [x+1]\},$$

where we have started omitting the subscript on the square brackets due to laziness.

In lecture, I went through the addition and multiplication table for this. I'll ask you to repeat this on the homework. The next example is similar, and slightly more involved.

Example 4.2.6. $\mathbb{Z}/3[x]/(x^2+1)$. Using the same argument as in the previous example we have

$$\mathbb{Z}/3[x]/(x^2+1) = \{0, 1, -1, [x], [x+1], [x-1], [-x], [-x+1], [-x-1]\}.$$

Suggestively call this ring $\mathbb{F}_9$, since it turns out to be a field with 9 elements and let $\alpha = [x]$. Then

$$\mathbb{F}_9 = \{0, 1, -1, \alpha, \alpha+1, \alpha-1, -\alpha, -\alpha+1, -\alpha-1\}$$

and $\alpha^2+1 = [x]^2+[1] = [x^2+1] = [0] = 0$ so that

$$\alpha^2 = -1 \quad \text{and} \quad \alpha^2 - 1 = 1.$$

This means that in $\mathbb{F}_9[y]$

$$\begin{aligned} y^9 - y &= y(y-1)(y+1)(y^2+1)(y^2-y-1)(y^2+y-1) \\ &= y(y-1)(y+1) \\ &\quad \cdot (y-\alpha)(y+\alpha)(y-(\alpha-1))(y+(\alpha+1))(y+(\alpha-1))(y-(\alpha+1)). \end{aligned}$$

See example 4.1.7 for the first equality. I have changed the polynomial indeterminant x to y to avoid confusion with the previous x.

It turns out that constructing the smallest thing containing $\mathbb{Z}/p$ where the polynomial $y^{p^n} - y$ factors into linear factors gives a field with p^n elements.

5 Division

5.1 The definition of "b divides a"

The definition of a commutative ring is based on the properties of $\mathbb{Z}$. In $\mathbb{Z}$ we have the notion of divisibility: 6 is divisible by 2. To say this another way: 2 divides 6. This notion makes sense in any commutative ring.

Definition 5.1.1. Suppose R is a commutative ring. If $a, b \in R$, we say b *divides* a and write $b|a$ if $a = bc$ for some $c \in R$.

Example 5.1.2. In $\mathbb{Z}$, $2|6$, $(-5)|100$, $3|(-30)$, 5 does not divide 26.

Example 5.1.3. In $\mathbb{Z}/5$, $[3]_5$ divides $[2]_5$ since $[2]_5 = [3]_5 \cdot [4]_5$. In $\mathbb{Z}/6$, $[3]_6$ does not divide $[2]_6$, since the only multiplies of $[3]_6$ are $[0]_6$ and $[3]_6$.

Example 5.1.4. In $\mathbb{Z}/3[x]$, $(x-1)|(x^2+x+1)$ since $x^2+x+1 = (x-1)^2$.

5.2 Division in $\mathbb{Z}$

In $\mathbb{Z}$ things are even better. Even when we cannot divide exactly we have the notion of remainder.

Theorem 5.2.1 (Division theorem)**.** *Suppose that $a \in \mathbb{N} \cup \{0\}$ and $b \in \mathbb{N}$. Then there exist unique $q, r \in \mathbb{N} \cup \{0\}$ with the properties that $a = qb + r$ and $0 \leq r < b$.*

Proof. First things first: the theorem stated in the book is wrong since it says $q > 0$; we must allow for $q = 0$.

How did we do this when we were in primary school? We added b to itself as many times as possible without it being bigger than a. This gave q and then we let $r = a - qb$.

Let's write this carefully in math. Let

$$S = \{x \in \mathbb{N} \cup \{0\} : xb \leq a\}.$$

Since $b \geq 1$, if $x > a$ then $xb > a$ and so $x \notin S$. By the contrapositive, if $x \in S$ then $x \leq a$ and so

$$S \subseteq \{0, 1, \ldots, a\}.$$

Let $q = \max S$ and $r = a - qb$. Since $q \in S$ we have $qb \leq a$ and so $r \geq 0$. If $r \geq b$ then $a - qb \geq b$ so that $(q+1)b \leq a$, giving $q + 1 \in S$, a contradiction. Thus, $0 \leq r < b$.

For uniqueness, suppose that $qb + r = q'b + r'$ where $q, q', r, r' \in \mathbb{N} \cup \{0\}$ and $0 \leq r, r' < b$. Without loss of generality assume that $r \leq r'$. Then

$$0 \leq r' - r \leq r' < b.$$

Moreover, $(q - q')b = r' - r$ so that $0 \leq (q - q')b < b$. Dividing by b gives $0 \leq (q - q') < 1$. Since $q, q' \in \mathbb{N} \cup \{0\}$, this shows $q = q'$ and so $r = r'$. □

Corollary 5.2.2. *Suppose $a, b \in \mathbb{Z}$ and that $b \neq 0$. Then there exist $q, r \in \mathbb{Z}$ with $a = qb + r$ and $|r| < |b|$.*

Proof. If we can do the case when $b > 0$, then we can do the case when $b < 0$ by changing the sign of q, so suppose $b \in \mathbb{N}$.

Since $|a| \in \mathbb{N} \cup \{0\}$ there exists $q, r \in \mathbb{N} \cup \{0\}$ with the properties that

$$|a| = qb + r$$

and $0 \leq r < b$. If $a \geq 0$, we are finished. If $a < 0$, then we change the sign of q and r. □

Perhaps you would prefer the following stronger corollary.

Corollary 5.2.3. *Suppose $a, b \in \mathbb{Z}$ and that $b \neq 0$. Then there exist unique $q, r \in \mathbb{Z}$ with $a = qb + r$ and $0 \leq r < |b|$.*

Proof. Homework. It is probably best to prove it directly; I have called it a corollary because the proof is so similar. □

One way in which division in $\mathbb{Z}$ helps us is that it allows us to express numbers in different bases. How does our usual base 10 number system work? To express x in base 10 we find n so that $10^n \leq x < 10^{n+1}$; divide by 10^n and take the remainder; divide the remainder by 10^{n-1} and take the remainder; and so on.

$$\begin{aligned} 53796 &= 5 \cdot 10^4 + 3796 \\ 3796 &= 3 \cdot 10^3 + 796 \\ 796 &= 7 \cdot 10^2 + 96 \\ 96 &= 9 \cdot 10^1 + 9 \\ 6 &= 6 \cdot 10^0 + 0 \end{aligned}$$

Similarly, for base 3.

$$\begin{aligned} 53796 &= 2 \cdot 3^9 + 14430 \\ 14430 &= 2 \cdot 3^8 + 1308 \\ 1308 &= 0 \cdot 3^7 + 1308 \\ 1308 &= 1 \cdot 3^6 + 579 \\ 579 &= 2 \cdot 3^5 + 93 \\ 93 &= 1 \cdot 3^4 + 12 \\ 12 &= 0 \cdot 3^3 + 12 \\ 12 &= 1 \cdot 3^2 + 3 \\ 3 &= 1 \cdot 3^1 + 0 \\ 0 &= 0 \cdot 3^0 + 0 \end{aligned}$$

In base 3, 53796_{10} looks like 2201210110_3.

5.3 Division in $F[x]$ when F is a field

In a polynomial ring with coefficients in a field we can also divide with remainder.

Recall the definition of the degree of a polynomial. The following lemma is intuitive.

Proposition 5.3.1. *Suppose F is a field and $f(x), g(x) \in F[x]$ are non-zero polynomials. Then*

$$\deg(f(x)g(x)) = \deg f(x) + \deg g(x).$$

Proof. Suppose $\deg(f(x)) = n$ and $\deg(g(x)) = m$. Then we have

$$f(x) = a_0 + a_1x + \ldots + a_nx^n,\ g(x) = b_0 + b_1x + \ldots + b_mx^m$$

where $a_n, b_m \neq 0$. $f(x)g(x) = c_0 + c_1x + \ldots + c_{n+m}x^{n+m}$ where $c_{n+m} = a_nb_m \neq 0$. This is because non-zero elements of a field multiply to be non-zero (homework). Thus, $f(x)g(x) \neq 0$. □

Theorem 5.3.2. *(Division theorem for polynomials) Suppose F is a field and that $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Then there exist unique $q(x), r(x) \in F[x]$ with the properties that $f(x) = q(x)g(x) + r(x)$, and either $r(x) = 0$, or $r(x) \neq 0$ and $\deg r(x) < \deg g(x)$.*

Proof. If $f(x) = 0$ or $f(x) \neq 0$ and $\deg f(x) < \deg g(x)$, then writing $f(x) = 0 \cdot g(x) + f(x)$ shows we can take $q(x) = 0$ and $r(x) = f(x)$.

Suppose $f(x) \neq 0$ and let $k = \deg f(x) - \deg g(x)$. We have just shown that we are done when $k < 0$. We proceed by induction on k. So suppose $k \geq 0$. Let $m = \deg g(x)$ and write

$$f(x) = a_{m+k}x^{m+k} + \ldots + a_1x + a_0,\ g(x) = b_mx^m + \ldots + b_1x + b_0$$

where $a_{m+k}, b_m \neq 0$.

Observe that

$$\frac{a_{m+k}}{b_m}x^k g(x) = a_{m+k}x^{m+k} + \frac{a_{m+k}}{b_m}b_{m-1}x^{m+k-1} + \ldots + \frac{a_{m+k}}{b_m}b_1x^{k+1} + \frac{a_{m+k}}{b_m}b_0x^k.$$

Thus, $\tilde{f}(x) = f(x) - \frac{a_{m+k}}{b_m}x^k g(x)$ satisfies $\deg \tilde{f}(x) < \deg f(x)$ and thus

$$\deg \tilde{f}(x) - \deg g(x) < \deg f(x) - \deg g(x) = k.$$

By induction we can find $\tilde{q}(x), r(x) \in F[x]$ with the properties that

$$\tilde{f}(x) = \tilde{q}(x)g(x) + r(x)$$

and either $r(x) = 0$, or $r(x) \neq 0$ and $\deg r(x) < \deg g(x)$. Now

$$f(x) = \frac{a_{m+k}}{b_m}x^k g(x) + \tilde{f}(x) = \left[\frac{a_{m+k}}{b_m}x^k + \tilde{q}(x)\right]g(x) + r(x),$$

so we can take $q(x) = \frac{a_{m+k}}{b_m}x^k + \tilde{q}(x)$.

For uniqueness, suppose that $q_0(x)g(x)+r_0(x) = q_1(x)g(x)+r_1(x)$ where $q_0(x), q_1(x), r_0(x), r_1(x) \in F[x]$ and $r_0(x), r_1(x)$ have the relevant properties. We have

$$(q_0(x) - q_1(x))g(x) = r_1(x) - r_0(x).$$

If $q_0(x) - q_1(x) \neq 0$ then $r_1(x) - r_0(x) \neq 0$ giving

$$\deg g(x) \leq \deg((q_0(x) - q_1(x))g(x)) = \deg(r_1(x) - r_0(x)) < \deg g(x),$$

a contradiction. Thus, $q_0(x) = q_1(x)$ and so $r_0(x) = r_1(x)$. □

A picture to accompany the proof of theorem 5.3.2.

$$
\begin{array}{rl}
 & \frac{a_{m+k}}{b_m}x^k \qquad\qquad \tilde{q}(x) \text{ ——————————————————} \\
\hline
b_m x^m + b_{m-1}x^{m-1} + \ldots + b_0 \;\Big|\; & a_{m+k}x^{m+k} + a_{m+k-1}x^{m+k-1} + \ldots + a_k x^k + a_{k-1}x^{k-1} + \ldots + a_0 \\
 & \frac{a_{m+k}}{b_m}b_m x^{m+k} + \frac{a_{m+k}}{b_m}b_{m-1}x^{m+k-1} + \ldots + \frac{a_{m+k}}{b_m}b_0 x^k + 0x^{k-1} + \ldots + 0 \\
 & 0 \qquad\qquad \tilde{f}(x) \text{ ——————————————————} \\
 & \qquad\qquad\qquad\qquad \vdots \\
 & \qquad\qquad\qquad\qquad\qquad\qquad r(x) \text{ ———}
\end{array}
$$

Corollary 5.3.3. *Suppose F is a field, that $f(x) = a_0 + a_1x + \ldots + a_nx^n \in F[x]$, and that $b \in F$. Then*

$$f(b) = a_0 + a_1b + \ldots + a_nb^n \in F \subset F[x]$$

is the remainder when dividing $f(x)$ by $(x - b)$. Thus $(x - b)|f(x)$ precisely when $f(b) = 0 \in F$.

Proof. Since $\deg(x - b) = 1$, the division theorem gives us $q(x) \in F[x]$ and $r \in F$ such that

$$f(x) = q(x)(x - b) + r.$$

Setting $x = b$ gives $f(b) = r$. □

Example 5.3.4. Let $f(x) = x^4 - 10x^2 + 10x - 1$ and $g(x) = x^2 + 3x - 2$ in $\mathbb{Q}[x]$.

$$\begin{array}{r|l} & x^2 - 3x + 1 \\ \hline x^2+3x-2 & x^4 + 0x^3 - 10x^2 + 10x - 1 \\ & x^4 + 3x^3 - 2x^2 \\ \hline & -3x^3 - 8x^2 + 10x \\ & -3x^3 - 9x^2 + 6x \\ \hline & +x^2 + 4x - 1 \\ & +x^2 + 3x - 2 \\ \hline & +x + 1 \end{array}$$

So $f(x) = q(x)g(x) + r(x)$ where $q(x) = x^2 - 3x + 1$ and $r(x) = x + 1$.

Example 5.3.5. Let $f(x) = x^4 - 1$ and $g(x) = x^2 - [2]_5x - [2]_5$ in $\mathbb{Z}/5[x]$.

$$\begin{array}{r|l} & x^2 + 2x + 1 \\ \hline x^2+3x+3 & x^4 + 0x^3 + 0x^2 + 0x + 4 \\ & x^4 + 3x^3 + 3x^2 \\ \hline & +2x^3 + 2x^2 + 0x \\ & +2x^3 + x^2 + 2x \\ \hline & +x^2 + 3x + 4 \\ & +x^2 + 3x + 3 \\ \hline & +1 \end{array}$$

So $f(x) = q(x)g(x) + r(x)$ where $q(x) = x^2 + 2x + 1$ and $r(x) = 1$.

If you look at these two examples carefully you'll notice the second is just the first taken modulo 5.

6 Zero divisors, units

In $\mathbb{Z}$ we are used to the notion of cancellation: if $ab = ac$, and $a \neq 0$, we can cancel to get $b = c$. The argument for this is really as follows. Suppose $ab = ac$. Then $a(b - c) = ab - ac = 0$. Since $a \neq 0$, then $b - c = 0$ and thus $b = c$. What we need is the fact that "$ab = 0$ implies either $a = 0$ or $b = 0$." This is not true in other commutative rings.

Example 6.1. In $\mathbb{Z}/6$ we have $[2]_6[3]_6 = [6]_6 = [0]_6$. Two non-zero numbers multiply to give 0.

We say $[2]_6$ and $[3]_6$ are zero divisors, since they divide 0 (and they're non-zero).

Definition 6.2. Suppose R is a commutative ring and that $a \in R$. We say that a is a *zero divisor* if $a \neq 0$ and there is a $b \in R$ with $b \neq 0$ such that $ab = 0$.

Rings where we are allowed to cancel are called integral domains.

Definition 6.3. A commutative ring R is said to be an *integral domain* if there are no zero divisors, i.e. if $a, b \in R$ and $a, b \neq 0$, then $ab \neq 0$; alternatively, if $a, b \in R$ and $ab = 0$, then either $a = 0$ or $b = 0$.

Proposition 6.4. *Fields are integral domains.*

Proof. Homework. □

Proposition 6.5. *If R is an integral domain then so is $R[x]$.*

Proof. Homework. □

Just as numbers which divide 0 have a special name, so do numbers which divide 1.

Definition 6.6. Suppose R is a commutative ring. We say $u \in R$ is a *unit* if there exists a $v \in R$ such that $uv = 1$.

Example 6.7. In $\mathbb{Z}$ the only units are 1 and -1.

Example 6.8. In $\mathbb{Z}/5$, $[1]_5$, $[2]_5$, $[3]_5$, $[4]_5$, are units since

$$[1]_5[1]_5 = [2]_5[3]_5 = [4]_5[4]_5 = 1.$$

Example 6.9. In a field every non-zero element is a unit.

Example 6.10. If F is a field, the units of $F[x]$ are the non-zero degree zero polynomials: indeed, non-zero elements of $F \subset F[x]$ are units; because $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$ and $\deg 1 = 0$, units better have degree zero. Similarly, the units of $\mathbb{Z}[x]$ are 1 and -1.

It is familiar that we can factor any natural number into primes uniquely (up to reordering). When we extend this result to the integers, it basically says the same thing, except we are allowed to multiply our primes by -1 so that

$$10 = 2 \cdot 5 = (-2) \cdot (-5), \ -35 = (-5) \cdot 7 = 5 \cdot (-7).$$

2 and -2, 5 and -5, 7 and -7 are called associates. In a commutative ring we make the following definition.

Definition 6.11. Suppose R is a commutative ring and $a, b \in R$. We say that a is an *associate* of b if there is a unit $u \in R$ with $a = bu$.

Remark 6.12. If R is a commutative ring, $a, b \in R$, and a is an associate of b, then b is an associate of a, and so we can say a, b are associates. [If $uv = 1$, then $a = bu$ if and only if $av = b$.]

7 Greatest common divisors

7.1 The definition

For the natural numbers prime decompositions help us to find the greatest common divisor of a collection of numbers: we just read off common prime factors and multiply them. In fact, we should prove such a prime decomposition exists and we will do this eventually.

We can define the concept of a gcd in a commutative ring.

Definition 7.1.1. Let R be a commutative ring and let $a_1, \ldots, a_n \in R$. Then we say an element $d \in R$ is a *greatest common divisor* or *gcd* of $a_1, \ldots, a_n$ if

1. $d|a_1, \ldots, d|a_n$ (it is a common divisor of $a_1, \ldots, a_n$).
2. If $c \in R$ satisfies $c|a_1, \ldots, c|a_n$, then $c|d$ (it is a *greatest* common divisor).

To get the definition straight here are two examples, assuming we know prime decompositions exist and are unique.

Example 7.1.2. In $\mathbb{Z}$, 6 and -6 are greatest common divisors for 30 and 42. Let's check 6.

1. $6|30$, $6|42$.
2. If $c|30$, c can only have prime factors 2, 3, and 5, and they can occur at most once. If $c|42$, c can only have prime factors 2, 3, and 7, and they can occur at most once. Thus, if $c|30$ and $c|42$, c can only have prime factors 2 and 3, and they can occur at most once. Thus c divides 6.

Example 7.1.3. In $\mathbb{Q}[x]$, $x(x-1)$ is a gcd for $(x+1)x(x-1)$ and $x(x-1)(x-2)$.

Remark 7.1.4. Let R be a commutative ring and suppose d is a gcd of $a_1, \ldots, a_n \in R$. Then so is any associate of d (homework).

In the case of the integers this says that if d is a greatest common divisor of $a_1, \ldots, a_n$, then so is $-d$, which is hopefully clear. In the case of polynomials over a field this says, that if $g(x)$ is a greatest common divisor of $f_1(x), \ldots, f_n(x)$, then so is $cg(x)$ for any nonzero element c of the field.

Remark 7.1.5. Let R be an integral domain and suppose d and d' are both gcds of $a_1, \ldots, a_n \in R$. Then d and d' are associates (homework).

So for integers greatest common divisors can only differ by a sign, and for polynomials defined over a field greatest common divisors can only differ by a nonzero element of the field.

As the remarks which follow will show, the existence of gcd's is not entirely obvious. It turns out that for $\mathbb{Z}$ and polynomials over a field, they do exist, and we we can make the following definition.

Definition 7.1.6. In $\mathbb{Z}$, we will write $\gcd(a_1, \ldots, a_n)$ for the *positive* gcd of $a_1, \ldots, a_n \in \mathbb{Z}$.

If F is a field, we will write $\gcd(f_1(x), \ldots, f_n(x))$ for the *monic* gcd of $f_1(x), \ldots, f_n(x) \in F[x]$, (when it is non-zero).

Remark 7.1.7. There are commutative rings where greatest common divisors do not always exist. Let $R = \mathbb{Z}/4[x]$ and let

$$f(x) = x^2 = (x + [2]_4)^2, \quad g(x) = x(x + [2]_4).$$

One can check that the gcd of $f(x)$ and $g(x)$ does not exist.

Remark 7.1.8. There are even integral domains in which greatest common divisors do not always exist. Here is an example. Let $R = \mathbb{Z}[\sqrt{-3}] = \{x + y\sqrt{-3} : x, y \in \mathbb{Z}\}$ and let

$$a = 4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}), \quad b = 2 \cdot (1 + \sqrt{-3}).$$

One can check that the gcd of a and b does not exist.

In light of the previous remarks, how will we calculate greatest common divisors or even know they exist? One way is to use prime decompositions, but in order to prove such things exist in the cases we care about we need to do quite a bit of work. The most elementary things we can say are the content of the next two lemmas.

The second lemma suggests that to show greatest common divisors exist for $\mathbb{Z}$ and polynomials defined over a field, we might need to make use of the division algorithms we proved.

Lemma 7.1.9. *Suppose R is a commutative ring and that $a \in R$. Then a is a* gcd *of a and* 0.

Proof. a is a common divisor of a and 0, since $a|a$ and $a|0$; it is a *greatest* common divisor for if $c|a$ and $c|0$, then $c|a$. □

Lemma 7.1.10. *Suppose R is a commutative ring and that $a, b, q, r \in R$ satisfy $a = qb + r$. Then d is* gcd *of a and b if and only if d is a* gcd *of b and r.*

Proof. This is because the equation $a = qb + r$ shows an element $c \in R$ divides a and b if and only it divides b and r.

To expand, suppose d is a gcd for a and b. Then, in particular, $d|a$ and $d|b$, which means $d|b$ and $d|r$, so that d is a common divisor of b and r. If $c|b$ and $c|r$, then $c|a$ and $c|b$. Since d is a gcd for a and b, this tells us that $c|d$. Thus, d is a gcd for b and r. The other implication is proved in an identical manner. □

7.2 Calculating gcds: the Euclidean algorithm

Definition 7.2.1. A *Euclidean domain* is an integral domain R endowed with a map

$$d : R \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\}$$

such that for all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ with $a = qb + r$ and either $r = 0$, or $r \neq 0$ and $d(r) < d(b)$.

Theorem 7.2.2.

1. *The integers $\mathbb{Z}$ are a Euclidean domain when we define*

   $$d : \mathbb{Z} \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\}$$

   by $d(n) = |n|$.

2. *Polynomials with field coefficients $F[x]$ are a Euclidean domain when we define*

   $$d : F[x] \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\}$$

   by $d(f(x)) = \deg f(x)$.

Proof. This follows from corollary 5.2.2 and theorem 5.3.2. □

Theorem 7.2.3. *Suppose R is a Euclidean domain and that $a, b \in R$. There are elements $c, x, y \in R$ such that*

1. *c is a* gcd *for a and b;*
2. *$ax + by = c$ (Bezout's identity).*

Proof. The proof is called the *Euclidean algorithm*.

Suppose that R is a Euclidean domain with Euclidean function

$$d : R \setminus \{0\} \longrightarrow \mathbb{N} \cup \{0\}$$

and let $a, b \in R$.

If $b = 0$, then a is a gcd for a and b (lemma 7.1.9) and $a \cdot 1 + b \cdot 0 = a$, so we can take $c = a$, $x = 1$, $y = 0$.

If $b \neq 0$, let $r_0 = a$ and $r_1 = b$ and perform division as many times as possible until the remainder is 0: the d-value of the remainder strictly decreases each time we perform division with a non-zero remainder; since d takes values in $\mathbb{N} \cup \{0\}$, eventually the remainder must be zero. Supposing we have to perform n divisions, we have

$$\begin{aligned} r_0 &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ r_2 &= q_3 r_3 + r_4 \\ &\vdots \\ r_{n-1} &= q_n r_n + r_{n+1} \end{aligned}$$

and $r_{n+1} = 0$.

Because $r_{n+1} = 0$, r_n is a gcd for r_n and r_{n+1} (lemma 7.1.9). The last equation, together with lemma 7.1.10, tells us that r_n is a gcd for r_{n-1} and r_n. Applying lemma 7.1.10 $n - 1$ more times we see that r_n is a gcd for r_0 and r_1, i.e. a gcd for a and b. We take $c = r_n$.

Let $x_0 = 1$ and $y_0 = 0$ so that we have $ax_0 + by_0 = r_0$. Let $x_1 = 0$ and $y_1 = 1$ so that we have $ax_1 + by_1 = r_1$.

If $n = 1$, the proof is finished. Otherwise, let $1 \leq k < n$ and suppose inductively that for each $j \in \{0, 1, \ldots, k\}$ we have $x_j, y_j \in R$, such that $ax_j + by_j = r_j$. Then

$$r_{k+1} = r_{k-1} - q_k r_k = (ax_{k-1} + by_{k-1}) - q_k(ax_k + by_k) = a(x_{k-1} - q_k x_k) + b(y_{k-1} - q_k y_k).$$

Thus letting $x_{k+1} = x_{k-1} - q_k x_k$ and $y_{k+1} = y_{k-1} - q_k y_k$ we complete the inductive step. Taking $x = x_n$ and $y = y_n$ completes the proof. □

Remark 7.2.4. The Euclidean algorithm gives a particular gcd for a and b; there are other gcds which are associates of this particular gcd (remark 7.1.4).

The Euclidean algorithm for the integers is unique if we insist on using the stronger corollary 5.2.3 instead of corollary 5.2.2. Using this corollary ensures all remainders are taken to be positive and so we obtain our favorite gcd, that is, the positive gcd. For this reason, we *will* insist on using the stronger corollary and this is what we'll mean by the Euclidean algorithm for $\mathbb{Z}$.

The Euclidean algorithm for polynomials defined over a field is unique since each division is so. However, it is unlikely that the algorithm will return our favorite gcd; to obtain the monic one, we will have to multiply by a nonzero element of the field.

Example 7.2.5. Let's calculate a gcd for 42 and 30.

$$\begin{aligned} 42 &= 1 \cdot 30 + 12 \\ 30 &= 2 \cdot 12 + 6 \\ 12 &= 2 \cdot 6 + 0 \end{aligned}$$

The last non-zero remainder was 6, so 6 is a gcd for 42 and 30.

Moreover, rewriting the first two equations we get the following.

$$\begin{aligned} 12 &= 42 - 1 \cdot 30 \\ 6 &= 30 - 2 \cdot 12 \end{aligned}$$

Substituting the first into the second gives Bezout's identity

$$6 = 30 - 2 \cdot (42 - 1 \cdot 30) = 42 \cdot (-2) + 30 \cdot 3.$$

Example 7.2.6. Let's calculate a gcd for $x^3 - x$ and $x^3 - 3x^2 + 2x$ (elements of $\mathbb{Q}[x]$).

$$\begin{aligned} x^3 - x &= 1 \cdot (x^3 - 3x^2 + 2x) + (3x^2 - 3x) \\ x^3 - 3x^2 + 2x &= \frac{1}{3}(x - 2) \cdot (3x^2 - 3x) + 0. \end{aligned}$$

The last non-zero remainder was $3x^2 - 3x$ so $3x^2 - 3x$ is a gcd for the two polynomials. Rescaling by 3, we have $\gcd(x^3 - x, x^3 - 3x^2 + 2x) = x^2 - x$.

Doing the algorithm the other way (swapping the polynomials) gives $-3x^2 + 3x$ as a gcd.

$$\begin{aligned} x^3 - 3x^2 + 2x &= 1 \cdot (x^3 - x) + (-3x^2 + 3x) \\ x^3 - x &= -\frac{1}{3}(x + 1) \cdot (-3x^2 + 3x) + 0. \end{aligned}$$

Bezout's identity is obtained by rearranging the first equation we wrote down.

$$3x^2 - 3x = (x^3 - x) - (x^3 - 3x^2 + 2x).$$

8 Consequences of the Euclidean algorithm

8.1 Greatest common divisors

First up, we make note of some basic properties of the greatest common divisors for $\mathbb{Z}$ and $F[x]$. There's nothing surpirsing here. Basically, these corollaries say that gcds do what they think they should, but we're careful about units (recall definition 7.1.6).

Corollary 8.1.1. *Suppose $a, b, c \in \mathbb{Z}$. Then $\gcd(ac, bc) = \gcd(a, b) \cdot |c|$.*

Proof. $\gcd(ac, bc) = \gcd(a|c|, b|c|)$ since $x|y$ if and only if $x| \pm y$.

Run the Euclidean algorithm to calculate $\gcd(a, b)$. By multiplying all equations by $|c|$ we run the Euclidean algorithm for $\gcd(a|c|, b|c|)$. Thus, $\gcd(a|c|, b|c|) = \gcd(a, b) \cdot |c|$. □

Corollary 8.1.2. *Suppose F is a field, $f_1(x), f_2(x) \in F[x]$ and that*

$$g(x) = b_m x^m + \ldots + b_1 x + b_0 \in F[x], \text{ with } b_m \neq 0.$$

Then

$$\gcd(f_1(x)g(x), f_2(x)g(x)) = \gcd(f_1(x), f_2(x)) \cdot \frac{g(x)}{b_m}.$$

Proof. Running the Euclidean algorithm on $f_1(x)$ and $f_2(x)$, we obtain $c \cdot \gcd(f_1(x), f_2(x))$ for some non-zero element $c \in F$. By multiplying all equations by $g(x)$ we run the Euclidean algorithm for $f_1(x)g(x)$ and $f_2(x)g(x)$. It gives $c \cdot \gcd(f_1(x), f_2(x)) \cdot g(x)$. We multiply this by $\frac{1}{c \cdot b_m}$ to obtain a monic polynomial. □

8.2 Bezout's theorem

Bezout's theorem is the key to solving linear congruences $ax \equiv d \pmod{b}$.

Theorem 8.2.1 (Bezout)**.** *Given $a, b, d \in \mathbb{Z}$, we can find $x, y \in \mathbb{Z}$ with $ax + by = d$ if and only if $\gcd(a, b)|d$.*

Proof. $\gcd(a, b)|a$ and $\gcd(a, b)|b$ and so, if $ax + by = d$, then $\gcd(a, b)|d$.

Conversely, the Euclidean algorithm gives us $\tilde{x}, \tilde{y} \in \mathbb{Z}$ so that $a\tilde{x} + b\tilde{y} = \gcd(a, b)$. If $\gcd(a, b)|d$, there is a $c \in \mathbb{Z}$ such that $\gcd(a, b) \cdot c = d$. Then

$$a(c\tilde{x}) + b(c\tilde{y}) = \gcd(a, b) \cdot c = d,$$

so we can take $x = c\tilde{x}$ and $y = c\tilde{y}$. □

There's a polynomial version useful for congruences involving polynomials.

Theorem 8.2.2 (Bezout)**.** *Let F be a field. Given $f_1(x), f_2(x), h(x) \in F[x]$, there are $g_1(x), g_2(x)$ in $F[x]$ with*

$$f_1(x)g_1(x) + f_2(x)g_2(x) = h(x).$$

if and only if $\gcd(f_1(x), f_2(x))$ divides $h(x)$.

Proof. Basically it is the same as the previous one but one should be more careful about units.

$\gcd(f_1(x), f_2(x))$ divides both $f_1(x)$ and $f_2(x)$, and so, if $f_1(x)g_1(x) + f_2(x)g_2(x) = h(x)$, then $\gcd(f_1(x), f_2(x))$ divides $h(x)$.

Conversely, the Euclidean algorithm gives us $\tilde{g}_1(x), \tilde{g}_2(x) \in F[x]$ so that $f_1(x)\tilde{g}_1(x)+f_2(x)\tilde{g}_2(x) = c \cdot \gcd(f_1(x), f_2(x))$ for some nonzero element c of the field F.

If $\gcd(f_1(x), f_2(x))|h(x)$, there is a $g(x) \in F[x]$ such that $\gcd(f_1(x), f_2(x)) \cdot g(x) = h(x)$. Then

$$f_1(x) \cdot \frac{g(x)\tilde{g}_1(x)}{c} + f_2(x) \cdot \frac{g(x)\tilde{g}_2(x)}{c} = \gcd(f_1(x), f_2(x)) \cdot g(x) = h(x)$$

so we can take $g_1(x) = \frac{g(x)\tilde{g}_1(x)}{c}$ and $g_2(x) = \frac{g(x)\tilde{g}_2(x)}{c}$. □

8.3 Examples

Example 8.3.1. Solve $30x \equiv 24 \pmod{42}$.

This one you could try some values and you would discover $30 \cdot 12 = 360 \equiv 24 \pmod{42}$. There are other solutions: since $30 \cdot 7 = 210 \equiv 0 \pmod{42}$, $x = 12 + 7k$ is a solution for all $k \in \mathbb{Z}$.

How could you have figured this out without trial and error? To solve $30x \equiv 24 \pmod{42}$ we need to find $x, y \in \mathbb{Z}$ with $30x + 42y = 24$. Bezout's theorem tells us that we can do this as long as $\gcd(42, 30)|24$.

In example 7.2.5, we calculated $\gcd(42, 30)$ using the Euclidean algorithm. We discovered

$$\gcd(42, 30) = 6 = 30 \cdot 3 - 42 \cdot 2.$$

First, since $6|24$, this tells us immediately that we can solve the equation we were given. Secondly, it tells us that $30 \cdot 3 \equiv 6 \pmod{42}$. Multiplying by 4 gives $30 \cdot 12 \equiv 24 \pmod{42}$.

Finally, $\frac{42}{\gcd(42,30)} = 7$. Thus

$$30 \cdot 7 = 30 \cdot \frac{42}{\gcd(42, 30)} = \frac{30}{\gcd(42, 30)} \cdot 42 \equiv 0 \pmod{42}.$$

It turns out that this $\frac{42}{\gcd(42,30)} = 7$ trick gives the minimal solution to $30x \equiv 0 \pmod{42}$. I don't think we'll need this, but we'll see something similar in theorem 10.1.11.

Example 8.3.2. Solve $30x \equiv 23 \pmod{42}$.

We cannot solve this because 6 does not divide 23.

Example 8.3.3. Solve $35x \equiv 1 \pmod{221}$.

We run the Euclidean algorithm to find the gcd of 35 and 221.

$$\begin{aligned} 221 &= 6 \cdot 35 + 11 \\ 35 &= 3 \cdot 11 + 2 \\ 11 &= 5 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0 \end{aligned}$$

Thus $\gcd(221, 35) = 1$ and the congruence is solvable. Rearranging the above equations gives

$$\begin{aligned} 11 &= 221 - 6 \cdot 35 \\ 2 &= 35 - 3 \cdot 11 \\ 1 &= 11 - 5 \cdot 2 \end{aligned}$$

Thus,

$$\begin{aligned} 1 &= 11 - 5 \cdot 2 \\ &= 11 - 5 \cdot (35 - 3 \cdot 11) = 16 \cdot 11 - 5 \cdot 35 \\ &= 16 \cdot (221 - 6 \cdot 35) - 5 \cdot 35 = 16 \cdot 221 - 101 \cdot 35 \end{aligned}$$

and we conclude that $35 \cdot (-101) \equiv 1 \pmod{221}$.

We just showed 35 is a unit in $\mathbb{Z}/221$ and found its inverse (c.f. homework 3, question 3)c)).

9 Units in our favorite rings

9.1 Coprimality

Before stating our main results about units in our favorite rings we need some terminology.

Definition 9.1.1. We say $a, b \in \mathbb{Z}$ are *coprime* if $\gcd(a, b) = 1$.

Remark 9.1.2. Bezout's theorem tells us that $a, b \in \mathbb{Z}$ are coprime if and only if there are $x, y \in \mathbb{Z}$ with $ax + by = 1$.

Definition 9.1.3. Suppose F is a field. We say $f(x), g(x) \in F[x]$ are *coprime* if $\gcd(f(x), g(x)) = 1$.

Remark 9.1.4. Bezout's theorem tells us that $f_1(x), f_2(x) \in F[x]$ are coprime if and only if there are $g_1(x), g_2(x) \in F[x]$ with $f_1(x)g_1(x) + f_2(x)g_2(x) = 1$.

The following result could probably be credited to Euclid.

Proposition 9.1.5 (Euclids's lemma)**.** *Suppose that either $R = \mathbb{Z}$, or $R = F[x]$ for some field F. Suppose, in addition, that $a, b, c \in R$, $a|bc$, and a and b are coprime. Then $a|c$.*

Proof. Since a and b are coprime there are elements $x, y \in R$ with $ax+by = 1$. Then $acx+bcy = c$. Since $a|a$ and $a|bc$, we conclude that $a|c$. □

9.2 Modular arithmetic

Suppose we ask the following question: given $[x]_n \in \mathbb{Z}/n$, is there a $[y]_n \in \mathbb{Z}/n$ with $[x]_n[y]_n = 1$? By definition of multiplication and what it means to be equal in $\mathbb{Z}/n$, the answer is "yes" precisely when there are $c, y \in \mathbb{Z}$ with

$$xy + cn = 1.$$

Bezout's theorem says this happens exactly when x and n are coprime.

Theorem 9.2.1. *$[x]_n$ is a unit in $\mathbb{Z}/n$ if and only if $x \in \mathbb{Z}$ and n are coprime.*

Example 9.2.2. The units in $\mathbb{Z}/221$ are

$$\{[x]_{221} : \gcd(x, 221) = 1\} = \{[x]_{221} : 13 \text{ and } 17 \text{ do not divide } x\}.$$

Corollary 9.2.3. *If $p \in \mathbb{N}$ is a prime, then $\mathbb{Z}/p$ is a field.*

In fact, $\mathbb{Z}/p$ is a field if and only if $p \in \mathbb{N}$ is a prime (quick exercise).

9.3 Modular arithmetic with polynomials

Similarly, we have the following theorem.

Theorem 9.3.1. *Suppose F is a field. Then $[f(x)]_{q(x)}$ is a unit in $F[x]/(q(x))$ if and only if $f(x)$ and $q(x)$ are coprime.*

Using a lemma which we'll see a little later (lemma 14.1.3), we also have following corollary. Here irreducible means "cannot be factored any further" (see definition 14.1.1).

Corollary 9.3.2. *Suppose F is a field and $q(x) \in F[x]$ is irreducible. Then $F[x]/(q(x))$ is a field.*

In fact, $F[x]/(q(x))$ is a field if and only if $q(x) \in F[x]$ is irreducible (homework).

10 The theorems of Fermat, Euler, and Lagrange

10.1 Exponents of units in $\mathbb{Z}/n$

Definition 10.1.1. If R is a commutative ring, write $U(R)$ for the set of units in R.

Definition 10.1.2 (Euler's totient function)**.** For $n \geq 2$, let $\varphi(n)$ be the number of units in $\mathbb{Z}/n$,

$$\text{i.e. } \varphi(n) = |U(\mathbb{Z}/n)|.$$

Remark 10.1.3. By theorem 9.2.1 $\varphi(n)$ is equal to the number of integers x with $1 \leq x < n$, that are coprime to n.

Example 10.1.4. If p is a prime then $\mathbb{Z}/p$ is a field so that

$$U(\mathbb{Z}/p) = \{[1]_p, \ldots, [p-1]_p\} \text{ and } \varphi(p) = p - 1.$$

Example 10.1.5.

$$U(\mathbb{Z}/4) = \{\pm 1\},\ U(\mathbb{Z}/6) = \{\pm 1\},\ U(\mathbb{Z}/8) = \{\pm 1, \pm 3\},\ U(\mathbb{Z}/9) = \{\pm 1, \pm 2, \pm 4\},$$
$$U(\mathbb{Z}/10) = \{\pm 1, \pm 3\},\ U(\mathbb{Z}/12) = \{\pm 1, \pm 5\},\ U(\mathbb{Z}/14) = \{\pm 1, \pm 3, \pm 5\}$$
$$U(\mathbb{Z}/15) = \{\pm 1, \pm 2, \pm 4, \pm 7\},\ U(\mathbb{Z}/16) = \{\pm 1, \pm 3, \pm 5, \pm 7\},\ U(\mathbb{Z}/18) = \{\pm 1, \pm 5, \pm 7\},$$

so $\varphi(4) = 2$, $\varphi(6) = 2$, $\varphi(8) = 4$, $\varphi(9) = 6$, $\varphi(10) = 4$, $\varphi(12) = 4$, $\varphi(14) = 6$, $\varphi(15) = 8$, $\varphi(16) = 8$, and $\varphi(18) = 6$.

Theorem 10.1.6 (Euler's theorem)**.** *Suppose x and n are coprime. Then $x^{\varphi(n)} \equiv 1\ (n)$.*

Proof. Since x and n are coprime, $[x]_n$ is a unit in $\mathbb{Z}/n$ (theorem 9.2.1). This means there is a $[y]_n$ with $[x]_n[y]_n = 1$. Let $U(\mathbb{Z}/n) = \{[x]_n : [x]_n \text{ is a unit }\}$. The map

$$U(\mathbb{Z}/n) \longrightarrow U(\mathbb{Z}/n), \quad u \longmapsto [x]_n \cdot u$$

is a bijection; the inverse is given by "multiplication by $[y]_n$." Thus,

$$[x]_n^{\varphi(n)} \cdot \left[\prod_{u \in U(\mathbb{Z}/n)} u\right] = \prod_{u \in U(\mathbb{Z}/n)} \left[[x]_p \cdot u\right] = \prod_{u \in U(\mathbb{Z}/n)} u = 1 \cdot \left[\prod_{u \in U(\mathbb{Z}/n)} u\right].$$

Cancelling $\left[\prod_{u \in U(\mathbb{Z}/n)} u\right]$ (multiplying by its multiplicative inverse) gives $[x]_n^{\varphi(n)} = 1$. □

Corollary 10.1.7 (Fermat's little theorem)**.** *If p is a prime and $x \in \mathbb{Z}$ is not divisible by p then $x^{p-1} \equiv 1$ (mod p).*

Proof. $\varphi(p) = p - 1$, since every number y with $1 \leq y < p$ is coprime to p or, alternatively, because $\mathbb{Z}/p$ is a field. □

Definition 10.1.8. Suppose x and n are coprime. The smallest $e \in \{1, \ldots, \varphi(n)\}$ with $x^e \equiv 1\ (n)$ is called *the order of x modulo n*. We'll write $\text{ord}_n(x)$ for the order of x modulo n.

Example 10.1.9. Let's calculate the order of some elements in $\mathbb{Z}/31$.

We have $\mathrm{ord}_{31}(1) = 1$. Here is the sequence $(3^n \pmod{31})_{n=1}^{30}$:

$$3,\ 9,\ -4,\ -12,\ -5,\ -15,\ -14,\ -11,\ -2,\ -6,\ 13,\ 8,\ -7,\ 10,$$

$$-1,\ -3,\ -9,\ 4,\ 12,\ 5,\ 15,\ 14,\ 11,\ 2,\ 6,\ -13,\ -8,\ 7,\ -10,\ 1.$$

This shows $\mathrm{ord}_{31}(3) = 30$. We also see that

$$\mathrm{ord}_{31}(9) = \mathrm{ord}_{31}(3^2) = 15,\ \mathrm{ord}_{31}(-4) = \mathrm{ord}_{31}(3^3) = 10,\ \mathrm{ord}_{31}(-5) = \mathrm{ord}_{31}(3^5) = 6,$$
$$\mathrm{ord}_{31}(-15) = \mathrm{ord}_{31}(3^6) = 5,\ \mathrm{ord}_{31}(-6) = \mathrm{ord}_{31}(3^{10}) = 3,\ \mathrm{ord}_{31}(-1) = \mathrm{ord}_{31}(3^{15}) = 2.$$

We can obtain the order of other elements using the theorem following the next.

Notice that all the elements have order dividing $30 = \varphi(31)$. This is not a coincidence.

Theorem 10.1.10. *Suppose x and n are coprime and that $x^d \equiv 1\ (n)$. Then the order of x modulo n divides d.*

Proof. Let e denote the order of x modulo n and suppose $x^d \equiv 1\ (n)$. Since $e \neq 0$, we can divide

$$d = qe + r$$

where $0 \leq r < e$ (corollary 5.2.3). Then

$$1 \equiv x^d \equiv x^{qe+r} \equiv (x^e)^q \cdot x^r \equiv 1 \cdot x^r \equiv x^r\ (n).$$

Since r is smaller than e and $x^r \equiv 1\ (n)$ we must have $r = 0$, since e is the smallest strictly positive number with $x^e \equiv 1\ (n)$. Thus, e divides d. □

Theorem 10.1.11. *Suppose x and n are coprime and that x has order e modulo n. Then the order of x^d modulo n is $e/\gcd(d, e)$. (Recall that $\gcd(d, e)$ denotes the positive gcd of d and e.)*

Proof. Suppose $(x^d)^s \equiv 1\ (n)$. Then the previous theorem says that $e|ds$. Thus, $\frac{e}{\gcd(d,e)}|\frac{ds}{\gcd(d,e)}$. Using corollary 8.1.1 we have

$$\gcd\left(\frac{d}{\gcd(d,e)}, \frac{e}{\gcd(d,e)}\right) = 1.$$

Thus, $\frac{e}{\gcd(d,e)}|s$, by proposition 9.1.5. In particular, if $s > 0$, then $s \geq \frac{e}{\gcd(d,e)}$. Finally, note that

$$(x^d)^{\frac{e}{\gcd(d,e)}} \equiv (x^e)^{\frac{d}{\gcd(d,e)}} \equiv 1^{\frac{d}{\gcd(d,e)}} \equiv 1\ (n).$$

□

10.2 Another proof of Fermat's little theorem

Let's give a different proof of Fermat's little theorem which makes use of the binomial theorem.

Theorem 10.2.1 (Binomial theorem)**.** *In $\mathbb{Z}[x, y]$ (the polynomial ring over $\mathbb{Z}$ in two variables) we have*

$$(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

where $\binom{n}{i} = \frac{n!}{i!(n-i)!}$.

This result has been stated for polynomials in two variables and this implies the result in any commutative ring. [In grown up talk, this is because $\mathbb{Z}[x, y]$ is the free ring on two variables but don't worry about this now.] That is, if R is a commutative ring and $x, y \in R$, then $(x + y)^n$ can be expressed in the same way.

Lemma 10.2.2. *If $p \in \mathbb{N}$ is prime and $i \in \{1, 2, \ldots, p - 1\}$, then p divides $\binom{p}{i}$.*

Proof. We use the "fundamental theorem of arithmetic" which says that any natural number can be factored uniquely into primes. We'll prove this in section 14.

$p|p!$ and so p appears in the prime decomposition of $p!$. Since $\binom{p}{i} \in \mathbb{N}$ we have $i!(p - i)!|p!$. p does not appear in the prime decomposition of $i!(p - i)!$ since, by definition of factorial, we have expressed it as a product of numbers less than p. Thus, p still appears in the prime decomposition of $\binom{p}{i} = \frac{p!}{i!(p-i)!}$. □

Corollary 10.2.3 (Freshman dream)**.** *In $\mathbb{Z}/p[x, y]$ we have $(x + y)^p = x^p + y^p$.*

As for the binomial theorem, we stated this for polynomials in two variables (defined over $\mathbb{Z}/p$). This is because it implies the result for any ring in which adding an element to itself p times gives 0. [In grown up talk, this is because $\mathbb{Z}/p[x, y]$ is the free ring of characteristic p on two variables but don't worry about this now.]

Theorem 10.2.4 (Fermat's little theorem)**.** *Suppose $x \in \mathbb{Z}$. Then $x^p \equiv x \pmod{p}$.*

Proof. It is enough to check the result for $x \in \mathbb{N}$, since any $x \in \mathbb{Z}$ is congruent to some $y \in \mathbb{N}$.

The result is true for 0 and 1 so we proceed by induction. Suppose that $x^p \equiv x \pmod{p}$. Then

$$(x + 1)^p \equiv x^p + 1^p \equiv x + 1 \pmod{p}$$

where the first congruence is the freshman dream, and the second is the inductive hypothesis. This completes the proof. □

10.3 A tiny bit of group theory

Definition 10.3.1. A *group* G is a set equipped with a multiplication

$$* : G \times G \longrightarrow G,\ (g, h) \longmapsto g * h$$

satisfying the following three axioms.

- (G1) $g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3$ for all $g_1, g_2, g_3 \in G$ (associativity)
- (G2) there is an element e with the property that $g * e = g = e * g$ for all $g \in G$ (identity)
- (G3) for each $g \in G$ there is an element $h \in G$ such that $g * h = e = h * g$ (inverses).

We say a group is *abelian* is $g * h = h * g$ for all $g, h \in G$.

Example 10.3.2.

1. Let $\Sigma_n = \{\sigma : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\} : \sigma \text{ is a bijection}\}$. Then $(\Sigma_n, \circ)$ is a group.
2. Let $R(\square) = \{\text{rotations of a cube}\}$. This is a group because we can do one rotation followed by another and we still get a rotation. In fact, $R(\square)$ is basically the same as Σ_4: follow what happens to the diagonals of the cube under a rotation.
3. Let $R(\Delta) = \{\text{rotations of a tetrahedron}\}$. In fact, $R(\Delta)$ is the same as something called A_4, which lives inside Σ_4: follow what happens to the vertices of the tetrahedron under a rotation.

More relevant for us is the following example.

Example 10.3.3.
1. Suppose R is a ring. Then $(R, +)$ is an abelian group.
2. Suppose R is a ring. Let $U(R) = \{u \in R : u \text{ is a unit in } R\}$. Then $(U(R), \cdot)$ is a group. If R is a commutative ring then $(U(R), \cdot)$ is an abelian group.

Theorem 10.3.4 (Lagrange)**.** *Suppose G is a finite group, that $|G| = n$ and $g \in G$. Then $g^n = e$.*

Since $\varphi(n) = |U(\mathbb{Z}/n)|$ we obtain Euler's theorem as a corollary.

Corollary 10.3.5 (Euler's theorem)**.** *Suppose $[x]_n \in U(\mathbb{Z}/n)$. Then $[x]_n^{\varphi(n)} = 1$.*

11 The chinese remainder theorem

11.1 $\mathbb{Z}/6 \cong \mathbb{Z}/2 \times \mathbb{Z}/3$

Here is the multiplication table for $\mathbb{Z}/6$.

$\cdot$	0	1	2	3	-2	-1
0	0	0	0	0	0	0
1	0	1	2	3	-2	-1
2	0	2	-2	0	2	-2
3	0	3	0	3	0	3
-2	0	-2	2	0	-2	2
-1	0	-1	-2	3	2	1

Let's relabel the elements.

$$0 \longmapsto (0,0),\ 1 \longmapsto (1,1),\ 2 \longmapsto (0,-1),\ 3 \longmapsto (1,0),\ -2 \longmapsto (0,1),\ -1 \longmapsto (1,-1).$$

We get

$\cdot$	$(0,0)$	$(1,1)$	$(0,-1)$	$(1,0)$	$(0,1)$	$(1,-1)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(1,1)$	$(0,0)$	$(1,1)$	$(0,-1)$	$(1,0)$	$(0,1)$	$(1,-1)$
$(0,-1)$	$(0,0)$	$(0,-1)$	$(0,1)$	$(0,0)$	$(0,-1)$	$(0,1)$
$(1,0)$	$(0,0)$	$(1,0)$	$(0,0)$	$(1,0)$	$(0,0)$	$(1,0)$
$(0,1)$	$(0,0)$	$(0,1)$	$(0,-1)$	$(0,0)$	$(0,1)$	$(0,-1)$
$(1,-1)$	$(0,0)$	$(1,-1)$	$(0,1)$	$(1,0)$	$(0,-1)$	$(1,1)$

This is the multiplication table for $\mathbb{Z}/2 \times \mathbb{Z}/3$.

You can check that the same thing happens with the addition tables. $\mathbb{Z}/6$ is basically the same as $\mathbb{Z}/2 \times \mathbb{Z}/3$.

11.2 $\mathbb{Z}/4 \ncong \mathbb{Z}/2 \times \mathbb{Z}/2$

The previous example might make you believe that $\mathbb{Z}/mn \cong \mathbb{Z}/m \times \mathbb{Z}/n$. In general, this is not true. Take any element of $\mathbb{Z}/2 \times \mathbb{Z}/2$ and add it to itself; you get zero. That is, for all $x \in \mathbb{Z}/2 \times \mathbb{Z}/2$

$$x + x = 0.$$

The element $1 \in \mathbb{Z}/4$, has $1 + 1 \neq 0$. So $\mathbb{Z}/4$ is not the same as $\mathbb{Z}/2 \times \mathbb{Z}/2$.

Let's do one more example to try and spot a pattern.

11.3 $\mathbb{Z}/12 \subset \mathbb{Z}/4 \times \mathbb{Z}/6$

Here is a quarter of the multiplication table for $\mathbb{Z}/12$.

$\cdot$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	-4	-2	0
3	0	3	6	-3	0	3	6
4	0	4	-4	0	4	-4	0
5	0	5	-2	3	-4	1	6
6	0	6	0	6	0	6	0

Let's relabel the elements

$$0 \longmapsto (0,0),\ 1 \longmapsto (1,1),\ 2 \longmapsto (2,2),\ 3 \longmapsto (-1,3),\ 4 \longmapsto (0,-2),\ 5 \longmapsto (1,-1),$$

$$6 \longmapsto (2,0),\ 7 \longmapsto (-1,1),\ 8 \longmapsto (0,2),\ 9 \longmapsto (1,3),\ 10 \longmapsto (2,-2),\ 11 \longmapsto (-1,-1).$$

We get a sub-table of the multiplication table for $\mathbb{Z}/4 \times \mathbb{Z}/6$. $\mathbb{Z}/12$ is basically the same as

$$\Big\{([a]_4, [b]_6) \in \mathbb{Z}/4 \times \mathbb{Z}/6 \ :\ 2|(a-b)\Big\}.$$

Notice that $12 = \frac{4 \cdot 6}{\gcd(4,6)}$ and $2 = \gcd(4,6)$. This will make more sense shortly (11.4.3).

11.4 Congruences

Suppose that we wish to solve the congruences

$$x \equiv a \pmod{m}$$
$$x \equiv b \pmod{n}.$$

If $x \in \mathbb{Z}$ is a solution then

$$x = a + cm$$
$$x = b + dn$$

for some $c, d \in \mathbb{Z}$. So $a + cm = b + dn$ giving $a - b = dn - cm$. Bezout tells us that $\gcd(m,n)|(a-b)$. Moreover, if $\gcd(m,n)|(a-b)$ such $c, d \in \mathbb{Z}$ exist and we can solve the congruences.

If $x' \in \mathbb{Z}$ is another solution to the congruences, then $y = x - x'$ satisfies the congruences

$$y \equiv 0 \pmod{m}$$
$$y \equiv 0 \pmod{n}.$$

This says that $m|y$ and $n|y$. By writing m, n, and y out in their prime decompositions (see section 14), we conclude that $\frac{mn}{\gcd(m,n)}|y$. If this bothers you, you can read the proof of the following lemma (shout out, Songlin).

Lemma 11.4.1. *Suppose $m, n, x \in \mathbb{Z}$, $m|x$, and $n|x$. Then $\frac{mn}{\gcd(m,n)}|x$.*

Proof. Since $m|x$ we have an $a \in \mathbb{Z}$ so that $x = ma$ and because $n|x = ma$ this gives

$$\frac{n}{\gcd(m,n)} \Bigg| \frac{m}{\gcd(m,n)} a.$$

Using corollary 8.1.1 followed by proposition 9.1.5 we obtain

$$\gcd\left(\frac{m}{\gcd(m,n)}, \frac{n}{\gcd(m,n)}\right) = 1$$

and $\frac{n}{\gcd(m,n)}|a$, and so there is a $b \in \mathbb{Z}$ with $a = \frac{n}{\gcd(m,n)}b$. We conclude that $x = ma = \frac{mn}{\gcd(m,n)}b$, i.e $\frac{mn}{\gcd(m,n)}|x$. □

We have proved the following theorem.

Theorem 11.4.2 (Chinese remainder theorem (Sun Ze))**.** *Suppose we wish to solve the congruences*

$$x \equiv a \ (mod\ m)$$
$$x \equiv b \ (mod\ n).$$

A solution exists if and only if $\gcd(m,n)|(a-b)$ *and in this case it can be found using the extended Euclidean algorithm.*

If x and x' are two solutions to the congruences then

$$x \equiv x' \ \left(mod\ \ \frac{mn}{\gcd(m,n)}\right).$$

This relates to the previous examples $\mathbb{Z}/6 \cong \mathbb{Z}/2 \times \mathbb{Z}/3$ and $\mathbb{Z}/12 \subset \mathbb{Z}/4 \times \mathbb{Z}/6$ in the following way.

Theorem 11.4.3 (Chinese remainder theorem (alternative statement))**.** *Let $m, n \in \mathbb{N}$ and*

$$l = \frac{mn}{\gcd(m,n)}.$$

The function $\mathbb{Z}/l \longrightarrow \mathbb{Z}/m \times \mathbb{Z}/n, \ [x]_l \longmapsto ([x]_m, [x]_n)$ is well-defined, has range

$$\{([a]_m, [b]_n) : \gcd(m,n)|(a-b)\}$$

and is one-to-one.

Corollary 11.4.4 (Chinese remainder theorem)**.** *Suppose that m and n are coprime then $\mathbb{Z}/mn \cong \mathbb{Z}/m \times \mathbb{Z}/n$. In particular, $x \equiv a \ (mod\ mn)$ if and only if $x \equiv a \ (mod\ m)$ and $x \equiv a \ (mod\ n)$. Also,*

$$\varphi(mn) = \varphi(m)\varphi(n).$$

Everything just said works for polynomials too. We might need this later.

Theorem 11.4.5 (Chinese remainder theorem for polynomials)**.** *Suppose that we wish to solve the congruences*

$$f(x) \equiv a(x) \ (mod\ p(x))$$
$$f(x) \equiv b(x) \ (mod\ q(x)).$$

A solution exists if and only if $\gcd(p(x), q(x))|(a(x) - b(x))$ *and in this case it can be found using the extended Euclidean algorithm.*

If $f_0(x)$ *and* $f_1(x)$ *are two solutions to the congruences then*

$$f_0(x) \equiv f_1(x) \quad \left(mod \ \ \frac{p(x)q(x)}{\gcd(p(x), q(x))}\right).$$

11.5 Examples

I truly messed up the following example in class. I'm sorry about this.

Example 11.5.1. Solve the congruences

$$x \equiv 3 \ (\text{mod } 13)$$
$$x \equiv 4 \ (\text{mod } 17).$$

For a solution to exist we need $\gcd(13, 17)|(4 - 3)$. This is fine since $1|1$.

In the previous subsection we said that we can use the extended Euclidean algorithm to find a solution. Let's see this in action.

$$17 = 1 \cdot 13 + 4$$
$$13 = 3 \cdot 4 + 1$$
$$4 = 4 \cdot 1 + 0$$

demonstrates that $\gcd(13, 17) = 1$. We also obtain

$$4 = 17 - 13$$
$$1 = 13 - 3 \cdot 4$$

so that $1 = 13 - 3(17 - 13) = 4 \cdot 13 - 3 \cdot 17$, the Bezout identity.

This is the point in lectures where I spaced out and became confused because a new 3 and 4 have shown up, different to the 3 and 4 in the original congruences we wish to solve. We note that

$$4 - 3 = 1 = 4 \cdot 13 - 3 \cdot 17$$

so that $3 + 4 \cdot 13 = 4 + 3 \cdot 17$. Let

$$x = 55 = 3 + 4 \cdot 13 = 4 + 3 \cdot 17.$$

Then $x \equiv 3 \ (13)$ and $x \equiv 4 \ (17)$. Since $\frac{13 \cdot 17}{\gcd(13,17)} = 221$ the solution is unique up to a multiple of 221, i.e. all the solutions are of the form

$$x = 55 + 221k \text{ for } k \in \mathbb{Z}.$$

Example 11.5.2. Solve the congruences

$$x \equiv 3 \pmod{13}$$
$$x \equiv 4 \pmod{17}.$$

Normally when you solve simultaneous equations you try substituting one equation into the other. We can do that here. Solutions to the second equation are of the form $4 + 17n$. So we can try and solve

$$4 + 17n \equiv 3 \pmod{13}.$$

This is equivalent to $4n \equiv -1 \pmod{13}$. We see that $n \equiv 3 \pmod{13}$ so $17n \equiv 17 \cdot 3 \pmod{13 \cdot 17}$ and

$$x \equiv 4 + 17 \cdot 3 \equiv 55 \pmod{221}.$$

Let's do another example.

Example 11.5.3. Solve

$$x \equiv 2 \pmod{35}$$
$$x \equiv 23 \pmod{49}.$$

For a solution to exist we need $\gcd(35, 49) | (23 - 2)$. This is fine since $7|21$.

The Euclidean algorithm for $\gcd(35, 49)$ looks as follows.

$$49 = 1 \cdot 35 + 14$$
$$35 = 2 \cdot 14 + 7$$
$$14 = 2 \cdot 7 + 0.$$

Thus,

$$14 = 49 - 35$$
$$7 = 35 - 2 \cdot 14,$$

which gives $7 = 35 - 2(49 - 35) = 3 \cdot 35 - 2 \cdot 49$, the Bezout identity.

This allows us to write

$$23 - 2 = 3 \cdot 7 = 9 \cdot 35 - 6 \cdot 49$$

so that $2 + 9 \cdot 35 = 23 + 6 \cdot 49$. Let

$$x = 317 = 2 + 9 \cdot 35 = 23 + 6 \cdot 49.$$

Then $x \equiv 2 \ (35)$ and $x \equiv 23 \ (49)$. Since $\frac{35 \cdot 49}{\gcd(35,49)} = 245$ the solution is unique up to a multiple of 245. Making use of the fact that $317 - 245 = 72$ we see that all solutions are of the form

$$72 + 245k \text{ for } k \in \mathbb{Z}.$$

Example 11.5.4. Solve

$$x \equiv 2 \pmod{35}$$
$$x \equiv 23 \pmod{49}.$$

We need to solve $23 + 49n \equiv 2 \pmod{35}$. This is equivalent to $14n \equiv 14 \pmod{35}$. So $n \equiv 1 \pmod{5}$ works (notice the 35 changed to a 5, since $14 \cdot 5 \equiv 0 \pmod{35}$). Thus,

$$x \equiv 23 + 49n \equiv 23 + 49 \equiv 72 \pmod{245}.$$

Example 11.5.5. Find all the solutions to the following congruences

$$x \equiv 2 \pmod{12}$$
$$x \equiv 8 \pmod{10}$$
$$x \equiv 9 \pmod{13}.$$

We ignore the third one for now. We can solve the first two since $\gcd(10, 12) = 2$, $8 - 2 = 6$ and $2|6$. We run the extended Euclidean algorithm (which is very short in this case) to write

$$2 = 12 - 10.$$

Thus, $8 - 2 = 3 \cdot 12 - 3 \cdot 10$ giving $x = 38 = 2 + 3 \cdot 12 = 8 + 3 \cdot 10$ as a solution. Since $\frac{10 \cdot 12}{\gcd(10,12)} = 60$ we see that all solutions of the first two congruences are given by

$$x = 38 + 60k \text{ for } k \in \mathbb{Z}.$$

In particular, we can replace the first two congruences by $x \equiv 38 \pmod{60}$, so we now have to solve

$$x \equiv 38 \pmod{60}$$
$$x \equiv 9 \pmod{13}.$$

Since $\gcd(60, 13) = 1$, $38 - 9 = 29$ and $1|29$, this is possible.

We could run the extended Euclidean algorithm to write $1 = 5 \cdot 60 - 23 \cdot 13$, giving $38 - 9 = (29 \cdot 5) \cdot 60 - (29 \cdot 23) \cdot 13$. We would obtain $x = -8662 = 38 - (29 \cdot 5) \cdot 60 = 9 - (29 \cdot 23) \cdot 13$ as a solution. Since $\frac{13 \cdot 60}{\gcd(13,60)} = 780$ and $698 = -8662 + 780 \cdot 12$ we see that $698 + 780k$ for $k \in \mathbb{Z}$ are the solutions.

Alternatively, solving these last two congruences is equivalent to finding an $n \in \mathbb{Z}$ such that $38 + 60n \equiv 9 \pmod{13}$. This is the same as solving $-1 - 5n \equiv -4 \pmod{13}$, which rearranges to $5n \equiv 3 \pmod{13}$. Multiplying by -5 gives $n \equiv -2 \pmod{13}$, so that $x = 38 - 60 \cdot 2 = -82$ is a solution. Since $-82 + 780 = 698$ this agrees with above.

12 RSA

RSA is a cryptosystem. Suppose two parties, call them Alice and Bob, wish to send messages back and forth to each other and want to be be incomprehensible to a third party, say Eve. We may as well assume our messages consist of numbers.

12.1 The procedure

Bob does the following:

- He chooses two different large prime numbers p and q.
- He lets m (the modulus) be the product of the primes pq.
- He chooses a fairly small number $e > 0$ (the encrypting exponent) which is coprime to $\varphi(m)$.
- He finds a number $d > 0$ (the decrypting exponent) such that $ed \equiv 1 \pmod{\varphi(m)}$.
- He tells Alice m and e but keeps everything else secret.

Alice does the following:

- Alice has a message that consists of a sequence of numerical words. What we mean by this is that each "word" is a number $w \in \{0, 1, \ldots, m-1\}$.
- To encrypt the word w she finds a number $c \in \{0, 1, \ldots, m-1\}$ such that

$$c \equiv w^e \pmod{m}.$$

- She sends Bob the encrypted words.

Bob does the following:

- For each encrypted word c which Bob recieves from Alice he finds a number

$$w' \in \{0, 1, \ldots, m-1\}$$

such that

$$w' \equiv c^d \pmod{m}.$$

- As if by magic, it turns out that $w' = w$.

12.2 An example

Let's do an example with smaller numbers than would be used in reality.

Example 12.2.1. Bob lets $p = 7$ and $q = 11$, so that $m = 77$. We have $\varphi(77) = 60$. He chooses $e = 7$, and finds $d = 43$. He tells Alice that $m = 77$ and $e = 7$.

Alice wishes to send the one word message $w = 2$ so she calculates

$$c = 2^7 = 128 \equiv 51 \pmod{77}$$

and sends Bob her encrypted word 51.

Bob wishes to decrypt Alices message and so needs to calculate 51^{43} modulo 77. Bob realizes that the Chinese remainder theorem and Fermat's Little Theorem makes his calculation easier. He finds

$$51^{43} \equiv 2^{43} \equiv (2^6)^7 \cdot 2 \equiv 1 \cdot 2 \equiv 2 \pmod 7$$
$$51^{43} \equiv (-4)^{43} = ((-4)^{10})^4 \cdot (-4)^3 \equiv 1 \cdot 2 \equiv 2 \pmod{11}$$

and so $w' = 2$. This is precisely the message Alice sent him.

12.3 Checking that the RSA decryption works

The proof that RSA works. Let's carefully examine the procedure. So let p, q, m, e, d, w, c, w' be as in section 12.1.

The first bit of math Bob has to do is to find a number d such that $ed \equiv 1 \pmod{\varphi(m)}$. This is possible since he chooses e to be coprime to $\varphi(m)$ and this ensures $[e]_{\varphi(m)}$ is a unit in $\mathbb{Z}/\varphi(m)$.

We just need to check that the decryption works out. Since $ed \equiv 1 \pmod{\varphi(m)}$ there is a $k \in \mathbb{Z}$ such that

$$ed = 1 + \varphi(m)k.$$

Thus,

$$w' \equiv c^d \equiv (w^e)^d \equiv w^{1+\varphi(m)k} \pmod m.$$

Since $w, w' \in \{0, 1, \ldots, m-1\}$, we just have to show that $w^{1+\varphi(m)k} \equiv w \pmod m$.

The Chinese remainder theorem tells us two things:

1. $\varphi(m) = (p-1)(q-1)$;
2. it is enough to check $w^{1+\varphi(m)k} \equiv w \pmod p$ and $w^{1+\varphi(m)k} \equiv w \pmod q$.

The requisite congruences follow from Fermat's Little Theorem since

$$w^{1+\varphi(m)k} = w^{1+(p-1)(q-1)k} = w \cdot (w^{p-1})^{(q-1)k} \equiv w \pmod p;$$
$$w^{1+\varphi(m)k} = w^{1+(p-1)(q-1)k} = w \cdot (w^{q-1})^{(p-1)k} \equiv w \pmod q.$$

□

12.4 Why is RSA effective?

It is often the case that Bob will wish to receive messages from different people, and he will actually give out the numbers m and e to the public. If Alice sends Bob a message and Eve intercepts it, what does Eve need to do to decrypt the message?

- Eve needs to find the e-th root of c. She can do this as long as she can find d, the decrypting exponent.
- To find d, Eve would need to solve the congruence $ed \equiv 1 \pmod{\varphi(m)}$.
- To solve such a congruence Eve would need to know $\varphi(m)$.
  - If p and q are large, then m is large and finding $\varphi(m)$ by counting would take too long.
  - If Eve could find p and q she would know $\varphi(m)$, but factoring large numbers into primes is also difficult.

12.5 Other examples

Example 12.5.1. Bob decides on $p = 11$, $q = 17$, so that $m = 187$. $\varphi(m) = 10 \cdot 16 = 160$, and he chooses $e = 3$. Bob calculates d using a short extended Eucidean algorithm.

$$160 = 53 \cdot 3 + 1$$

so $3 \cdot (-53) \equiv 1 \pmod{160}$ and he takes $d = 160 - 53 = 107$.

Alice wants to send Bob $w = 127$ so she calculates

$$c \equiv 127^3 \equiv (-60)^3 \equiv -216,000 \equiv -216,000 + 1156 \cdot 187 \equiv 172 \pmod{187}$$

and transmits 172 to Bob.

Bob decodes using the Chinese remainder theorem and Fermat's Little Theorem.

$$172^{107} \equiv (-4)^7 \equiv (-4)(-4)^2(-4)^4 \equiv -4 \cdot 5 \cdot 3 \equiv -5 \pmod{11},$$
$$172^{107} \equiv 2^{11} \equiv 2 \cdot 2^2 \cdot 2^8 \equiv 2 \cdot 4 \cdot 1 \equiv 8 \pmod{17},$$

and $8 + 17n \equiv -5 \pmod{11}$ gives $n \equiv 7 \pmod{11}$ so that

$$172^{107} \equiv 8 + 17 \cdot 7 \equiv 127 \pmod{187}.$$

He concludes that Alice's message is 127.

Example 12.5.2. Bob decides on $p = 23$, $q = 29$, so that $m = 667$. $\varphi(m) = 22 \cdot 28 = 616$, and he chooses $e = 5$. Bob calculates d using a short extended Euclidean algorithm.

$$616 = 5 \cdot 123 + 1.$$

So $5 \cdot (-123) \pmod{616}$ and he takes $d = 616 - 123 = 493$.

Alice encrypts her word w using e and sends Bob $c = 168$.

Bob decodes using the Chinese remainder theorem and Fermat's little theorem.

$$168^{493} \equiv 7^9 \equiv 7 \cdot 7^8 \equiv 7 \cdot 12 \equiv 15 \pmod{23},$$
$$168^{493} \equiv (-6)^{17} \equiv -6 \cdot 6^{16} \equiv -6 \cdot 7 \equiv 16 \pmod{29};$$

$16 + 29n \equiv 15 \pmod{23}$ gives $6n \equiv -1 \pmod{23}$ and $n \equiv -4 \pmod{23}$ so that

$$168^{493} \equiv 16 + 29 \cdot (-4) \equiv -100 \equiv 567 \pmod{667}.$$

He concludes that Alice's message is 567.

The calculation Alice must have done is

$$567^5 \equiv (-100)^5 \equiv -10,000,000,000 \equiv 168 \pmod{667}.$$

12.6 The not-crazy-hard part of Shor's algorithm for cracking RSA

We have described why RSA is difficult to decrypt: doing so depends on factoring numbers which is very difficult. However, there is an algorithm due to Peter Shor making use of quantum computers (they don't exist yet) that would destroy the safe encryption that RSA attempts to guarantee.

Here's the important theorem.

Theorem 12.6.1. *Suppose that p and q are distinct primes and that $m = pq$. Suppose, in addition, that $\gcd(x, m) = 1$, that $ord_m(x) = 2k$ and that $x^k + 1 \not\equiv 0 \pmod{m}$. Then either*

$$\gcd(x^k - 1, m) = p \text{ or } \gcd(x^k - 1, m) = q.$$

Proof. Since $\mathrm{ord}_m(x) = 2k$ we know that $x^{2k} \equiv 1 \pmod{m}$ and so

$$x^{2k} \equiv 1 \pmod{p}$$
$$x^{2k} \equiv 1 \pmod{q}.$$

Thus,

$$(x^k - 1)(x^k + 1) \equiv 0 \pmod{p}$$
$$(x^k - 1)(x^k + 1) \equiv 0 \pmod{q},$$

and since $\mathbb{Z}/p$ and $\mathbb{Z}/q$ are fields we obtain

$$x^k \equiv \pm 1 \pmod{p}$$
$$x^k \equiv \pm 1 \pmod{q}.$$

We cannot have

$$x^k \equiv -1 \pmod{p}$$
$$x^k \equiv -1 \pmod{q}$$

since, in this case, the Chinese remainder theorem would give $x^k \equiv -1 \pmod{m}$, but we supposed that $x^k + 1 \not\equiv 0 \pmod{m}$.

If $x^k \equiv 1 \pmod{p}$ then we see that $p|(x^k - 1)$. If $x^k \equiv 1 \pmod{q}$ then we see that $q|(x^k - 1)$. Since $p, q|m$ we conclude that either $p|\gcd(x^k - 1, m)$ or $q|\gcd(x^k - 1, m)$. We cannot have $m|\gcd(x^k - 1, m)$ since this would tell us that $m|(x^k - 1)$ and thus $x^k \equiv 1 \pmod{m}$; this would contradict the fact that $\mathrm{ord}_m(x) = 2k$. □

Peter Shor's algorithm for factoring $m = pq$ is approximately as follows.

1. Pick $x \in \{2, 3, \ldots, m-2\}$ at random.

   If $\gcd(x, m) = p$ or q, then you have won. If $\gcd(x, m) = 1$, continue to step 2.

2. Calculate $\mathrm{ord}_m(x)$ and continue to step 3.

3. If $\mathrm{ord}_m(x)$ is odd, go back to step 1. If $\mathrm{ord}_m(x) = 2k$, continue to step 4.

4. If $x^k + 1 \equiv 0 \pmod{m}$, go back to step 1.

   If $x^k + 1 \not\equiv 0 \pmod{m}$, then $\gcd(x^k - 1, m) = p$ or q, and you have won.

Step 1 uses the Euclidean algorithm, step 2 requires a difficult quantum computer algorithm to run in a reasonable speed, step 3 uses nothing, and step 4 uses the Euclidean algorithm.

Example 12.6.2. Suppose we wish to factor 21 and the random number we pick is 2. Then

1. $\gcd(2, 21) = 1$.

2. $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, $2^4 \equiv -5 \pmod{21}$, $2^5 \equiv -10 \pmod{21}$, $2^6 \equiv 1 \pmod{21}$. So $\mathrm{ord}_{21}(2) = 6$.

3. $\mathrm{ord}_{21}(2) = 2 \cdot 3$.

4. $2^3 + 1 = 9 \not\equiv 0 \pmod{21}$. We find that

   $$\gcd(2^3 - 1, 21) = 7$$

   and factor $21 = 3 \cdot 7$.

12.7 Pollard's $p - 1$ algorithm

Another factoring algorithm that doesn't require a quantum computer is Pollard's $(p-1)$ algorithm. The reason it does not render RSA unsafe is that it is only effective when $(p - 1)$ has small prime factors.

Pollard's $(p - 1)$ algorithm makes use of Fermat's Little Theorem, in particular, the following corollary to Fermat's Little Theorem.

Corollary 12.7.1. *Suppose p is an odd prime factor of m and $(p-1)|B$. Then* $\gcd(2^B - 1, m) > 1$.

Proof. Since $2 \not\equiv 0 \pmod{p}$ Fermat's Little Theorem tells us that $2^{p-1} \equiv 1 \pmod{p}$ and because $(p - 1)|B$, we obtain

$$2^B \equiv 1 \pmod{p}.$$

This says that $p|(2^B - 1)$ and so $p| \gcd(2^B - 1, m)$. □

Definition 12.7.2. Let $k \in \mathbb{N}$ with $k \geq 2$. We say that a number m is k-smooth if every prime divisor of m is less than or equal to k.

Example 12.7.3.

$$\begin{aligned}
180 &= 2^2 \cdot 3^2 \cdot 5 \\
181 &= 181 \\
182 &= 2 \cdot 7 \cdot 13 \\
183 &= 3 \cdot 61 \\
184 &= 2^3 \cdot 23 \\
185 &= 5 \cdot 37 \\
186 &= 2 \cdot 3 \cdot 31 \\
187 &= 11 \cdot 17 \\
188 &= 2^2 \cdot 47 \\
189 &= 3^3 \cdot 7 \\
190 &= 2 \cdot 5 \cdot 19
\end{aligned}$$

None are 2-smooth or 3-smooth. 180 is 5-smooth. 180 and 189 are 7-smooth.

180, 182, 187, 189 and 190 are 19-smooth.

Definition 12.7.4. Let $e_{m,q} \in \mathbb{N} \cup \{0\}$ be defined by the property that $q^{e_{m,q}} \leq m < q^{e_{m,q}+1}$ and let

$$B_{m,k} = \prod_{q \text{ prime, } q \leq k} q^{e_{m,q}}.$$

In words, $B_{m,k}$ is the product of all prime powers q^e where $q \leq k$ and $q^e \leq m < q^{e+1}$.

Lemma 12.7.5. *If p is a prime factor of m and $p-1$ is k-smooth, then $(p-1)|B_{m,k}$.*

Proof. We *have* to make use of the fundamental theorem of arithmetic which is proved in section 14. Let q be a prime with $q|(p-1)$. Since $p-1$ is k-smooth we have $q \leq k$. Moreover,

$$p - 1 < m < q^{e_{m,q}+1}$$

so that if $q^e|(p-1)$, we have $e \leq e_{m,q}$ and so $q^e|B_{m,k}$. □

Theorem 12.7.6. *If p is an odd prime factor of m and $p-1$ is k-smooth. Then*

$$\gcd(2^{B_{m,k}} - 1, m) > 1.$$

Proof. The previous lemma says that $(p-1)|B_{m,k}$ and so the first lemma of the subsection says that $\gcd(2^{B_{m,k}} - 1, m) > 1$. □

Pollard's $(p-1)$ algorithm for finding a factor of m is as follows.

1. Pick some smoothness bound k.
2. Find $x \in \{0, 1, \ldots, m-1\}$ with $x \equiv 2^{B_{m,k}} \pmod{m}$.
3. Calculate $\gcd(x-1, m)$.

Step 2 is a quick calculation on a computer by writing the exponent in base 2. Step 3 is a quick calculation on a computer when one uses the Euclidean algorithm. Take a look at pages 215 and 216 of the textbook for examples.

13 Diffie-Hellman and El Gamal

If you like the previous cryptography material then you should look up the Diffie-Hellman and El Gamal schemes. They make use of the fact that calculating the discrete logarithm

$$\log_b : U(\mathbb{Z}/p) \longrightarrow \mathbb{Z}/(p-1)$$

is difficult. Here b must be an element with $\mathrm{ord}_p(b) = p - 1$ (the existence of such an element is part of the primitive element theorem) and I have mentioned this to some of you in office hours.

These are interesting topics, but I have decided that, since we spent a while experimenting with finite fields at the beginning of the class, it might be fun to see some applications of them instead.

14 Unique factorization

14.1 Irreducibles and primes

In order to talk about factoring natural numbers into primes we first have to define primes. In a commutative ring there are at least two sensible definitions one can make. The term irreducible is motivated by the fact that a natural number prime can not be factored any further except in the trivial way by using 1. The second is motivated by a property that natural number primes p have: if $p|ab$ then either $p|a$ or $p|b$ (Euclid's lemma).

Definition 14.1.1. Suppose R is a commutative ring and that $a \in R$ is non-zero and non-unital.

1. a is said to be *irreducible* if whenever $a = bc$ either b or c is a unit.
2. a is said to be *prime* if whenever $a|bc$ then either $a|b$ or $a|c$.

As long as we can cancel, primes are irreducible.

Proposition 14.1.2. *In an integral domain primes are irreducible.*

Proof. Let R be an integral domain, and suppose $a \in R$ is prime. We wish to show a is irreducible, so suppose that $a = bc$. We wish to show that either b or c is a unit.

Since $a = bc$, we trivially have $a|bc$ and so, because a is prime, either $a|b$ or $a|c$. Suppose without loss of generality that $a|b$. This means that there is a $d \in R$ with $b = ad$. Then $a = bc = (ad)c = a(dc)$ so that $a(1 - dc) = 0$. Since a is prime, $a \neq 0$. Since R is an integral domain, we deduce that $1 - dc = 0$, i.e. $dc = 1$, and so c is a unit. □

Lemma 14.1.3. *Suppose R is a commutative ring, that $a, b \in R$, that a is irreducible, and that a does not divide b. Then a and b are coprime.*

Proof. We have to show that 1 is a gcd of a, b.

1. That $1|a$ and $1|b$ is clear;
2. Suppose $c|a$ and $c|b$. The first division says that $a = cd$ for some $d \in R$. Since a is irreducible we deduce that either c is a unit or that d is a unit. If c is a unit we get $c|1$, what we want. If d is a unit, c is an associate of a, and, because $c|b$, this gives $a|b$, a contradiction.

□

Theorem 14.1.4. *Suppose that either $R = \mathbb{Z}$, or $R = F[x]$ for some field F, and that $a \in R$ is irreducible. Then a is prime.*

Proof. Suppose that $a|bc$ and that a does not divide b; we wish to show that $a|c$. By proposition 9.1.5, it is enough to show that a and b are coprime. The previous lemma finishes the proof. □

In all of the following examples primes and irreducibles coincide by the previous theorem.

Example 14.1.5. In $\mathbb{Z}$ the primes are $\pm 2, \pm 3, \pm 5, \pm 7, \pm 11, \pm 13, \pm 17, \ldots$

Example 14.1.6. In $\mathbb{Q}[x]$, $x - 1$, $x^2 - 2$, $x^2 + 1$, $x^2 + x + 1$ are some irreducibles.

Example 14.1.7. In $\mathbb{Z}/2[x]$ the irreducibles of degree less than or equal to 2 are x, $x+1$, x^2+x+1.

It was necessary to prove the previous two theorems carefully as the following remark shows.

Remark 14.1.8. There are strange integral domains where irreducibles are not necessarily primes. Here is an example. Let $R = \mathbb{Z}[\sqrt{-3}] = \{x + y\sqrt{-3} : x, y \in \mathbb{Z}\}$ and let

$$a = 2,\ b = 1 + \sqrt{-3},\ c = 1 - \sqrt{-3}.$$

We have $2|4 = bc$, and yet 2 does not divide b or c. Thus 2 is not prime. However, 2 is irreducible (exercise).

Remark 14.1.9. There are commutative rings where primes are not necessarily irreducible. $\mathbb{Z}/6$ is an example. 2 is prime. This because the multiples of 2 are $\{-2, 0, 2\}$ and the only way to express these as products is as follows.

$$\begin{aligned} -2 &= 1 \cdot (-2) = (-1) \cdot 2 = 2 \cdot 2 \\ 0 &= 0 \cdot 0 = 0 \cdot (\pm 1) = 0 \cdot (\pm 2) = 0 \cdot 3 = (\pm 2) \cdot 3 \\ 2 &= 1 \cdot 2 = (-1) \cdot (-2) = 2 \cdot (-2) \end{aligned}$$

This also shows 2 is not irreducible, since $2 = 2 \cdot (-2)$. (Shout out, Sitara.)

14.2 Euclidean domains are unique factorization domains

We finally come to factorizing elements into their prime decomposition.

Definition 14.2.1. Let R be an integral domain. We say that R is a *unique factorization domain* if the following two conditions hold.

1. Whenever $a \in R$ is non-zero and non-unital there exist irreducible elements $b_1, \ldots, b_n \in R$ such that
$$a = b_1 \cdots b_n.$$
2. If $b_1, \ldots b_n, c_1, \ldots c_m \in R$ are irreducible with
$$b_1 \cdots b_n = c_1 \cdots c_m$$
then $n = m$ and there exists a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that b_k and $c_{\sigma(k)}$ are associates for each $k \in \{1, \ldots, n\}$.

Theorem 14.2.2. *$\mathbb{Z}$ is a unique factorization domain.*

Proof. First, we verify existence of factorizations: let $n \in \mathbb{Z}$ be non-zero and non-unital; we must show that n can be written as a finite product of irreducibles. We induct on $|n|$. The cases $|n| = 0$ and $|n| = 1$ are not relevant since we took n to be non-zero and non-unital. The base case is when $|n| = 2$, which follows since ± 2 are irreducible, so suppose $|n| > 2$. If n is irreducible we are done. If n is not irreducible, then we can write $n = lm$ where neither l or m is a unit; this means that $|l|, |m| > 1$, and so $|l|, |m| < |n|$, and the result follows by induction.

We now turn to showing the uniqueness of such a factorization. Let $b_1, \ldots b_n, c_1, \ldots c_m \in \mathbb{Z}$ be irreducibles with

$$b_1 \cdots b_n = c_1 \cdots c_m.$$

Since b_1 is irreducible, it is prime (theorem 14.1.4). Because $b_1|b_1 \cdots b_n = c_1 \cdots c_m$ we see that $b_1|c_j$ for some $j \in \{1, \ldots, m\}$. By reordering we can assume $b_1|c_1$. So $c_1 = b_1 u$ for some $u \in \mathbb{Z}$, which must necessarily be a unit as c_1 is irreducible and b_1 is not a unit. We can now cancel b_1 to obtain

$$b_2 \cdots b_n = \pm c_2 \cdots c_m$$

and continuing in this way gives the result. □

Theorem 14.2.3. *Let F be a field. Then $F[x]$ is a unique factorization domain.*

Proof. First, we verify the existence of factorizations: let $f(x) \in F[x]$ be non-zero and non-unital; we must show that $f(x)$ can be written as a finite product of irreducibles. We induct on $\deg f(x)$. The case $\deg f(x) = 0$ is not relevant since we took $f(x)$ to be non-zero and non-unital. The base case is when $\deg f(x) = 1$, in which case $f(x)$ is irreducible, so suppose $\deg f(x) > 1$. If $f(x)$ is irreducible we are done. If $f(x)$ is not irreducible, then we can write $f(x) = g(x)h(x)$ where neither $g(x)$ or $h(x)$ is a unit; this means that $\deg g(x), \deg h(x) > 0$, and so $\deg g(x), \deg h(x) < \deg f(x)$, and the result follows by induction.

The proof of uniqueness is the same as for $\mathbb{Z}$. The main point is that it also uses theorem 14.1.4 to say irreducibles are prime. □

Example 14.2.4. If one examines the previous proofs, one sees we used the Euclidean functions for $\mathbb{Z}$ and $F[x]$ to justify the existence of a factorization. The uniqueness used the fact that irreducibles are prime, which depended on the existence of gcds and Bezout's identity. One can imagine a place where Bezout's identity works so that factorizations are unique if they exist, but where it is possible that factorizations do not exist. Consider

$$\left\{ f(x) = \sum_{n=0}^{\infty} a_n x^n : f(x) \text{ has infinite radius of convergence} \right\}.$$

This is such a ring. Since $\exp(x)$ can be factored more and more

$$\exp(x) = \exp\left(\frac{x}{2}\right) \cdot \exp\left(\frac{x}{2}\right) = \exp\left(\frac{x}{2}\right) \cdot \exp\left(\frac{x}{4}\right) \cdot \exp\left(\frac{x}{4}\right) = \cdots$$

factorizations do not necessarily exist. However, (I have not checked this) it is a Bezout domain.

14.3 The Gaussian integers $\mathbb{Z}[i]$

Consider the following subset of the complex numbers

$$\mathbb{Z}[i] = \{x + yi : x, y \in \mathbb{Z}\}.$$

The addition and multiplication of complex numbers make $\mathbb{Z}[i]$ into a commutative ring. Moreover, $\mathbb{Z}[i]$ is an integral domain.

Definition 14.3.1. Elements of the commutative ring $\mathbb{Z}[i]$ are called *Gaussian integers*. We define

$$d : \mathbb{Z}[i] \longrightarrow \mathbb{N} \cup \{0\}$$

by $d(a) = a\overline{a} = |a|^2$.

Lemma 14.3.2. *For $a, b \in \mathbb{Z}[i]$, $d(ab)$=$d(a)d(b)$.*

Proof. This follows from the facts that $\mathbb{C}$ is a *commutative* ring and that $\overline{ab} = \overline{a}\overline{b}$. □

Lemma 14.3.3. *The units in $\mathbb{Z}[i]$ are $1, -1, i, -i$.*

Proof. Suppose that $ab = 1$. Then $d(a)d(b) = d(ab) = d(1) = 1$. Thus $d(a) = 1$ and

$$a = 1, -1, i, \text{ or } -i.$$

□

Theorem 14.3.4. *The Gaussian integers form a Euclidean domain.*

Proof. We just have to show that if $a, b \in \mathbb{Z}[i]$ with $b \neq 0$, then there exist $q, r \in \mathbb{Z}[i]$ with $a = qb + r$ and $d(r) < d(b)$.

Let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Since $a, b \in \mathbb{C}$ we can consider $\frac{a}{b} \in \mathbb{C}$. Choose $q \in \mathbb{Z}[i]$ with

$$\left|\frac{a}{b} - q\right|^2 \leq \frac{1}{2}$$

and let $r = a - qb$. Then it is automatic that $a = qb + r$. Moreover,

$$d(r) = |r|^2 = |a - qb|^2 = \left|\frac{a}{b} \cdot b - q \cdot b\right|^2 = \left|\frac{a}{b} - q\right|^2 \cdot |b|^2 = \left|\frac{a}{b} - q\right|^2 \cdot d(b) \leq \frac{d(b)}{2} < d(b).$$

□

Because $\mathbb{Z}[i]$ is a Euclidean domain, there is a Euclidean algorithm, gcd's make sense, Bezout's theorem holds, and proposition 9.1.5 (Euclid's lemma) is true. Moreover, theorem 14.1.4 holds for $\mathbb{Z}[i]$ too: that is, irreducibles are primes.

Definition 14.3.5. If an element $a \in \mathbb{Z}[i]$ is prime (equivalently irreducible) then we say that a is a *Gaussian prime*.

Example 14.3.6. 2 is prime when considered as an element of $\mathbb{Z}$. We say it is a *rational prime*. It is not a Gaussian prime, because it is not irreducible:

$$2 = (1 - i)(1 + i) = -i(1 + i)^2 = i(1 - i)^2.$$

Example 14.3.7. It turns out that the rational primes 3, 7, 11, 19, 23, 31, are all Gaussian primes. What is so special about them? You'll see on the homework.

Example 14.3.8. The rational primes 5, 13, 17, 29, 37, 41 are not Gaussian primes since

$$\begin{aligned} 5 &= 1^2 + 2^2 = (1-2i)(1+2i) \\ 13 &= 2^2 + 3^2 = (2-3i)(2+3i) \\ 17 &= 1^2 + 4^2 = (1-4i)(1+4i) \\ 29 &= 2^2 + 5^2 = (2-5i)(2+5i) \\ 37 &= 1^2 + 6^2 = (1-6i)(1+6i) \\ 41 &= 4^2 + 5^2 = (4-5i)(4+5i) \end{aligned}$$

Theorem 14.3.9. *Suppose $a \in \mathbb{Z}[i]$ and that $d(a)$ is a prime in $\mathbb{Z}$. Then a is a Gaussian prime.*

Proof. Suppose that $a = bc$ so $d(a) = d(bc) = d(b)d(c)$. Because $d(a)$ is a prime in $\mathbb{Z}$, it is non-zero and non-unital and either $d(b) = 1$ or $d(c) = 1$. This means that a is non-zero and non-unital and either b is a unit or c is a unit. Thus, a is irreducible. □

Example 14.3.10. The factors appearing above are Gaussian primes. That is,

$$1 \pm 2i,\ 2 \pm 3i,\ 1 \pm 4i,\ 2 \pm 5i,\ 1 \pm 6i,\ 4 \pm 5i$$

are Gaussian primes.

Theorem 14.3.11. *$\mathbb{Z}[i]$ is a unique factorization domain.*

Proof. The same as for $\mathbb{Z}$ inducting on $|x|^2$ instead of $|x|$ (which is not always an integer). □

Example 14.3.12. Suppose we wish to factor $a = 3 + 21i$. Here's a trick for hunting down the prime factors. Notice that

$$a\overline{a} = 3^2 + 21^2 = 450 = 2 \cdot 3^2 \cdot 5^2 = i(1-i)^2 \cdot 3^2 \cdot (1-2i)^2(1+2i)^2.$$

This tells us that the possible prime factors of a are $(1-i)$, 3, $(1-2i)$ and $1+2i$. We see that

$$a = 3(1+7i).$$

Then we spot (by thinking about arguments, maybe) that $1 + 7i = (1-i)(1+2i)^2$ so that

$$a = 3(1-i)(1+2i)^2.$$

15 BCH codes

Suppose we are trying to transmit messages consisting of zeros and ones, but, due to noise, it is likely that there will be errors in our transmission. However, suppose we know that with very high probability there will be at most t errors in our transmitted message. Can we think of a way to send enough data that our original message can be recovered, regardless of whether there are 0, 1, $\ldots$, $t-1$, or t errors? BCH codes answer this question with a big fat "yes." Moreover, the "big fat yes" does something more efficient than writing out our message $(2t+1)$ times.

15.1 Sending four bits and accounting for up to one error

Suppose we are trying to transmit messages consisting of four bits (four numbers which are either zero or one), but, due to noise, it is likely that there will be errors in our transmission. However, suppose we know that there will be at most one error in our transmitted message. Using seven bits, we can guarantee that our original message can be recovered, irrespective of whether there are no errors or there is one error. There are simple ways to do this, but here is one which will generalize.

$$\begin{array}{rcl} \alpha^0 &=& 1 \\ \alpha^1 &=& \alpha \\ \alpha^2 &=& \alpha^2 \\ \alpha^3 &=& \alpha+1 \\ \alpha^4 &=& \alpha^2+\alpha \\ \alpha^5 &=& \alpha^2+\alpha+1 \\ \alpha^6 &=& \alpha^2+1 \end{array} \qquad \begin{array}{rcl} 1 &=& \alpha^0 \\ \alpha &=& \alpha^1 \\ \alpha+1 &=& \alpha^3 \\ \alpha^2 &=& \alpha^2 \\ \alpha^2+1 &=& \alpha^6 \\ \alpha^2+\alpha &=& \alpha^4 \\ \alpha^2+\alpha+1 &=& \alpha^5 \end{array}$$

Recall that $m(x) = x^3 + x + 1 \in \mathbb{F}_2[x]$ is irreducible, so that

$$\mathbb{F}_8 = \mathbb{F}_2[x]/(x^3+x+1)$$

is a field. Label $[x]_{x^3+x+1}$ by α. Then (homework)

$$\mathbb{F}_8 = \{0, 1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6\}.$$

We can go between this description of elements and the usual description using the exp and log-table displayed above.

THE PROCEDURE

1. Suppose we have a word of length 4, $\mathbf{w} = (w_6, w_5, w_4, w_3)$. Form the polynomial

$$W(x) = w_6x^6 + w_5x^5 + w_4x^4 + w_3x^3 \in \mathbb{F}_2[x].$$

2. Using the division algorithm we can write

$$W(x) = u(x)m(x) + V(x)$$

for some polynomials $u(x)$, $V(x)$ where $\deg V(x) < 3$. Let $C(x) = W(x) + V(x)$.

3. Writing $C(x) = c_6x^6 + c_5x^5 + \ldots + c_1x + c_0$ the code for our word is the length 7 vector

$$\mathbf{c} = (c_6, c_5, \ldots, c_1, c_0).$$

Notice that since $\deg V(x) < 3$, we have

$$\mathbf{w} = (w_6, w_5, w_4, w_3) = (c_6, c_5, c_4, c_3). \tag{15.1.1}$$

4. We send $\mathbf{c}$ and the vector which is received is

$$\mathbf{r} = (r_6, r_5, \ldots, r_1, r_0).$$

We have to describe how to reconstruct $\mathbf{w}$ from $\mathbf{r}$. In light of (15.1.1) it is enough to reconstruct $\mathbf{c}$ from $\mathbf{r}$. Here's what we do. Let $R(x) = r_6x^6 + r_5x^5 + \ldots + r_1x + r_0 \in \mathbb{F}_8[x]$.

(a) If $R(\alpha) = 0$, then there are no errors, we have $\mathbf{c} = \mathbf{r}$ and $\mathbf{w} = (r_6, r_5, r_4, r_3)$.

(b) If $R(\alpha) \neq 0$, then there is one error. Finding $e \in \{0, 1, 2, 3, 4, 5, 6\}$ such that $R(\alpha) = \alpha^e$ tells us the error was at r_e.

Remark 15.1.2. Why does this work? Well, the first thing to notice is that when we view $m(x)$ as an element of $\mathbb{F}_8[x]$ we have $m(\alpha) = 0$. In step 2, above, we wrote

$$W(x) = u(x)m(x) + V(x)$$

and then set $C(x) = W(x) + V(x)$. So $C(x) = u(x)m(x)$ and since α is a root of $m(x)$, it is also a root of $C(x)$.

The error between the recieved vector and the coded word is stored by the polynomial

$$E(x) = C(x) + R(x).$$

Either there are no errors and $E(x) = 0$, or there is one error at position e and $E(x) = x^e$. In the first case, $R(\alpha) = E(\alpha) = 0$. In the second case, $R(\alpha) = E(\alpha) = \alpha^e$.

Example 15.1.3. Suppose we wish to encode $\mathbf{w} = (1, 0, 1, 0)$. We form the polynomial

$$W(x) = x^6 + x^4.$$

We write $W(x) = u(x)m(x) + V(x)$ where $u(x) = x^3 + 1$ and $V(x) = x + 1$. We let

$$C(x) = W(x) + V(x) = x^6 + x^4 + x + 1.$$

The code for our word $\mathbf{w}$ is $\mathbf{c} = (1, 0, 1, 0, 0, 1, 1)$. Suppose that we send $\mathbf{c}$ and the vector which is received is $\mathbf{r} = (1, 1, 1, 0, 0, 1, 1)$. Then $R(x) = x^6 + x^5 + x^4 + x + 1$ and

$$R(\alpha) = \alpha^6 + \alpha^5 + \alpha^4 + \alpha + 1 = (\alpha^2 + 1) + (\alpha^2 + \alpha + 1) + (\alpha^2 + \alpha) + \alpha + 1 = \alpha^2 + \alpha + 1 = \alpha^5.$$

The receiver would conclude that there was one error occuring in the fifth position: this is correct!

$$
\begin{array}{rcl}
\alpha^0 &=& 1 \\
\alpha^1 &=& \alpha \\
\alpha^2 &=& \alpha^2 \\
\alpha^3 &=& \alpha^3 \\
\alpha^4 &=& \alpha + 1 \\
\alpha^5 &=& \alpha^2 + \alpha \\
\alpha^6 &=& \alpha^3 + \alpha^2 \\
\alpha^7 &=& \alpha^3 + \alpha + 1 \\
\alpha^8 &=& \alpha^2 + 1 \\
\alpha^9 &=& \alpha^3 + \alpha \\
\alpha^{10} &=& \alpha^2 + \alpha + 1 \\
\alpha^{11} &=& \alpha^3 + \alpha^2 + \alpha \\
\alpha^{12} &=& \alpha^3 + \alpha^2 + \alpha + 1 \\
\alpha^{13} &=& \alpha^3 + \alpha^2 + 1 \\
\alpha^{14} &=& \alpha^3 + 1
\end{array}
\qquad
\begin{array}{rcl}
1 &=& \alpha^0 \\
\alpha &=& \alpha^1 \\
\alpha + 1 &=& \alpha^4 \\
\alpha^2 &=& \alpha^2 \\
\alpha^2 + 1 &=& \alpha^8 \\
\alpha^2 + \alpha &=& \alpha^5 \\
\alpha^2 + \alpha + 1 &=& \alpha^{10} \\
\alpha^3 &=& \alpha^3 \\
\alpha^3 + 1 &=& \alpha^{14} \\
\alpha^3 + \alpha &=& \alpha^9 \\
\alpha^3 + \alpha + 1 &=& \alpha^7 \\
\alpha^3 + \alpha^2 &=& \alpha^6 \\
\alpha^3 + \alpha^2 + 1 &=& \alpha^{13} \\
\alpha^3 + \alpha^2 + \alpha &=& \alpha^{11} \\
\alpha^3 + \alpha^2 + \alpha + 1 &=& \alpha^{12}
\end{array}
$$

15.2 Sending seven bits and accounting for up to two errors

Suppose we are trying to transmit messages consisting of seven bits accounting for up to two errors in our transmitted message.

Recall that $x^4 + x + 1 \in \mathbb{F}_2[x]$ is irreducible, so that

$$\mathbb{F}_{16} = \mathbb{F}_2[x]/(x^4 + x + 1)$$

is a field. Label $[x]_{x^4+x+1}$ by α. Then

$$\mathbb{F}_{16} = \{0, 1, \alpha, \alpha^2, \ldots, \alpha^{14}\}.$$

We can go between this description of elements and the usual description using the exp and log-table displayed above. Let

$$m(x) = x^8 + x^7 + x^6 + x^4 + 1.$$

THE PROCEDURE

1. Suppose we have a word of length 7, $\mathbf{w} = (w_{14}, w_{13}, w_{12}, w_{11}, w_{10}, w_9, w_8)$. Form the polynomial
$$W(x) = w_{14}x^{14} + w_{13}x^{13} + \ldots + w_9x^9 + w_8x^8 \in \mathbb{F}_2[x].$$
2. Using the division algorithm we can write
$$W(x) = u(x)m(x) + V(x)$$
for some polynomials $u(x)$, $V(x)$ where $\deg V(x) < 8$. Let $C(x) = W(x) + V(x)$.
3. Writing $C(x) = c_{14}x^{14} + c_{13}x^{13} + \ldots + c_1x + c_0$ the code for our word is the length 15 vector
$$\mathbf{c} = (c_{14}, c_{13}, \ldots, c_1, c_0).$$
Notice that since $\deg V(x) < 8$, we have
$$\mathbf{w} = (w_{14}, w_{13}, w_{12}, w_{11}, w_{10}, w_9, w_8) = (c_{14}, c_{13}, c_{12}, c_{11}, c_{10}, c_9, c_8). \tag{15.2.1}$$

4. We send $\mathbf{c}$ and the vector which is received is

$$\mathbf{r} = (r_{14}, r_{13}, \ldots, r_1, r_0).$$

We have to describe how to reconstruct $\mathbf{w}$ from $\mathbf{r}$. In light of (15.2.1) it is enough to reconstruct $\mathbf{c}$ from $\mathbf{r}$. Here's what we do.

Let $R(x) = r_{14}x^{14} + r_{13}x^{13} + \ldots + r_1x + r_0 \in \mathbb{F}_{16}[x]$ and consider the matrix

$$\mathbf{S} = \begin{pmatrix} R(\alpha) & R(\alpha^2) \\ R(\alpha^2) & R(\alpha^3) \end{pmatrix}.$$

(a) If $\mathbf{S} = \mathbf{0}$, then there are no errors, we have $\mathbf{c} = \mathbf{r}$ and $\mathbf{w} = (r_{14}, r_{13}, r_{12}, r_{11}, r_{10}, r_9, r_8)$.

(b) If $\mathbf{S} \neq 0$, but $\det(\mathbf{S}) = 0$, then there is one error. It will turn out that $R(\alpha) \neq 0$ and finding $e \in \{0, 1, 2, 3, 4, 5, 6\}$ such that $R(\alpha) = \alpha^e$ tells us the error was at r_e.

(c) If $\det(\mathbf{S}) \neq 0$, then let

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \end{pmatrix} = \begin{pmatrix} R(\alpha) & R(\alpha^2) \\ R(\alpha^2) & R(\alpha^3) \end{pmatrix}^{-1} \begin{pmatrix} R(\alpha^3) \\ R(\alpha^4) \end{pmatrix}.$$

Find the roots of $\Sigma(x) = x^2 + \sigma_1 x + \sigma_0 \in \mathbb{F}_{16}[x]$. They will be of the form α^{e_1} and α^{e_2} where $e_1, e_2 \in \{0, 1, \ldots, 13, 14\}$. e_1 and e_2 are where the errors occur.

Remark 15.2.2. Why does this work? Well, the first thing to notice is that when we view $m(x)$ as an element of $\mathbb{F}_{16}[x]$ we have $m(\alpha) = m(\alpha^2) = m(\alpha^3) = m(\alpha^4) = 0$. In step 2, above, we wrote

$$W(x) = u(x)m(x) + V(x)$$

and then set $C(x) = W(x) + V(x)$. So $C(x) = u(x)m(x)$ and since $\alpha,\ \alpha^2,\ \alpha^3$, and α^4 are roots of $m(x)$, they are also roots of $C(x)$.

The error between the received vector and the coded word is stored by the polynomial

$$E(x) = C(x) + R(x).$$

Either there are no errors and $E(x) = 0$, or there is one error at position e and $E(x) = x^e$, or there are two errors at positions e_1 and e_2 and $E(x) = x^{e_1} + x^{e_2}$.

In the first case, we have for $i = 1, 2, 3$, $R(\alpha^i) = E(\alpha^i) = 0$, so that $\mathbf{S} = \mathbf{0}$.

In the second case, we have

$$\mathbf{S} = \begin{pmatrix} \alpha^e & \alpha^{2e} \\ \alpha^{2e} & \alpha^{3e} \end{pmatrix} \neq \mathbf{0}.$$

The second row is α times the first row, so $\det(\mathbf{S}) = 0$, and $R(\alpha) = \alpha^e$.

In the third case, one can check that

$$\mathbf{S} = \begin{pmatrix} 1 & 1 \\ \alpha^{e_1} & \alpha^{e_2} \end{pmatrix} \begin{pmatrix} \alpha^{e_1} & 0 \\ 0 & \alpha^{e_2} \end{pmatrix} \begin{pmatrix} 1 & \alpha^{e_1} \\ 1 & \alpha^{e_2} \end{pmatrix},$$

the product of three invertible matrices. Thus $\det(\mathbf{S}) \neq 0$. We will check that $\Sigma(\alpha^{e_1}) = \Sigma(\alpha^{e_2}) = 0$ in more generality a little later.

Example 15.2.3. Suppose we wish to encode $\mathbf{w} = (1, 1, 1, 0, 1, 0, 1)$. We form the polynomial

$$W(x) = x^{14} + x^{13} + x^{12} + x^{10} + x^8.$$

We write $W(x) = u(x)m(x) + V(x)$ where $u(x) = x^6 + 1$ and $V(x) = x^7 + x^4 + 1$. We let

$$C(x) = W(x) + V(x) = x^{14} + x^{13} + x^{12} + x^{10} + x^8 + x^7 + x^4 + 1.$$

The code for our word $\mathbf{w}$ is $\mathbf{c} = (1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1)$. Suppose that we send $\mathbf{c}$ and the vector which is received is $\mathbf{r} = (1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1)$. Then

$$R(x) = x^{14} + x^{12} + x^{10} + x^8 + x^7 + x^4 + x^3 + 1.$$

We calculate

$$\begin{aligned} R(\alpha) &= \alpha^{14} + \alpha^{12} + \alpha^{10} + \alpha^8 + \alpha^7 + \alpha^4 + \alpha^3 + 1 \\ &= (\alpha^3 + 1) + (\alpha^3 + \alpha^2 + \alpha + 1) + (\alpha^2 + \alpha + 1) + (\alpha^2 + 1) + (\alpha^3 + \alpha + 1) + (\alpha + 1) + \alpha^3 + 1 \\ &= \alpha^2 + 1 = \alpha^8 \end{aligned}$$

Thus, using the mod 2 Freshman dream we have

$$R(\alpha^2) = R(\alpha)^2 = \alpha^{16} = \alpha, \text{ and } R(\alpha^4) = R(\alpha^2)^2 = \alpha^2.$$

Finally,

$$\begin{aligned} R(\alpha^3) &= \alpha^{42} + \alpha^{36} + \alpha^{30} + \alpha^{24} + \alpha^{21} + \alpha^{12} + \alpha^9 + 1 \\ &= \alpha^{12} + \alpha^6 + 1 + \alpha^9 + \alpha^6 + \alpha^{12} + \alpha^9 + 1 = 0, \end{aligned}$$

and so

$$\mathbf{S} = \begin{pmatrix} \alpha^8 & \alpha \\ \alpha & 0 \end{pmatrix}.$$

$\mathbf{S} \neq 0$ and $\det(\mathbf{S}) = \alpha^2 \neq 0$. We have

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \end{pmatrix} = \begin{pmatrix} \alpha^8 & \alpha \\ \alpha & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \alpha^2 \end{pmatrix} = \alpha^{-2} \begin{pmatrix} 0 & \alpha \\ \alpha & \alpha^8 \end{pmatrix} \begin{pmatrix} 0 \\ \alpha^2 \end{pmatrix} = \begin{pmatrix} 0 & \alpha \\ \alpha & \alpha^8 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha^8 \end{pmatrix}$$

so that $\Sigma(x) = x^2 + \alpha^8 x + \alpha$. We find that

$$\begin{aligned} \Sigma(\alpha^3) &= \alpha^6 + \alpha^8\alpha^3 + \alpha = \alpha^6 + \alpha^{11} + \alpha = (\alpha^3 + \alpha^2) + (\alpha^3 + \alpha^2 + \alpha) + \alpha = 0 \\ \Sigma(\alpha^{13}) &= \alpha^{26} + \alpha^{11}\alpha^{13} + \alpha = \alpha^{11} + \alpha^9 + \alpha = (\alpha^3 + \alpha^2 + \alpha) + (\alpha^3 + \alpha^2) + \alpha = 0 \end{aligned}$$

and so the receiver would conclude that there were two errors occuring in the third and thirteenth position: this is correct!

15.3 General BCH codes

Let $q = 2^r$, let $\mathbb{F} = \mathbb{F}_q$ be the field with q elements, and let α be a primitive (see section 16) element in $\mathbb{F}^\times$, so that

$$\mathbb{F} = \{0, 1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}.$$

Let $m(x)$ be the polynomial of smallest degree in $\mathbb{F}_2[x]$ with

$$\alpha, \alpha^2, \ldots, \alpha^{2t}$$

as roots. Let $d = \deg m(x)$. We hope that $d < q - 1$. Suppose it is and let $l = q - 1 - d > 0$. We will be able to transmit words of length l. Here's how...

THE PROCEDURE

1. Suppose we have a word of length l, $\mathbf{w} = (w_{q-2}, w_{q-3}, \ldots, w_{d+1}, w_d)$. Form the polynomial

$$W(x) = w_{q-2}x^{q-2} + w_{q-3}x^{q-3} + \ldots + w_{d+1}x^{d+1} + w_d x^d.$$

2. Using the division algorithm we can write

$$W(x) = u(x)m(x) + V(x)$$

for some polynomials $u(x)$, $V(x)$ where $\deg V(x) < d$. Let $C(x) = W(x) + V(x)$.

3. Writing $C(x) = c_{q-2}x^{q-2} + c_{q-3}x^{q-3} + \ldots + c_1 x + c_0$ the code for our word is the vector

$$\mathbf{c} = (c_{q-2}, c_{q-3}, \ldots, c_1, c_0).$$

Notice that since $\deg V(x) < d$, we have

$$\mathbf{w} = (w_{q-2}, w_{q-3}, \ldots, w_{d+1}, w_d) = (c_{q-2}, c_{q-3}, \ldots, c_{d+1}, c_d). \tag{15.3.1}$$

4. We send $\mathbf{c}$ and the vector which is received is

$$\mathbf{r} = (r_{q-2}, r_{q-3}, \ldots, r_1, r_0).$$

We have to describe how to reconstruct $\mathbf{w}$ from $\mathbf{r}$. In light of (15.3.1) it is enough to reconstruct $\mathbf{c}$ from $\mathbf{r}$.

THINGS WE KNOW

1. In step 2, above, we wrote $W(x) = u(x)m(x) + V(x)$ and then set $C(x) = W(x) + V(x)$. So

$$C(x) = u(x)m(x).$$

   Since $\alpha, \alpha^2, \ldots, \alpha^{2t}$ are roots of $m(x)$, they are also roots of $C(x)$.

2. If we let $R(x) = r_{q-2}x^{q-2} + r_{q-3}x^{q-3} + \ldots + r_1x + r_0$, then the error between the received vector and the coded word is stored by the polynomial

$$E(x) = C(x) + R(x).$$

   We can write

$$E(x) = x^{e_1} + x^{e_2} + \ldots + x^{e_r}$$

   where $e_1, \ldots, e_r$ are distinct numbers less than $q-1$; they are the locations of the errors. We (with high probability) assumed that there were going to be at most t errors. So $r \leq t$.

WHAT WE HAVE TO FIGURE OUT

1. The number of errors r.
2. The locations of the errors $e_1, \ldots, e_r$.

15.4 Determining r, the number of errors

Theorem 15.4.1. *Let*

$$S_1 = R(\alpha),\ S_2 = R(\alpha^2),\ \ldots,\ S_{2t-1} = R(\alpha^{2t-1}),\ S_{2t} = R(\alpha^{2t}).$$

Then the number of errors r is given by the rank of the following matrix.

$$\boldsymbol{S} = \begin{pmatrix} S_1 & S_2 & \ldots & S_t \\ S_2 & S_3 & \ldots & S_{t+1} \\ \vdots & & & \vdots \\ S_t & S_{t+1} & \ldots & S_{2t-1} \end{pmatrix}$$

Moreover, in this case

$$\boldsymbol{U}_r = \begin{pmatrix} S_1 & S_2 & \ldots & S_r \\ S_2 & S_3 & \ldots & S_{r+1} \\ \vdots & & & \vdots \\ S_r & S_{r+1} & \ldots & S_{2r-1} \end{pmatrix}$$

is invertible.

Proof. First we note that, since $C(\alpha) = C(\alpha^2) = \ldots = C(\alpha^{2t}) = 0$, we have, for $j = 1, \ldots, 2t$,

$$S_j = R(\alpha^j) = E(\alpha^j) = \sum_{k=1}^{r} \alpha^{je_k}.$$

Define vectors $\mathbf{x}_1, \ldots, \mathbf{x}_r \in \mathbb{F}^t$ by

$$\mathbf{x}_k = (\alpha^{e_k}, \alpha^{2e_k}, \alpha^{3e_k}, \ldots, \alpha^{te_k}).$$

Then, for $j = 1, \ldots, t$, we have

$$\alpha^{(j-1)e_k}\mathbf{x}_k = (\alpha^{je_k}, \alpha^{(j+1)e_k}, \alpha^{(j+2)e_k}, \ldots, \alpha^{(j+t-1)e_k}),$$

and thus,

$$\sum_{k=1}^{r} \alpha^{(j-1)e_k}\mathbf{x}_k = (S_j, S_{j+1}, S_{j+2}, \ldots, S_{j+(t-1)}).$$

This says that $\mathbf{x}_1, \ldots, \mathbf{x}_r$ span the row space of $\mathbf{S}$ and so the rank of $\mathbf{S}$ is less than or equal to r. To deduce the rank is equal to r it is enough to show $\mathbf{U}_r$ is invertible so we turn to this.

Define vectors $\mathbf{y}_1, \ldots, \mathbf{y}_r \in \mathbb{F}^r$ by

$$\mathbf{y}_j = (\alpha^{(j-1)e_1}, \alpha^{(j-1)e_2}, \alpha^{(j-1)e_3}, \ldots, \alpha^{(j-1)e_r}).$$

Then

$$\mathbf{y}_i \ \mathrm{diag}(\alpha^{e_1}, \alpha^{e_2}, \alpha^{e_3}, \ldots, \alpha^{e_r}) \ \mathbf{y}_j^T = \sum_{k=1}^{r} \alpha^{(i-1)e_k}\alpha^{e_k}\alpha^{(j-1)e_k} = \sum_{k=1}^{r} \alpha^{(i+j-1)e_k} = S_{i+j-1}.$$

Thus, if $\mathbf{A}$ is the $r \times r$ matrix whose j-th row is $\mathbf{y}_j$, then

$$\mathbf{A} \ \mathrm{diag}(\alpha^{e_1}, \alpha^{e_2}, \alpha^{e_3}, \ldots, \alpha^{e_r}) \ \mathbf{A}^T = \mathbf{U}_r.$$

This gives

$$\det(\mathbf{U}_r) = \det(\mathbf{A})^2 \cdot \prod_{k=1}^{r} \alpha^{e_k} = \pm \prod_{\substack{i \neq j, \\ 1 \leq i,j \leq r}} (\alpha^{e_i} - \alpha^{e_j}) \ \cdot \ \prod_{k=1}^{r} \alpha^{e_k} \neq 0$$

where we have made use of a formula for the determinant of a Vandermonde matrix. □

15.5 Determining the error locations

Theorem 15.5.1. *Suppose there are r errors. As in the previous theorem let*

$$S_1 = R(\alpha),\ S_2 = R(\alpha^2),\ \ldots,\ S_{2r-1} = R(\alpha^{2r-1}),\ S_{2r} = R(\alpha^{2r}).$$

Define $\sigma_0, \sigma_1, \ldots, \sigma_{r-1} \in \mathbb{F}$ by

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \\ \vdots \\ \sigma_{r-1} \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & \ldots & S_r \\ S_2 & S_3 & \ldots & S_{r+1} \\ \vdots & & & \vdots \\ S_r & S_{r+1} & \ldots & S_{2r-1} \end{pmatrix}^{-1} \begin{pmatrix} S_{r+1} \\ S_{r+2} \\ \vdots \\ S_{2r} \end{pmatrix}.$$

Then the polynomial

$$\Sigma(x) = x^r + \sigma_{r-1}x^{r-1} + \sigma_{r-2}x^{r-2} + \ldots + \sigma_1 x + \sigma_0 \in \mathbb{F}[x]$$

factors as

$$(x + \alpha^{e_1})(x + \alpha^{e_2}) \cdots (x + \alpha^{e_r}).$$

Proof. Let

$$\tau(x) = (x + \alpha^{e_1})(x + \alpha^{e_2}) \cdots (x + \alpha^{e_r}) = x^r + \tau_{r-1}x^{r-1} + \tau_{r-2}x^{r-2} + \ldots + \tau_1 x + \tau_0.$$

Plugging in α^{e_k} gives

$$0 = \alpha^{re_k} + \tau_{r-1}\alpha^{(r-1)e_k} + \ldots + \tau_1\alpha^{e_k} + \tau_0.$$

Multiplying by α^{je_k} gives

$$0 = \alpha^{(r+j)e_k} + \tau_{r-1}\alpha^{(r+j-1)e_k} + \ldots + \tau_1\alpha^{(j+1)e_k} + \tau_0\alpha^{je_k}.$$

Summing from $k = 1$ to r gives

$$0 = S_{r+j} + \tau_{r-1}S_{r+j-1} + \ldots + \tau_1 S_{j+1} + \tau_0 S_j$$

so that, for $j = 1, \ldots, r$,

$$S_{r+j} = \tau_{r-1}S_{r+j-1} + \ldots + \tau_1 S_{j+1} + \tau_0 S_j.$$

Writing these equations in matrix form gives

$$\begin{pmatrix} S_1 & S_2 & \ldots & S_r \\ S_2 & S_3 & \ldots & S_{r+1} \\ \vdots & & & \vdots \\ S_r & S_{r+1} & \ldots & S_{2r-1} \end{pmatrix} \begin{pmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{r-1} \end{pmatrix} = \begin{pmatrix} S_{r+1} \\ S_{r+2} \\ \vdots \\ S_{2r} \end{pmatrix}.$$

Thus, $\tau_j = \sigma_j$ for $j = 0, 1, \ldots, r-1$, and

$$\Sigma(x) = \tau(x) = (x + \alpha^{e_1})(x + \alpha^{e_2}) \cdots (x + \alpha^{e_r}).$$

□

15.6 Other applications

We can send eleven bits accounting for up to four errors, using the field

$$\mathbb{F}_{32} = \mathbb{F}_2[x]/(x^5 + x^2 + 1),$$

$\alpha = [x]_{x^5+x^2+1}$, and the polynomial

$$m_0(x) = (x^5 + x^2 + 1)(x^5 + x^4 + x^3 + x^2 + 1)(x^5 + x^4 + x^2 + x + 1)(x^5 + x^3 + x^2 + x + 1).$$

We can send six bits accounting for up to seven errors, using $\alpha \in \mathbb{F}_{32}$ and the polynomial

$$m_1(x) = m_0(x)(x^5 + x^4 + x^3 + x + 1).$$

A table showing the capabilities of $\mathbb{F}_{64}$ follows.

information bits	errors allowed
30	6
24	7
18	10
16	11
10	13
7	15

16 Finite fields

Throughout this final section $\mathbb{F}$ will always be a finite field.

16.1 The size of finite fields

Definition 16.1.1. If $n \in \mathbb{Z}$, then we can use n to denote an element of $\mathbb{F}$. If $n \geq 0$, then it stands for $1 + 1 + \ldots + 1$ where there are $|n|$ ones. If $n < 0$ then it stands for $(-1) + (-1) + \ldots + (-1)$ where there are $|n|$ minus ones.

Lemma 16.1.2. *If $\mathbb{F}$ is a finite field then there exists an $n \in \mathbb{N}$ such that $n = 0$ in $\mathbb{F}$.*

Proof. The set $\{n \in \mathbb{F} : n \in \mathbb{N}\} \subseteq \mathbb{F}$ is finite, so for some $n, m \in \mathbb{N}$ with $n > m$ we have $n = m$ in $\mathbb{F}$. Thus, $n - m = 0$ in $\mathbb{F}$. □

Definition 16.1.3. If $\mathbb{F}$ is a finite field the *characteristic* of $\mathbb{F}$ is given by

$$\text{char } \mathbb{F} = \min\{n \in \mathbb{N} : n = 0 \text{ in } \mathbb{F}\}.$$

Proposition 16.1.4. *If $\mathbb{F}$ is a finite field then the characteristic of $\mathbb{F}$ is prime.*

Proof. Let $n = \text{char } \mathbb{F}$ and suppose for contradiction that $n = ab$ where $a, b \in \mathbb{N}$ and $a, b > 1$. By definition of characteristic, we have $n = 0$ in $\mathbb{F}$. Using the distributivity of addition, one can check that $n = ab$ in $\mathbb{F}$. Since $\mathbb{F}$ is a field, either $a = 0$ or $b = 0$ in $\mathbb{F}$. Since $a, b < n$, this contradicts the definition of the characteristic as the minimum n with $n = 0$ in $\mathbb{F}$. □

Theorem 16.1.5. *If $\mathbb{F}$ is a finite field then $|\,\mathbb{F}\,| = p^n$ for some prime p and some $n \in \mathbb{N}$.*

Proof. Let $p = \text{char } \mathbb{F}$. Then $\mathbb{Z}/p \subseteq \mathbb{F}$ is a subfield of $\mathbb{F}$. In fact, $\mathbb{F}$ is a vector space over $\mathbb{Z}/p$. Let $n = \dim_{\mathbb{Z}/p} \mathbb{F}$. Counting gives $|\,\mathbb{F}\,| = p^n$. □

16.2 The order of elements in $\mathbb{F}^\times$

Notation 16.2.1. If $\mathbb{F}$ is a field we write $\mathbb{F}^\times$ for the group of units $U(\mathbb{F}) = \mathbb{F} \setminus \{0\}$.

The following theorem, definition, and the propositions follow exactly those of section 10.1.

Theorem 16.2.2. *Suppose that $\mathbb{F}$ is a field of size q. For each $a \in \mathbb{F}^\times$, we have $a^{q-1} = 1$.*

Definition 16.2.3. Suppose $a \in \mathbb{F}^\times$. The smallest $e \in \{1, \ldots, |\,\mathbb{F}\,| - 1\}$ with $a^e = 1$ is called *the order of a*. We write $\text{ord}_{\mathbb{F}}(a)$ for the order of a.

Proposition 16.2.4. *Suppose that $a \in \mathbb{F}^\times$ and $a^d = 1$. Then the order of a divides d.*

Proposition 16.2.5. *Suppose that $a \in \mathbb{F}^\times$ and that a has order e. The order of a^d is $e/\gcd(d, e)$.*

Corollary 16.2.6. *If $\mathbb{F}$ is a field of size q. Then $x^q - x \in \mathbb{F}[x]$ factors as*

$$\prod_{a \in \mathbb{F}} (x - a)$$

and elements of $\mathbb{F}^\times$ have order dividing $q - 1$.

Proof. The previous theorem tells us that for each $a \in \mathbb{F}^\times$, we have $a^{q-1} = 1$. Multiplying by a gives $a^q = a$, and this is true for $a = 0$, too. Thus, each $a \in \mathbb{F}$ is a root of $x^q - x$, so that for each $a \in \mathbb{F}$, $x - a$ divides $x^q - x$. Since $\mathbb{F}[x]$ is a unique factorization domain, this gives the factorization stated. The final statement follows from the theorem and the first proposition. □

16.3 Two big theorems: existence of finite fields and primitive elements

The following proposition is fairly mind-blowing. It is easy enough to prove if you know finite fields already. With the proof of it I have in mind, using it for the theorem which follows would be cyclic. But we note it because it is useful for small p^n, and just generally kind of awesome.

Proposition 16.3.1. *Suppose $n \in \mathbb{N}$ and write $n = p_1^{n_1} \cdots p_r^{n_r}$, where $2 \leq p_1 < \ldots < p_r$ are primes and $n_1, \ldots, n_r \geq 1$ (the p_j and n_j are unique by the fundamental theorem of arithmetic).*

Let $p \in \mathbb{N}$ be prime, $q = p^n$, and $q_j = p^{n/p_j}$. Let $f_0(x) = x^q - x \in \mathbb{Z}/p[x]$, for $1 \leq j \leq r$ let

$$f_j(x) = \frac{f_{j-1}(x)}{\gcd(f_{j-1}(x), x^{q_j} - x)},$$

and let $f(x) = f_r(x)$.

Then $\deg f(x) \geq n$ and $f(x)$ is a product of all the monic irreducibles in $\mathbb{Z}/p[x]$ of degree n. Thus, to test irreducibility of a degree n polynomial $g(x) \in \mathbb{Z}/p[x]$, you can calculate $\gcd(f(x), g(x))$.

Example 16.3.2. Notice that $64 = 2^6 = 2^{2 \cdot 3}$, $2^2 = 4$, and $2^3 = 8$. In $\mathbb{F}_2[x]$ we have

$$\frac{(x^{64} + x)(x^2 + x)}{(x^4 + x)(x^8 + x)} = (x^{42} + x^{21} + 1)(x^{12} + x^{11} + x^9 + x^8 + x^6 + x^4 + x^3 + x + 1).$$

This polynomial is the product of the nine monic irreducibles in $\mathbb{F}_2[x]$ of degree six.

The following theorem gives the existence and uniqueness of finite fields of all possible sizes.

Theorem 16.3.3. *If $p \in \mathbb{N}$ is a prime and $n \in \mathbb{N}$, then there exists a finite field $\mathbb{F}$ of size p^n.*

Moreover, any two such fields are isomorphic.

Proof idea. We do not have the necessary technology available to us to prove this yet. However, some useful things can be said in the direction of a proof.

Let $q = p^n$. The corollary of the last section inspires the proof. If such a field $\mathbb{F}$ existed, then the corollary would tell us that the elements of $\mathbb{F}$ are precisely the roots of $x^q - x$.

1. We know that $x^2 + 1 \in \mathbb{R}[x]$ does not factor but, by letting $\mathbb{C} = \mathbb{R}[x]/(x^2 + 1)$ and setting $i = [x]_{x^2+1}$, we can factor it as $(x - i)(x + i)$.

2. We have seen that $x^2 + x + 1 \in \mathbb{F}_2[x]$ does not factor but, by letting $\mathbb{F}_4 = \mathbb{F}_2[x]/(x^2 + x + 1)$ and setting $\alpha = [x]_{x^2+x+1}$, we can factor it as $(x + \alpha)(x + \alpha^2)$.

3. We have seen that $x^4 + x + 1 \in \mathbb{F}_2[x]$ does not factor but, by letting $\mathbb{F}_{16} = \mathbb{F}_2[x]/(x^4 + x + 1)$ and setting $\alpha = [x]_{x^4+x+1}$, we can factor it as $(x + \alpha)(x + \alpha^2)(x + \alpha^4)(x + \alpha^8)$.

4. We know that $x^2 - 2 \in \mathbb{Q}[x]$ does not factor but, by considering $\mathbb{Q}[x]/(x^2 - 2)$ and setting $\sqrt{2} = [x]_{x^2-2}$, we can factor it as $(x - \sqrt{2})(x + \sqrt{2})$.

This suggests setting $\mathbb{F}_q = \mathbb{F}[x]/(x^q - x)$ but this doesn't work.

1. $i \in \mathbb{C}$ is also a root of $x^4 - 1$ but we don't let $\mathbb{C} = \mathbb{R}[x]/(x^4 - 1)$; this would not even be a field since $x^4 - 1$ is not irreducible.

2. $\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$ is also a root of $x^4 - 4$ but we don't bother with $Q[x]/(x^4 - 4)$; again this is not even a field.

What we want is the smallest field containing $\mathbb{Z}/p$ where $x^q - x$ does factor. Such a thing is called *the splitting field of $x^q - x$ over $\mathbb{Z}/p$* and there is a general procedure for constructing it. We cannot go into this now but the above examples suggest the strategy.

Take an irreducible factor $f(x)$ of $x^q - x$ with degree bigger than 1, and construct $\mathbb{Z}/p[x]/(f(x))$. Then see what the status of $x^q - x$ is. If it factors completely we are done; otherwise keep going like this until it does factor. Eventually you'll get a field which you can show has q elements.

In fact, if one uses the lemma above, one can find an irreducible $f(x) \in \mathbb{Z}/p[x]$ of degree n and the process above will be one step long.

Splitting fields are unique up to isomorphism, and this allows one to prove that any two finite fields of the same size are isomorphic. □

The following theorem can be expressed by saying that the multiplicative group of a finite field is cyclic.

Theorem 16.3.4. *If $\mathbb{F}$ is a finite field of size q, then there exists an $\alpha \in \mathbb{F}^\times$ with order $q - 1$.*

Definition 16.3.5. If $\mathbb{F}$ is a finite field of size q, an element $\alpha \in \mathbb{F}^\times$ with order $q - 1$ is said to be *primitive*. The theorem just stated is called *the primitive element theorem*.

Proof of primitive element theorem. Finally, we wish we had defined the *lowest common multiple*. Suppose that we did; you can figure out the definition. Define

$$e := \operatorname{lcm}\{\operatorname{ord}_{\mathbb{F}}(a) : a \in \mathbb{F}^\times\}.$$

For each $a \in \mathbb{F}^\times$ we have $a^e = 1$, so that each element of $\mathbb{F}^\times$ is a root of $x^e - 1$. This means that $e \geq (q - 1)$ and so it is enough to show that an element $a \in \mathbb{F}^\times$ has order e.

Write $e = p_1^{n_1} \cdots p_r^{n_r}$, where $2 \leq p_1 < \ldots < p_r$ are primes and $n_1, \ldots, n_r \geq 1$ (the p_j and n_j are unique by the fundamental theorem of arithmetic). By definition of e, we have an element $b_j \in \mathbb{F}^\times$ with $p_j^{n_j} | \operatorname{ord}_{\mathbb{F}}(b_j)$; a suitable power $a_j \in \mathbb{F}^\times$ of b_j has $\operatorname{ord}_{\mathbb{F}}(a_j) = p_j^{n_j}$. We will show that $a = a_1 a_2 \cdots a_r$ has order e.

Suppose that $m \in \mathbb{N}$ and $a^m = 1$. For $1 \leq j \leq r$ we have

$$a_j^m = a_1^{-m} \cdots a_{j-1}^{-m} a_{j+1}^{-m} \cdots a_r^{-m}.$$

So, if $q_j = p_1^{n_1} \cdots p_{j-1}^{n_{j-1}} p_{j+1}^{n_{j+1}} \cdots p_r^{n_r}$, then $a_j^{mq_j} = 1$. Since $\operatorname{ord}_{\mathbb{F}}(a_j) = p_j^{n_j}$, this gives $p_j^{n_j} | (mq_j)$, and since p_j and q_j are coprime, $p_j^{n_j} | m$. This holds for all j, so that $e | m$. In particular, $m \geq e$ so that $\operatorname{ord}_{\mathbb{F}}(a) = e$, as required. □