Math 131AH - Week 1
Textbook pages covered: none

- What is analysis?
- Why do analysis?
- The natural number system
- The principle of induction
- The integers
- The rationals

* * * * *

What is analysis?

- This course is an honors-level introduction to *real analysis*: the analysis of the real numbers, sequences and series of real numbers, and real-valued functions. This is related to, but is distinct from, *complex analysis*, which concerns the analysis of the complex numbers and complex functions, *harmonic analysis*, which concerns the analysis of harmonics (waves) such as sine waves, and how they synthesize other functions via the Fourier transform, *functional analysis*, which focuses much more heavily on functions (and how they form things like vector spaces), and so forth. *Analysis* is the rigorous study of such objects, with a focus on trying to pin down precisely and accurately the qualitative and quantitative behavior of these objects. Real analysis is the theoretical foundation which underlies *calculus*, which is the collection of computational algorithms which one uses to manipulate functions.
- In this course we will be studying many objects which will be familiar to you from lower-division mathematics: numbers, sequences, series, limits, functions, definite integrals, derivatives, and so forth. You already have a great deal of experience knowing how to *compute* with these objects; however here we will be focused more on the underlying theory for these objects. We will be concerned with questions such as the following:

- 1. What is a real number? Is there a largest real number? After 0, what is the "next" real number (i.e. what is the smallest positive real number)? Can you cut a real number into pieces infinitely many times? Why does a number such as 2 have a square root, while a number such as -2 does not? If there are infinitely many reals and infinitely many rationals, how come there are "more" real numbers than rational numbers?

- 2. How do you take the limit of a sequence of real numbers? Which sequences have limits and which ones don't? If you can stop a sequence from escaping to infinity, does this mean that it must eventually settle down and converge? Can you add infinitely many real numbers together and still get a finite real number? Can you add infinitely many rational numbers together and end up with a non-rational number? If you rearrange the elements of an infinite sum, is the sum still the same?

- 3. What is a function? What does it mean for a function to be continuous? differentiable? integrable? bounded? can you add infinitely many functions together? What about taking limits of sequences of functions? Can you differentiate an infinite series of functions? What about integrating? If a function $f(x)$ takes the value of $f(0) = 3$ when $x = 0$ and $f(1) = 5$ when $x = 1$, does it have to take every intermediate value between 3 and 5 when x goes between 0 and 1? Why?

- You may already know how to answer some of these questions from your lower-division classes, but most likely these sorts of issues were only of secondary importance to those courses; the emphasis was on getting you to perform computations, such as computing the integral of $x \sin(x^2)$ from $x = 0$ to $x = 1$. But now that you are comfortable with these objects and already know how to do all the computations, we will go back to the theory and try to *really* understand what is going on.

* * * * *

Why do we need analysis?

- It is a fair question to ask, "why bother?", when it comes to analysis. There is a certain philosophical satisfaction in knowing *why* things

work, but a pragmatic person may argue that one only needs to know *how* things work to do real-life problems. The calculus training you receive in lower division is certainly adequate for you to begin solving many problems in physics, chemistry, biology, economics, computer science, finance, engineering, or whatever else you end up doing - and you can certainly use things like the chain rule, LHôpital's rule, or integration by parts without knowing why these rules work, or whether there are any exceptions to these rules. However, one can get into trouble if one applies rules without knowing where they came from and what the limits of their applicability are. Let me give some examples in which several of these familiar rules, if applied blindly without knowledge of the underlying analysis, can lead to disaster.

- **Division by zero.** This is a very familiar one to you: the cancellation law $ac = bc \Rightarrow a = b$ does not work when $c = 0$. For instance, the identity $1 \times 0 = 2 \times 0$ is true, but if one blindly cancels the 0 then one obtains $1 = 2$, which is false. In this case it was obvious that one was dividing by zero; but in other cases it can be more hidden.

- **Divergent series.** You have probably seen geometric series such as the infinite sum

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots.$$

You have probably seen the following trick to sum this series: if we call the above sum S, then if we multiply both sides by 2, we obtain

$$2S = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 2 + S$$

and hence $S = 2$, so the series sums to 2. However, if you apply the same trick to the series

$$S = 1 + 2 + 4 + 8 + 16 + \ldots$$

one gets nonsensical results:

$$2S = 2 + 4 + 8 + 16 + \ldots = S - 1 \Rightarrow S = -1.$$

So the same reasoning that shows that $1 + \frac{1}{2} + \frac{1}{4} + \ldots = 2$ also gives that $1+2+4+8+\ldots = -1$. Why is it that we trust the first equation but not the second? A similar example arises with the series

$$S = 1 - 1 + 1 - 1 + 1 - 1 + \ldots;$$

we can write

$$S = 1 - (1 - 1 + 1 - 1 + \ldots) = 1 - S$$

and hence that $S = 1/2$; or instead we can write

$$S = (1 - 1) + (1 - 1) + (1 - 1) + \ldots = 0 + 0 + \ldots$$

and hence that $S = 0$; or instead we can write

$$S = 1 + (-1 + 1) + (-1 + 1) + \ldots = 1 + 0 + 0 + \ldots$$

and hence that $S = 1$. Which one is correct?

- **Interchanging sums** Consider the following fact of arithmetic. Consider any matrix of numbers, e.g.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

and compute the sums of all the rows and the sums of all the columns, and then total all the row sums and total all the column sums. In both cases you will get the same number - the total sum of all the entries in the matrix:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad \begin{pmatrix} 6 \\ 15 \\ 24 \end{pmatrix}$$
$$\begin{pmatrix} 12 & 15 & 18 \end{pmatrix} \quad 45$$

To put it another way, if you want to add all the entries in a $m \times n$ matrix together, it doesn't matter whether you sum the rows first or sum the columns first, you end up with the same answer. (Before the invention of computers, accountants and bookkeepers would use this

fact to guard against making errors when balancing their books). In series notation, this fact would be expressed as

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij},$$

if a_{ij} denoted the entry in the i^{th} row and j^{th} column of the matrix.

- Now one might think that this rule should extend easily to infinite series:

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

Indeed, if you use infinite series a lot in your work, you will find yourself having to switch summations like this fairly often. Another way of saying this fact is that in an infinite matrix, the sum of the row-totals should equal the sum of the column-totals. However, despite the reasonableness of this statement, it is actually false! Here is a counterexample:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ 0 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

If you sum up all the rows, and then add up all the row totals, you get 1; but if you sum up all the columns, and add up all the column totals, you get 0! So, does this mean that summations for infinite series should not be swapped, and that any argument using such a swapping should be distrusted?

- **Interchanging integrals** The interchanging of integrals is a trick which occurs just as commonly as integrating by sums in mathematics. Suppose one wants to compute the volume under a surface $z = f(x, y)$ (let us ignore the limits of integration for the moment). One can do it by slicing parallel to the x axis: for each fixed value of y, we can

compute an area $\int f(x,y)\ dx$, and then we integrate the area in the y variable to obtain the volume

$$V = \int \int f(x,y) dx dy.$$

Or we could slice parallel to the y axis to obtain an area $\int f(x,y)\ dy$, and then integrate in the x axis to obtain

$$V = \int \int f(x,y) dy dx.$$

This seems to suggest that one should always be able to swap integral signs:

$$\int \int f(x,y)\ dx dy = \int \int f(x,y)\ dy dx.$$

And indeed, people swap integral signs all the time, because sometimes one variable is easier to integrate in first than the other. However, just as infinite sums sometimes cannot be swapped, integrals are also sometimes dangerous to swap. An example is with the integrand $e^{-xy} - xye^{-xy}$. Suppose we believe that we can swap the integrals:

$$\int_0^\infty \int_0^1 e^{-xy} - xye^{-xy}\ dy\ dx = \int_0^1 \int_0^\infty e^{-xy} - xye^{-xy}\ dx\ dy.$$

Since

$$\int_0^1 e^{-xy} - xye^{-xy}\ dy = ye^{-xy}|_{y=0}^{y=1} = e^{-x}$$

the left-hand side is $\int_0^\infty e^{-x}\ dx = -e^{-x}|_0^\infty = 1$. But since

$$\int_0^\infty e^{-xy} - xye^{-xy}\ dx = xe^{-xy}|_{x=0}^{x=\infty} = 0$$

the right-hand side is $\int_0^1 0\ dx = 0$. Clearly $1 \neq 0$, so there is an error somewhere; but you won't find one anywhere except in the step where we interchanged the integrals. So how do we know when to trust the interchange of integrals?

- **Limits and lengths.** When you learn about integration and how it relates to the area under a curve, you were probably presented with some picture in which the area under the curve was approximated by a bunch of rectangles, whose area was given by a Riemann sum, and then one somehow "took limits" to replace that Riemann sum with an integral, which then presumably matched the actual area under the curve. Perhaps a little later in your lower division class, you learnt how to compute the length of a curve by a similar method - approximate the curve by a bunch of line segments, compute the length of all the line segments, then take limits again to see what you get.

- However, it should come as no surprise by now that this approach also can lead to nonsense if used incorrectly. Consider the right-angled triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$, and suppose we wanted to compute the length of the hypotenuse of this triangle. Pythagoras's theorem tells us that this hypotenuse has length $\sqrt{2}$, but suppose for some reason that we did not know about Pythagoras's theorem, and wanted to compute the length using calculus methods. Well, one way to do so is to approximate the hypotenuse by horizontal and vertical edges. Pick a large number N, and approximate the hypotenuse by a "staircase" consisting of N horizontal edges of equal length, alternating with N vertical edges of equal length. Clearly these edges all have length $1/N$, so the total length of the staircase is $2N/N = 2$. If one takes limits as N goes to infinity, the staircase clearly approaches the hypotenuse, and so in the limit we should get the length of the hypotenuse. However, as $N \to \infty$, the limit of $2N/N$ is 2, not $\sqrt{2}$, so we have an incorrect value for the length of the hypotenuse. How did this happen?

- The analysis you learn in the Math 131 series will help you resolve these questions, and will let you know when these rules (and others) are justified, and when they are illegal, thus separating the useful applications of these rules from the nonsense. Thus they can prevent you from making mistakes, and can help you place these rules in a wider context. Moreover, as you learn analysis you will develop an "analytical way of thinking", which will help you whenever you come into contact with any new rules of mathematics, or when dealing with situations

which are not quite covered by the standard rules (e.g. what if your functions are complex-valued instead of real-valued? What if you are working on the sphere instead of the plane? What if your functions are not continuous, but are instead things like square waves and delta functions? What if your functions, or limits of integration, or limits of summation, are occasionally infinite?). You will develop a sense of *why* a rule in mathematics (e.g. the chain rule) works, how to adapt it to new situations, and what its limitations (if any) are; this will allow you to apply the mathematics you have already learnt more confidently and correctly.

* * * * *

Starting at the beginning: the natural numbers

- In this honors analysis class, we will want to go over much of the material you have learnt in high school and in lower-division classes, but to do so as rigorously as possible. To do so we will have to begin at the very basics - indeed, we will go back to the concept of *numbers* and what their properties are. Of course, you have dealt with numbers for over ten years and you know very well how to manipulate the rules of algebra to simplify any expression involving numbers, but we will now turn to a more fundamental issue, which is *why* the rules of algebra work... for instance, why is it true that $a(b+c)$ is equal to $ab + ac$ for any three numbers a, b, c? This is not an arbitrary choice of rule; it can be proven from more primitive, and more fundamental, properties of the number system. This will teach you a new skill - how to prove complicated properties from simpler ones. You will find that even though a statement may be "obvious", it may not be easy to prove; the material here will give you plenty of practice in doing so, and in the process will lead you to think about *why* an obvious statement really is obvious. One skill in particular that you will pick up here is the use of *mathematical induction*, which is a basic tool in proving things in many areas of mathematics.

- So in the first few lectures we will re-acquaint you with various number systems that are used in real analysis. In increasing order of sophistication, they are the *natural numbers* $\mathbf{N}$; the *integers* $\mathbf{Z}$; the *rationals* $\mathbf{Q}$, and the *real numbers* $\mathbf{R}$. (There are other number systems such as the

complex numbers $\mathbf{C}$, but we will not study them in this course). The natural numbers are the most primitive of the number systems, but they are used to build the integers, which in turn are used to build the rationals, which in turn build the real numbers. Thus to begin at the very beginning, we must look at the natural numbers. We will consider the following question: how does one actually *define* the natural numbers? (This is a very different question as to how to *use* the natural numbers, which is something you of course know how to do very well. It's like the difference between knowing how to use, say, a computer, versus knowing how to *build* that computer).

- This question is more difficult to answer than it looks. The basic problem is that you have used the natural numbers for so long that they are embedded deeply into your mathematical thinking, and you can make various implicit assumptions about these numbers (e.g. that $a+b$ is always equal to $b+a$) without even thinking; it is difficult to let go for a moment and try to inspect this number system as if it is the first time you have seen it. So in what follows I will have to ask you to perform a rather difficult task: try to set aside, for the moment, everything you know about the natural numbers; forget that you know how to count, to add, to multiply, to manipulate the rules of algebra, etc. We will try to introduce these concepts one at a time and try to identify explicitly what our assumptions are as we go along - and not allow ourselves to use more "advanced" tricks - such as the rules of algebra - until we have actually proven them. This may seem like an irritating constraint, especially as we will spend a lot of time proving statements which are "obvious", but it is necessary to do this suspension of known facts to avoid circularity (e.g. using an advanced fact to prove a more elementary fact, and then later using the elementary fact to prove the advanced fact). Also, it is an excellent exercise for really affirming the foundations of your mathematical knowledge, and practicing your proofs and abstract thinking here will be invaluable when we move on to more advanced concepts, such as real numbers, then functions, then sequences and series, then differentials and integrals, and so forth. In short, the results here may seem trivial, but the journey is much more important than the destination, for now. (After this week we can resume using the laws of algebra etc. without having to rederive them

each time).

- We will also forget that we know the decimal system, which of course is an extremely convenient way to manipulate numbers, but it is not something which is fundamental to what numbers are. (For instance, one could use an octal or binary system instead of the decimal system, or even the Roman numeral system, and still get exactly the same set of numbers). Besides, if one tries to fully explain what the decimal number system is, it isn't as natural as you might think. Why is 00423 the same number as 423, but 32400 isn't the same number as 324? How come $123.4444\ldots$ is a real number, but $\ldots 444.321$ isn't? And why do we have to do all this carrying of digits when adding or multiplying? Why is $0.999\ldots$ the same number as 1? What is the smallest positive real number? Isn't it just $0.00\ldots001$? So to set aside these problems, we will not try to assume any knowledge of the decimal system (though we will of course still refer to numbers by their familiar names such as 1,2,3, etc. instead of using other notation such as I,II,III or 0++, (0++)++, ((0++)++)++ (see below) so as not to be needlessly artificial).

- One informal definition of the natural numbers $\mathbf{N}$ is that they are simply the collection of numbers

$$\mathbf{N} := \{0, 1, 2, 3, 4, \ldots\};$$

  thus the natural numbers are what you get by starting at 0 and then counting forward indefinitely. (In some texts the natural numbers start at 1 instead of 0, but this is a matter of notational convention more than anything else. In this course we shall refer to the set $\{1, 2, 3, \ldots\}$ as the *positive integers* $\mathbf{Z}^+$ rather than the natural numbers).

- In particular: we will postulate the existence of some number system $\mathbf{N}$, whose elements we shall now refer to as *natural numbers*. This solves the problem of what a natural number is - it's an element of $\mathbf{N}$ - but of course we now have to work out what $\mathbf{N}$ is.

- (Remark: This is not the only way to define the natural numbers. Another approach is to talk about the cardinality of finite sets, for instance one could take a set of five elements and define 5 to be the number of elements in that set. But this requires a substantial amount

of set theory to set up - in particular, one needs to define "cardinality" and "finite", so we will not pursue this approach here, although we will of course introduce these notions later on in this course).

- This definition of "start at 0 and count indefinitely" seems like an intuitive enough definition of $\mathbf{N}$, but it is not entirely acceptable, because it leaves many questions unanswered. For instance: how do we know we can keep counting indefinitely? Could we ever cycle back to 0? Also, how do you perform operations such as addition? multiplication? exponentiation? etc.

- We can answer the latter question first: we can define complicated operations in terms of simpler operations. Exponentiation is nothing more than repeated multiplication: 5^3 is nothing more than three fives multiplied together. Multiplication is nothing more than repeated addition; 5×3 is nothing more than three fives added together. (Subtraction and division will not be covered here, because they are not operations which are well-suited to the natural numbers; they will have to wait for the integers and rationals, respectively). And addition? It is nothing more than the repeated operation of *counting forward*, or *incrementing*. If you add three to five, what you are doing is incrementing five three times. On the other hand, incrementing seems to be a pretty fundamental operation, not reducible to any simpler operation; indeed, it is the first operation one learns on numbers, even before learning to add.

- Thus, to define the natural numbers, we will use two fundamental concepts: the zero number 0, and the increment operation. In deference to modern computer languages, we will use $n{+}{+}$ to denote the increment or *successor* of n, thus for instance $3{+}{+} = 4$, $(3{+}{+}){+}{+} = 5$, etc. (This is slightly different usage from that in computer languages such as C, where $n{+}{+}$ actually *redefines* the value of n to be its successor; however in mathematics we try not to define a variable more than once in any given setting, as it can often lead to confusion (as many of the statements which were true for the old value of the variable now become false)).

- So, it seems like we want to say that $\mathbf{N}$ consists of 0 and everything

which can be obtained from 0 by incrementing: **N** should consist of $0, 0{+}{+}, (0{+}{+}){+}{+}, ((0{+}{+}){+}{+}){+}{+}$, etc.. If we start writing down what this means about the natural numbers, we thus see that we should have the following:

- **Axiom I.** 0 is a natural number.
- **Axiom II.** If n is a natural number, then $n{+}{+}$ is also a natural number.
- Thus for instance, from Axiom I and two applications of Axiom II, we see that $(0{+}{+}){+}{+}$ is a natural number. Of course, this notation will begin to get unwieldy, so we adopt a convention to write these numbers in more familiar notation:
- **Definition** We define 1 to be the number $0{+}{+}$, 2 to be the number $(0{+}{+}){+}{+}$, 3 to be the number $((0{+}{+}){+}{+}){+}{+}$, etc. (In other words, $1 := 0{+}{+}$, $2 := 1{+}{+}$, $3 := 2{+}{+}$, etc. In this course I use "$x := y$" to denote the statement that x is defined to equal y.)
- Thus for instance, we have
- **Proposition** 3 is a natural number.
- **Proof** By Axiom I, 0 is a natural number. By Axiom II, $0{+}{+} = 1$ is a natural number. By Axiom II again, $1{+}{+} = 2$ is a natural number. By Axiom II again, $2{+}{+} = 3$ is a natural number. □
- It may seem that this is enough to describe the natural numbers. However, have not pinned down completely the behavior of **N**. For instance, what may happen is that after starting with 0 and incrementing, one might wrap around back to 0: it might be that $0{+}{+}$ is equal to 1, $1{+}{+}$ is equal to 2, $2{+}{+}$ is equal to 3, but $3{+}{+}$ is equal to 0 (and also equal to 4, by definition). This is in fact what happens when one uses a computer to try to store a natural number: if one starts at 0 and increments indefinitely, eventually the computer will overflow its memory and the number will wrap around back to 0. However, we believe that in the ideal world of mathematics, this does nto happen. To prevent this we will impose another axiom:

- **Axiom III.** 0 is not the successor of any natural number; i.e. we have $n{+}{+} \neq 0$ for every natural number n.

- Now we can show that certain types of wraparound do not occur: for instance we have

- **Proposition.** 4 is not equal to 0.

- Don't laugh! Because of the way we have defined 4 - it is the increment of the increment of the increment of the increment of 0 - it is not necessarily true *a priori* that this number is not the same as zero, even if it is "obvious". ("a priori" is Latin for "beforehand" - it refers to what one already knows or assumes to be true before one begins a proof or argument. The opposite is "a posteriori" - what one knows to be true after the proof or argument is concluded). Note that in a standard two-byte computer representation of a natural number, for instance, 65536 is equal to 0 (using our definition of 65536 as equal to 0 incremented sixty-five thousand, five hundred and thirty-six times).

- **Proof.** By definition, $4 = 3{+}{+}$. By Axioms I and II, 3 is a natural number. Thus by Axiom III, $3{+}{+} \neq 0$, i.e. $4 \neq 0$. □

- However, it is still possible that our number system behaves in other pathological ways. For instance, the incrementing might hit a "ceiling" at, say, 4: $0{+}{+} = 1$, $1{+}{+} = 2$, $2{+}{+} = 3$, $3{+}{+} = 4$, but $4{+}{+} = 4$ (or in other words that $5 = 4$). This does not contradict Axioms I,II,III. Or it might wrap around from 4 to 1: $0{+}{+} = 1$, $1{+}{+} = 2$, $2{+}{+} = 3$, $3{+}{+} = 4$, $4{+}{+} = 1$ (so $5 = 1$). There are many ways to prohibit this from happening, but one of the simplest is to assume the following axiom:

- **Axiom IV.** Different natural numbers must have different successors; i.e. if n, m are natural numbers and $n \neq m$, then $n{+}{+} \neq m{+}{+}$. Equivalently, if $n{+}{+} = m{+}{+}$, then we must have $n = m$.

- Thus, for instance, we have

- **Proposition.** 6 is not equal to 2.

- **Proof.** Suppose for contradiction that $6 = 2$. Then $5{+}{+} = 1{+}{+}$, so by Axiom IV $5 = 1$, so that $4{+}{+} = 0{+}{+}$. By Axiom IV again we then have $4 = 0$, which contradicts our previous proposition. □

- As one can see from this proposition, it now looks like we can keep all of the natural numbers distinct from each other. There is however still one more problem: while the axioms (particularly Axioms I and II) allow us to confirm that $0, 1, 2, 3, \ldots$ are elements of $\mathbf{N}$, there is the problem that there may be other "rogue" elements in our number system which are not of this form. For instance, we cannot preclude at this time that $\mathbf{N}$ could in fact look like the following collection of integers and half-integers:

  $$\mathbf{N} := \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, \ldots\};$$

  one can check that Axioms I-IV are still satisfied for this set. (This example is cheating a little since we are using real numbers, which we're not supposed to use yet; but this is only an example, not part of the main discussion).

- What we want is some axiom which says that the only numbers in $\mathbf{N}$ are those which can be obtained from 0 and incrementation - in order to exclude elements such as 0.5. But it is difficult to quantify what we mean by "can be obtained from" without already using the natural numbers, which we are trying to define. Fortunately, there is an ingenious solution to try to capture this fact:

- **Axiom V. (Principle of induction).** Let $P(n)$ be any property pertaining to a natural number n. Suppose that $P(0)$ is true, and suppose that whenever $P(n)$ is true, $P(n{+}{+})$ is also true. Then $P(n)$ is true for every natural number n.

- We are a little vague on what "property" means at this point, but some possible examples of $P(n)$ might be "n is even"; "n is equal to 3"; "n solves the equation $(n+1)^2 = n^2 + 2n + 1$"; and so forth. Of course we haven't defined many of these concepts yet, but when we do, Axiom V will apply to these properties. (A logical remark: Because this axiom refers not just to *variables*, but also *properties*, it is of a different nature

than the other four axioms; the first four axioms use what is called *first-order logic*, whereas Axiom V is a statement phrased using *second-order logic*. To discuss this distinction further is far beyond the scope of this course, though, and falls in the realm of philosophy and logic).

- The intuition behind this axiom is the following. Suppose $P(n)$ is such that $P(0)$ is true, and such that whenever $P(n)$ is true, then $P(n{+}{+})$ is true. Then since $P(0)$ is true, $P(0{+}{+}) = P(1)$ is true. Since $P(1)$ is true, $P(1{+}{+}) = P(2)$ is true. Repeating this indefinitely, we see that $P(0)$, $P(1)$, $P(2)$, $P(3)$, etc. are all true - however this line of reasoning will never let us conclude that $P(0.5)$, for instance, is true. Thus Axiom V should not hold for number systems which contain unnecessary elements such as 0.5. (Indeed, one can give a "proof" of this fact. Apply Axiom V to the property $P(n) = n$ "is not a half-integer", i.e. an integer plus 0.5. Then $P(0)$ is true, and if $P(n)$ is true, then $P(n{+}{+})$ is true (if n is not a half-integer. Then Axiom V asserts that $P(n)$ is true for all natural numbers n, i.e. no natural number can be a half-integer. In particular, 0.5 cannot be a natural number. This "proof" is not quite genuine, because we have not defined such notions as "integer", "half-integer", and "0.5" yet, but it should give you some idea as to how the principle of induction is supposed to prohibit any numbers other than the "true" natural numbers from appearing in $\mathbf{N}$.)

- The principle of induction gives us a way to prove that a property $P(n)$ is true for every natural number n. Thus in the rest of this course we will see many proofs which have a form like this:

- **Claim.** A certain property $P(n)$ is true for every natural number n.

- **Proof.** We use induction. We first verify the base case $n = 0$, i.e. we prove $P(0)$. (Insert proof of $P(0)$ here). Now suppose inductively that n is a natural number, and $P(n)$ has already been proven. We now prove $P(n{+}{+})$. (Insert proof of $P(n{+}{+})$, assuming that $P(n)$ is true, here). This closes the induction, and thus $P(n)$ is true for all numbers n. □

- Of course we will not necessarily use the exact template, wording, or

order in the above type of proof, but the proofs using induction will generally be something like the above form.

- Axioms I-V are known as the *Peano axioms* for the natural numbers. They are all very plausible, and so we shall make

- **Assumption.** There exists a number system $\mathbf{N}$, whose elements we will call *natural numbers*, for which Axioms I-V are true.

- We will refer to this number system $\mathbf{N}$ as *the* natural number system. (One could of course consider the possibility that there is more than one natural number system, e.g. we could have the Hindu-Arabic number system $\{0, 1, 2, 3, \ldots\}$ and the Roman number system $\{0, I, II, III, IV, V, VI, \ldots\}$, and if we really wanted to be annoying we could view these number systems as different (though equivalent). But there is no point in doing so; we only need one natural number system in order to do mathematics).

- We will not prove the above assumption, and it will be the only assumption we will ever make about our numbers. (It is possible to prove this Assumption using some machinery from set theory, but this only defers the problem, because set theory itself relies on some axioms, and at some point you have to make an assumption that there is a universe of sets which obeys those axioms. This is too far afield for this course, though; see Math 112 for more details). The remarkable thing is that from these five very primitive axioms, and a little dash of set theory, we can build all the other number systems, create functions, and do all the algebra and calculus that we are used to.

- One interesting feature about the natural numbers is that while each individual natural number is finite, the *set* of natural numbers is infinite; i.e. $\mathbf{N}$ is infinite but consists of individually finite elements. (The whole is greater than any of its parts). There are no infinite natural numbers; one can even prove this using Axiom V, provided one is comfortable with the notions of finite and infinite. (Clearly 0 is finite. Also, if n is finite, then clearly $n{+}{+}$ is also finite. Hence by Axiom V, all natural numbers are finite). So the natural numbers can *approach* infinity, but never actually reach it; infinity is not one of the natural numbers. (There are other number systems which admit "infinite"

numbers, such as the cardinals, ordinals, and p-adics, but they do not obey the principle of induction, and in any event are beyond the scope of this course).

- Note that our definition of the natural numbers is *axiomatic* rather than *constructive*. We have not told you what the natural numbers *are* (so we do not address such questions as what the numbers are made of, are they physical objects, what do they measure, etc.) - we have only listed some things you can do with them (in fact, the only operation we have defined on them right now is incrementation) and some of the properties that they have. This is how mathematics works - it treats its objects *abstractly*, caring only about what properties the objects have, not what the objects are or what they mean. If one wants to do mathematics, it does not matter whether a natural number means a certain arrangement of beads on an abacus, or a certain organization of bits in a computer's memory, or some more abstract concept with no physical substance; as long as you can increment them, see if two of them are equal, and later on do other arithmetic operations such as add and multiply, they qualify as numbers for mathematical purposes (provided they obey the requisite axioms, of course). (It is possible to construct the natural numbers from other mathematical objects - from sets, for instance - but there are multiple ways to construct a working model of the natural numbers, and it is pointless, at least from a mathematician's standpoint, as to argue about which model is the "true" one - as long as it obeys all the axioms and does all the right things, that's good enough to do maths).

- (A historical note: the realization that numbers could be treated axiomatically is very recent, not much more than a hundred years old. Before then, numbers were generally understood to be inextricably connected to some external concept, such as counting the cardinality of a set, measuring the length of a line segment, or the mass of a physical object, etc. This worked reasonably well, until one was forced to move from one number system to another; for instance, understanding numbers in terms of counting sheep or rocks is great for conceptualizing the numbers 3 and 5, but doesn't work so well for -3 or $1/3$ or $\sqrt{2}$ or $3+4i$; thus each great advance in the theory of numbers - negative numbers,

irrational numbers, complex numbers, even the number zero - led to a great deal of unnecessary philosophical anguish. The great discovery of the late nineteenth century was that numbers can be understood abstractly via axioms, without necessarily needing a conceptual model; of course a mathematician can use any of these models when it is convenient, to aid his or her intuition and understanding, but they can also be just as easily discarded when they begin to get in the way.)

- One consequence of the axioms is that we can now define sequences *recursively*. Suppose we want to build a sequence $a_0, a_1, a_2, \ldots$ by first defining a_0 to be some base value, e.g. $a_0 := c$, and then letting a_1 be some function of a_0, $a_1 := f_0(a_0)$, a_2 be some function of a_1, $a_2 := f_1(a_1)$, and so forth - in general, we set $a_{n++} := f_n(a_n)$ for some function f_n from $\mathbf{N}$ to $\mathbf{N}$. By using all the axioms together we will now conclude that this procedure will give a single value to the sequence element a_n for each natural number n. (More precisely, there is a unique function $a(n)$ from $\mathbf{N}$ to $\mathbf{N}$ such that $a(0) = c$ and $a(n{+}{+}) = f_n(a(n))$ for each natural number n).

- **Proof:** We use induction. We first observe that this procedure gives a single value to a_0, namely c. (None of the other definitions $a_{n++} := f_n(a_n)$ will redefine the value of a_0, because of Axiom III). Now suppose inductively that the procedure gives a single value to a_n. Then it gives a single value to a_{n++}, namely $a_{n++} := f_n(a_n)$. (None of the other definitions $a_{m++} := f_m(a_m)$ will redefine the value of a_{n++}, because of Axiom IV). This completes the induction, and so a_n is defined for each natural number n, with a single value assigned to each a_n. □

- Note how all of the axioms had to be used here. In a system which had some sort of wraparound, recursive definitions would not work because some elements of the sequence would constantly be redefined. For instance, in the system where $3{+}{+} = 0$, then there would be (at least) two conflicting definitions for a_0, either c or $f_3(a_3)$). In a system which had superfluous elements such as 0.5, the element $a_{0.5}$ would never be defined.

- Recursive definitions are very powerful; for instance, we can use them to define addition and multiplication, to which we now turn.

* * * * *

Addition

- The natural number system is very sparse right now: we have only one operation - incrementation - and a handful of axioms. But now we can build up more complex operations, such as addition.

- The way it works is the following. To add three to five should be the same as incrementing five three times - this is one increment more than adding two to five, which is one increment more than adding one to five, which is one increment more than adding zero to five, which should just give five. So we give a recursive definition for addition as follows.

- **Definition**. Let m be a natural number. To add zero to m, we define $0+m := m$. Now suppose inductively that we have defined how to add n to m. Then we can add $n{+}{+}$ to m by defining $(n{+}{+})+m := (n+m){+}{+}$.

- Thus $0 + m$ is m, $1 + m = (0{+}{+}) + m$ is $m{+}{+}$; $2 + m = (1{+}{+}) + m = (m{+}{+}){+}{+}$; and so forth. Thus for instance $2 + 3 = (3{+}{+}){+}{+} = 4{+}{+} = 5$. From our discussion of recursion in the previous section we see that we have defined $n + m$ for every integer n (here we are specializing the previous general discussion to the setting where $a_n = n + m$ and $f_n(a_n) = a_n{+}{+}$). Note that this definition is asymmetric: $3 + 5$ is incrementing 5 three times, while $5 + 3$ is incrementing 3 five times. It's not a priori clear why these two operations should be the same, but we will prove it shortly.

- Notice that we can prove easily, using Axioms I, II, and induction (Axiom IV), that the sum of two natural numbers is again a natural number. (Why?).

- Right now we only have two facts about addition: that $0 + m = m$, and that $(n{+}{+}) + m = (n + m){+}{+}$. Remarkably, this turns out to be enough to deduce everything else we know about addition.

- **Lemma 1.** For any natural number n, $n + 0 = n$.

- Note that we cannot deduce this immediately from $0 + m = m$ because we do not know yet that $a + b = b + a$.

- **Proof.** We use induction. The base case $0 + 0 = 0$ follows since we know that $0 + m = m$ for every natural number m, and 0 is a natural number. Now suppose inductively that $n + 0 = n$. We wish to show that $(n{+}{+}) + 0 = n{+}{+}$. But by definition of addition, $(n{+}{+}) + 0$ is equal to $(n + 0){+}{+}$, which is equal to $n{+}{+}$ since $n + 0 = n$. This closes the induction. □

- **Lemma 2.** For any natural numbers n and m, $n+(m{+}{+}) = (n+m){+}{+}$.

- Again, we cannot deduce this yet from $(n{+}{+})+m = (n+m){+}{+}$ because we do not know yet that $a + b = b + a$.

- **Proof.** We induct on n (keeping m fixed). We first consider the base case $n = 0$. In this case we have to prove $0 + (m{+}{+}) = (0 + m){+}{+}$. But by definition of addition, $0 + (m{+}{+}) = m{+}{+}$ and $0 + m = m$, so both sides are equal to $m{+}{+}$ and are thus equal to each other. Now we assume inductively that $n+(m{+}{+}) = (n+m){+}{+}$; we now have to show that $(n{+}{+}) + (m{+}{+}) = ((n{+}{+}) + m){+}{+}$. The left-hand side is $(n + (m{+}{+})){+}{+}$ by definition of addition, which is equal to $((n + m){+}{+}){+}{+}$ by the inductive hypothesis. Similarly, we have $(n{+}{+})+m = (n+m){+}{+}$ by the definition of addition, and so the right-hand side is also equal to $((n + m){+}{+}){+}{+}$. Thus both sides are equal to each other, and we have closed the induction. □

- As a particular corollary of Lemma 1 and Lemma 2 we see that $n{+}{+} = n + 1$ (why?).

- Finally, we can prove that $a + b = b + a$.

- **Proposition 3. (Addition is commutative)** For any natural numbers n and m, $n + m = m + n$.

- (A linguistic remark: From a logical point of view, there is no difference between a Lemma, Proposition, Theorem, or Corollary - they are all claims waiting to be proved. However, we use these terms to suggest different levels of importance and difficulty. A Lemma is an easily proved claim which is helpful for proving other Propositions and Theorems, but is not in itself particularly interesting. A Proposition is a statement which is interesting in its own right, while a Theorem is

a more important statement than a Proposition which says something definitive on the subject, and often takes more effort to prove than a Proposition or Lemma. A Corollary is a quick consequence of the previously proved Proposition or Theorem, which may or may not be important).

- **Proof.** We shall use induction on n (keeping m fixed). First we do the base case $n = 0$, i.e. we show $0 + m = m + 0$. By the definition of addition, $0 + m = m$, while by Lemma 1, $m + 0 = m$. Thus the base case is done. Now suppose inductively that $n + m = m + n$, now we have to prove that $(n{+}{+}) + m = m + (n{+}{+})$ to close the induction. By the definition of addition, $(n{+}{+}) + m = (n + m){+}{+}$. By Lemma 2, $m + (n{+}{+}) = (m + n){+}{+}$, but this is equal to $(n + m){+}{+}$ by the inductive hypothesis $n + m = m + n$. Thus $(n{+}{+}) + m = m + (n{+}{+})$ and we have closed the induction. □

- **Proposition 4. (Addition is associative)** For any natural numbers a, b, c, we have $(a + b) + c = a + (b + c)$.

- **Proof.** See Homework 1. □

- Because of this associativity we can write sums such as $a + b + c$ without having to worry about which order the numbers are being added together.

- Now we develop a cancellation law.

- **Proposition 5. (Cancellation law)** Let a, b, c be natural numbers such that $a + b = a + c$. Then we have $b = c$.

- Note that we cannot use subtraction or negative numbers yet to prove this Proposition, because we have not developed these concepts yet. In fact, this Cancellation law is crucial in letting us define subtraction (and the integers) later on in these notes, because it allows for a sort of "virtual subtraction" even before subtraction is officially defined.

- **Proof.** We prove this by induction on a. First consider the base case $a = 0$. Then we have $0 + b = 0 + c$, which by definition of addition implies that $b = c$ as desired. Now suppose inductively that we have

the cancellation law for a (so that $a+b = a+c$ implies $b = c$); we now have to prove the cancellation law for $a{+}{+}$. In other words, we assume that $(a{+}{+}) + b = (a{+}{+}) + c$ and imply that $b = c$. By the definition of addition, $(a{+}{+}) + b = (a+b){+}{+}$ and $(a{+}{+}) + c = (a+c){+}{+}$ and so we have $(a+b){+}{+} = (a+c){+}{+}$. By Axiom IV, we have $a+b = a+c$. Since we already have the cancellation law for a, we thus have $b = c$ as desired. This closes the induction. □

- We now discuss how addition interacts with positivity.

- **Definition** A natural number n is said to be *positive* iff it is not equal to 0. ("iff" is shorthand for "if and only if").

- **Proposition 6.** If a is positive and b is a natural number, then $a+b$ is positive (and hence $b+a$ is also, by Proposition 3).

- **Proof.** We use induction on b. If $b = 0$, then $a+b = a+0 = a$, which is positive, so this proves the base case. Now suppose inductively that $a+b$ is positive. Then $a+(b{+}{+}) = (a+b){+}{+}$, which cannot be zero by Axiom III, and is hence positive. This closes the induction. □

- **Corollary 7**. If a and b are natural numbers such that $a+b = 0$, then $a = 0$ and $b = 0$.

- **Proof.** Suppose for contradiction that $a \neq 0$ or $b \neq 0$. If $a \neq 0$ then a is positive, and hence $a+b = 0$ is positive by Proposition 6, contradiction. Similarly if $b \neq 0$ then b is positive, and again $a+b = 0$ is positive by Proposition 6, contradiction. Thus a and b must both be zero. □

- Once we have a notion of addition, we can begin defining a notion of *order*.

- **Definition** Let n and m be natural numbers. We say that n is *greater than or equal to* m, and write $n \geq m$ or $m \leq n$, iff we have $n = m + a$ for some natural number a. We say that n is *strictly greater than* m, and write $n > m$ or $m < n$, iff $n \geq m$ and $n \neq m$.

- Thus for instance $8 > 5$, because $8 = 5 + 3$ and $8 \neq 5$. Also note that $n{+}{+} > n$ for any n; thus there is no largest natural number n, because the next number $n{+}{+}$ is always larger still.

- **Proposition 8. (Basic properties of order)** Let a, b, c be natural numbers. Then $a \geq a$. Also, if $a \geq b$ and $b \geq c$, then $a \geq c$. If $a \geq b$ and $b \geq a$, then $a = b$. We have $a \geq b$ if and only if $a + c \geq b + c$, and we have $a < b$ if and only if $a{+}{+} \leq b$.

- **Proof.** See Homework 1. □

- From this Proposition it is easy to show that $n > m$ if and only if $n = m + a$ for some *positive* number a. (This can be used as an alternate definition for $n > m$).

- **Proposition 9. (Trichotomy of order)** Let a and b be natural numbers. Then exactly one of the following statements is true: $a < b$, $a = b$, or $a > b$.

- **Proof.** This is only a sketch of the proof; the gaps will be filled in Homework 1.

  First we show that we cannot have more than one of the statements $a < b$, $a = b$, $a > b$ holding at the same time. If $a < b$ then $a \neq b$ by definition, and if $a > b$ then $a \neq b$ by definition. If $a > b$ and $a < b$ then by Proposition 8 we have $a = b$, a contradiction. Thus no more than one of the statements is true.

  Now we show that at least one of the statements is true. We keep b fixed and induct on a. When $a = 0$ we have $0 \leq b$ for all b (why?), so we have either $0 = b$ or $0 < b$, which proves the base case. Now suppose we have proven the Proposition for a, and now we prove the proposition for $a{+}{+}$. From the trichotomy for a, there are three cases: $a < b$, $a = b$, and $a > b$. If $a > b$, then $a{+}{+} > b$ (why?). If $a = b$, then $a{+}{+} > b$ (why?). Now suppose that $a < b$. Then by Proposition 8, we have $a{+}{+} \leq b$. Thus either $a{+}{+} = b$ or $a{+}{+} < b$, and in either case we are done. This closes the induction. □

* * * * *

Multiplication

- In the previous section we have proven basically everything that we know to be true about addition and order. To save space and to avoid belaboring the obvious, we will now allow ourselves to use all the rules of algebra concerning addition and order that we are familiar with, without further comment. (Thus for instance we may write things like $a+b+c = c+b+a$ without supplying any further justification). Now we introduce multiplication. Just as addition is iterated incrementation, multiplication is incremented addition:

- **Definition**. Let m be a natural number. To multiply zero to m, we define $0 \times m := 0$. Now suppose inductively that we have defined how to multiply n to m. Then we can multiply $n{+}{+}$ to m by defining $(n{+}{+}) \times m := (n \times m) + m$.

- Thus for instance $0 \times m = 0$, $1 \times m = 0 + m$, $2 \times m = 0 + m + m$, etc. By induction one can easily verify that the product of two natural numbers is a natural number.

- By mimicking the proofs of Lemmas 1,2 and Proposition 3 one can easily show (see exercises) that $n \times 0 = 0$, $n \times (m{+}{+}) = n \times m + n$, and $n \times m = m \times n$ for all natural numbers n, m. Thus multiplication is commutative.

- We will now abbreviate $n \times m$ as nm, and use the usual convention that multiplication takes precedence over addition, thus for instance $ab + c$ means $(a \times b) + c$, not $a \times (b + c)$. (We will also use the usual notational conventions of precedence for the other arithmetic operations when they are defined later, to save on using parentheses all the time). By using Proposition 6 and induction, one can show that if a and b are positive, then ab is also positive; this implies that if $ab = 0$, then one has either $a = 0$ or $b = 0$ (or both). We leave this to the exercises.

- **Proposition 10. (Distributive law)** For any natural numbers a, b, c, we have $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$.

- **Proof.** Since multiplication is commutative we only need to show the first identity $a(b + c) = ab + ac$. We keep a and b fixed, and use induction on c. Let's prove the base case $c = 0$, i.e. $a(b + 0) = ab + a0$. The left-hand side is ab, while the right-hand side is $ab + 0 = ab$, so

we are done with the base case. Now let us suppose inductively that $a(b+c) = ab + ac$, and let us prove that $a(b + (c{+}{+})) = ab + a(c{+}{+})$. The left-hand side is $a((b+c){+}{+}) = a(b+c) + a$, while the right-hand side is $ab + ac + a = a(b+c) + a$ by the induction hypothesis, and so we can close the induction. □

- Using the distributive law, one can then mimic the proof of Proposition 4 to show that multiplication is associative: $(a \times b) \times c = a \times (b \times c)$. Again, we leave this to the exercises.

- **Proposition 11.** If a, b are natural numbers such that $a < b$, and c is positive, then $ac < bc$.

- **Proof.** Since $a < b$, we have $b = a + d$ for some positive d. Multiplying by c and using the distributive law we obtain $bc = ac + dc$. Since d is positive, and c is non-zero (hence positive), dc is positive, and hence $ac < bc$ as desired. □

- **Corollary 12. (Cancellation law)** Let a, b, c be natural numbers such that $ac = bc$ and c is non-zero. Then $a = b$.

- Just as Proposition 5 will allow for a "virtual subtraction" which will eventually let us define genuine subtraction, this Corollary provides a "virtual division" which will be needed to define genuine division later on.

- **Proof.** By the trichotomy of order, we have three cases: $a < b$, $a = b$, $a > b$. Suppose first that $a < b$, then by Proposition 11 we have $ac < bc$, a contradiction. We can obtain a similar contradiction when $a > b$. Thus the only possibility is that $a = b$, as desired. □

- With these propositions it is easy to deduce all the familiar rules of algebra involving addition and multiplication; for instance, you may check that if we define $n^2 := n \times n$, that you now have enough tools to show that $(a+b)^2 = a^2 + 2ab + b^2$.

- Now that we have the familiar operations of addition and multiplication, the more primitive notion of incrementation will begin to fall by the wayside, and we will see it rarely from now on. In any event we can always use addition to describe incrementation, since $n{+}{+} = n + 1$.

- **Proposition 13. (Euclidean algorithm)** Let n be a natural number, and let q be a positive number. Then there exists natural numbers m, r such that $0 \leq r < q$ and $n = mq + r$.

- In other words, we can divide a natural number n by a positive number q to obtain a quotient m (which is another natural number) and a remainder r (which is less than q). This algorithm marks the beginning of *number theory*, which is a beautiful and important subject but one which is beyond the scope of this course.

- **Proof.** See Homework 1. □

* * * * *

The integers

- We have now built up most of the basic properties of the natural number system, but are reaching the limits of what one can do with just addition and multiplication. We would now like to introduce a new operation, that of subtraction, but to do that properly we will have to pass from the natural number system to a larger number system, that of the *integers*.

- Informally, the integers are what you can get by subtracting two natural numbers; for instance, $3-5$ should be an integer, as should $6-2$. This is not a complete definition of the integers, because (a) it doesn't say when two differences are equal (for instance we should know why $3-5$ is equal to $2-4$, but is not equal to $1-6$), and (b) it doesn't say how to do arithmetic on these differences (how does one add $3-5$ to $6-2$?). Furthermore, (c) this definition is circular because it requires a notion of subtraction, which we can only adequately define once the integers are constructed. Fortunately, because of our prior experience with integers we know what the answers to these questions should be. To answer (a), we know from our advanced knowledge in algebra that $a-b=c-d$ happens exactly when $a+d=c+b$, so we can characterize equality of differences using only the concept of addition. Similarly, to answer (b) we know from algebra that $(a-b)+(c-d) = (a+c)-(b+d)$ and that $(a-b)(c-d) = (ac+bd)-(ad+bc)$. So we will take advantage of our foreknowledge by building all this into the *definition* of the integers, as we shall do shortly.

- We still have to resolve (c). To get around this problem we will use the following workaround: we will temporarily write integers not as a difference $a-b$, but instead use a new notation $a—b$ to define integers, where the — is a meaningless place-holder (similar to the comma in the Cartesian co-ordinate notation (x, y) for points in the plane). Later when we define subtraction we will see that $a—b$ is in fact equal to $a - b$, and so we can discard the notation —; it is only needed right now to avoid circularity. (These devices are similar to the scaffolding used to construct a building; they are temporarily essential to make sure the building is built correctly, but once the building is completed they are thrown away and never used again). This may seem unnecessarily complicated in order to define something that we already are very familiar with, but we will use this device again to construct the rationals, and knowing these kinds of constructions will be very helpful in later weeks.

- **Definition** An *integer* is an expression of the form $a—b$, where a and b are natural numbers. Two integers are considered to be equal, $a—b = c—d$, if and only if $a + d = c + b$. We let $\mathbf{Z}$ denote the set of all integers.

- Thus for instance $3—5$ is an integer, and is equal to $2—4$, because $3+4 = 2+5$. On the other hand, $3—5$ is not equal to $2—3$ because $3+3 \neq 2+5$. (This notation is strange looking, and has a few deficiencies; for instance, 3 is not yet an integer, because it is not of the form $a—b$! We will rectify these problems later).

- We have to check that this is a legitimate notion of equality. There are three axioms that equality must satisfy: (a) x is always equal to x; (b) if x is equal to y, then y is equal to x; and (c) if $x = y$ and $y = z$, then $x = z$. (There is actually a fourth: (d) if $x = y$, then one has $f(x) = f(y)$ for any function or operation f, but we will get to that later. If you think about it, whenever you ever use the fact that two objects are equal you are using one of the above four axioms). The first two are easy to check and are left to the reader; let us verify the third. Suppose we know that $a—b = c—d$ and $c—d = e—f$. Then we have $a + d = c + b$ and $c + f = d + e$. Adding the two equations together we obtain $a + d + c + f = c + b + d + e$. By Proposition 5 we

can cancel the c and d, obtaining $a+f=b+e$, i.e. $a{--}b=e{--}f$. Thus the cancellation law was needed to make sure that our notion of equality is sound.

- Now we define basic arithmetic operations on integers: addition and multiplication. (Incrementation is not so important for the integers, and we will remark on it later).

- **Definition** The sum of two integers, $(a{—}b)+(c{—}d)$, is defined by the formula

$$(a{—}b)+(c{—}d) := (a+c){—}(b+d).$$

The product of two integers, $(a{—}b)\times(c{—}d)$, is defined by

$$(a{—}b)\times(c{—}d) := (ac+bd){—}(ad+bc).$$

- Thus for instance, $(3{—}5)+(1{—}4)$ is equal to $(4{—}9)$. There is however one thing we have to check before we can accept these definitions - we have to check that if we replace one of the integers by an equal integer, that the sum or product does not change. For instance, $(3{—}5)$ is equal to $(2{—}4)$, so $(3{--}5)+(1{—}4)$ ought to have the same value as $(2{--}4)+(1{—}4)$, otherwise this would not give a consistent definition of addition. Fortunately, this is the case:

- **Lemma 14.** Let a,b,a',b',c,d be natural numbers. If $(a{—}b)=(a'{--}b')$, then $(a{—}b)+(c{—}d)=(a'{—}b')+(c{—}d)$ and $(a{—}b)\times(c{—}d)=(a'{—}b')\times(c{—}d)$, and also $(c{—}d)+(a{—}b)=(c{—}d)+(a'{--}b')$ and $(c{—}d)\times(a{—}b)=(c{--}d)\times(a'{--}b')$. Thus addition and multiplication are well-defined operations (equal inputs give equal outputs).

- **Proof.** To prove that $(a{—}b)+(c{—}d)=(a'{—}b')+(c{—}d)$, we evaluate both sides as $(a+c){—}(b+d)$ and $(a'+c){—}(b'+d)$. Thus we need to show that $a+c+b'+d=a'+c+b+d$. But since $(a{—}b)=(a'{--}b')$, we have $a+b'=a'+b$, and so by adding $c+d$ to both sides we obtain the claim. Now we show that $(a{--}b)\times(c{—}d)=(a'{--}b')\times(c{—}d)$. Both sides evaluate to $(ac+bd){—}(ad+bc)$ and $(a'c+b'd){--}(a'd+b'c)$, so we have to show that $ac+bd+a'd+b'c=$

$a'c+b'd+ad+bc$. But the left-hand side factors as $c(a+b')+d(a'+b)$, while the right factors as $c(a'+b)+d(a+b')$. Since $a+b'=a'+b$, the two sides are equal. The other two identities are proven similarly. □

- The integers n——0 behave in the same way as the natural numbers n; indeed one can check that $(n——0)+(m——0)=(n+m)——0$ and $(n——0)\times(m——0)=nm——0$. Furthermore, $(n——0)$ is equal to $(m——0)$ if and only if $n=m$. (The mathematical term for this is that there is an *isomorphism* between the natural numbers n and those integers of the form $n——0$). Thus we may *identify* the natural numbers with integers by setting $n \equiv n——0$; this does not affect our definitions of addition or multiplication or equality since they are consistent with each other. Thus for instance the natural number 3 is now considered to be the same as the integer $3——0$: $3=3——0$. In particular 0 is equal to $0——0$ and 1 is equal to $1——0$. Of course, if we set n equal to $n——0$, then it will also be equal to any other integer which is equal to $n——0$, for instance 3 is equal not only to $3——0$, but also to $4——1$, $5——2$, etc.

- We can now define incrementation on the integers by defining $x++ := x+1$ for any integer x; this is of course consistent with our definition of incrementation of natural numbers.

- We define a new operation on the integers: negation. If $(a——b)$ is an integer, we define the negation $-(a——b)$ to be the integer $(b——a)$. For instance $-(3——5)=(5——3)$. It is easy to see that this definition is well-defined in the sense that if $(a——b)=(a'——b')$, then $-(a——b)=-(a'——b')$ (so equal integers have equal negations). In particular if $n=n——0$ is a positive natural number, we can define its negation $-n=0——n$.

- We can now show that the integers correspond exactly to what we expect.

- **Lemma 15.** Let x be an integer. Then exactly one of the following three statements is true: (a) x is zero; (b) x is equal to a positive natural number n; or (c) x is the negation $-n$ of a positive natural number n.

- If n is a positive natural number, we call $-n$ a *negative integer*. Thus every integer is positive, zero, or negative, but not more than one of these at a time.

- **Proof.** We first show that at least one of (a), (b), (c) is true. By definition, $x = a—b$ for some natural numbers a, b. We have three cases: $a > b$, $a = b$, or $a < b$. If $a > b$ then $a = b + c$ for some positive natural number c, which means that $a—b = c—0 = c$, which is (a). If $a = b$, then $a—b = a—a = 0—0 = 0$, which is (b). If $a < b$, then $b > a$, so that $b—a = n$ for some natural number n by the previous reasoning, and thus $a—b = -n$, which is (c).

  Now we show that no more than one of (a), (b), (c) can hold at a time. By definition, a positive natural number is non-zero, so (a) and (b) cannot simultaneously be true. If (a) and (c) were simultaneously true, then $0 = -n$ for some positive natural n; thus $(0—0) = (0—n)$, so that $0 + n = 0 + 0$, so that $n = 0$, a contradiction. If (b) and (c) were simultaneously true, then $n = -m$ for some positive n, m, so that $(n—0) = (0—m)$, so that $n + m = 0 + 0$, which contradicts Proposition 6. Thus exactly one of (a), (b), (c) is true for any integer x. □

- One could well ask why we don't use Lemma 15 to *define* the integers; i.e. why didn't we just say an integer is anything which is either a positive natural number, zero, or the negative of a natural number. The reason is that if we did so, the rules for adding and multiplying integers would split into many different cases (e.g. negative times positive equals positive; negative plus positive is either negative, positive, or zero, depending on which term is larger, etc.) and to verify all the properties ends up being much messier than doing it this way.

- We now summarize the algebraic properties of the integers.

- **Proposition 16.** Let x, y, z be integers. Then the following laws of

algebra hold:

$$\begin{aligned} x+y &= y+x \\ (x+y)+z &= x+(y+z) \\ x+0=0+x &= x \\ x+(-x)=(-x)+x &= 0 \\ xy &= yx \\ (xy)z &= x(yz) \\ x1=1x &= x \\ x(y+z) &= xy+xz \\ (y+z)x &= yx+zx. \end{aligned}$$

- The above set of nine identities have a name; they are asserting that the integers form a *commutative ring*. (If one deleted the identity $xy = yx$, then they would only assert that the integers form a *ring*). Note that some of these identities were already proven for the natural numbers, but this does not automatically mean that they also hold for the integers because the integers are a larger set than the natural numbers.

- **Proof.** There are two ways to prove these identities. One is to use Lemma 15 and split into a lot of cases depending on whether x, y, z are zero, positive, or negative. This becomes very messy. A shorter way is to write $x = (a{-}{-}b)$, $y = (c{-}{-}d)$, and $z = (e{-}{-}f)$ for some natural numbers a, b, c, d, e, f, and expand these identities in terms of a, b, c, d, e, f and use the algebra of the natural numbers. This allows each identity to be proven in a few lines. We shall just prove the longest one, namely $(xy)z = x(yz)$:

$$\begin{aligned} (xy)z &= ((a{-}b)(c{-}d))(e{-}f) = ((ac+bd){-}(ad+bc))(e{-}f) \\ &= ((ace+bde+adf+bcf){-}(acf+bdf+ade+bce)) \\ x(yz) &= (a{-}{-}b)((c{-}d)(e{-}f)) = (a{-}b)((ce+df){-}{-}(cf+de)) \\ &= ((ace+adf+bcf+bde){-}(acf+ade+bcd+bdf)) \end{aligned}$$

and so one can see that $(xy)z$ and $x(yz)$ are equal. The other identities are proven in a similar fashion and are left to the reader. (Note that one can save some work by using some identities to prove others. For

instance, once you know that $xy = yx$, you get for free that $x1 = 1x$, and once you also prove $x(y+z) = xy + xz$, you automatically get $(y+z)x = yx + zx$ for free). □

- We now define the operation of *subtraction* $x - y$ of two integers by the formula

$$x - y := x + (-y).$$

One can easily check now that if a and b are natural numbers, then

$$a - b = a + -b = (a\text{—}0) + (0\text{—}b) = a\text{—}\text{—}b,$$

and so $a\text{—}b$ is just the same thing as $a - b$. Because of this we can now discard the — notation, and use the familiar operation of subtraction instead. (As remarked before, we could not use subtraction immediately because it would be circular).

- A basic property of the integers is that they contain no zero divisors:

- **Proposition 17.** Let a and b be integers such that $ab = 0$. Then either $a = 0$ or $b = 0$ (or both).

- **Proof.** See Homework 1. □

- As an easy corollary of this proposition we have the cancellation law: if a, b, c are integers such that $ac = bc$ and c is non-zero, then $a = b$. To see this, note from $ac = bc$ that $(a-b)c = ac - bc = 0$; since c is non-zero, we see from Proposition 17 that $a - b$ must be zero, i.e. $a = b$. (An alternate way to prove this cancellation law comes from combining Corollary 12 with Lemma 15).

- We now extend the notion of order, which was defined on the natural numbers, to the integers by repeating the definition verbatim:

- **Definition** Let n and m be integers. We say that n is *greater than or equal to* m, and write $n \geq m$ or $m \leq n$, iff we have $n = m + a$ for some natural number a. We say that n is *strictly greater than* m, and write $n > m$ or $m < n$, iff $n \geq m$ and $n \neq m$.

- Thus for instance $5 > -3$, because $5 = -3+8$ and $5 \neq -3$. Clearly this definition is consistent with the notion of order on the natural numbers, since we are using the same definition.

- Using the laws of algebra in Proposition 16 it is not hard to show the following properties of order:

- **Lemma 18.** If a and b are integers, then $a > b$ if and only if $a - b$ is a positive natural number. If $a > b$, then $a + c > b + c$ for any integer c. If $a > b$, then $ac > bc$ for any positive natural number c. If $a > b$, then $-a < -b$. If $a > b$ and $b > c$, then $a > c$. If $a \geq b$ and $b \geq a$, then $a = b$.

- **Proof.** See Homework 1. □

- One final warning about the integers: the principle of induction (Axiom V) does not apply directly to the integers: if you want to prove that a property $P(n)$ is true for all integers n, it is not enough to first prove that $P(0)$ and verify that $P(n)$ implies $P(n{+}{+})$ for all n. This is enough to obtain $P(n)$ for all natural numbers n, but not all integers n. Thus induction is not as useful a tool for dealing with the integers as it is with the natural numbers. (The situation becomes even worse with the rationals and reals).

* * * * *

The rationals

- We have now constructed the integers, with the operations of addition, subtraction, multiplication, and order and verified all the expected algebraic and order-theoretic properties. Now we will make a similar construction to build the rationals, adding division to our mix of operations.

- Just like the integers were constructed by subtracting two natural numbers, the rationals can be constructed by dividing two integers, though of course we have to make the usual caveat that the denominator should be non-zero. (There is no reasonable way we can divide by zero, since one cannot have both the identities $(a/b) * b = a$ and $c * 0 = 0$ hold simultaneously if b is allowed to be zero. However, we can eventually

get a reasonable notion of dividing by a quantity which *approaches* zero - think of l'Hôpital's rule, which suffices for doing things like defining differentiation). Of course, just as two differences $a-b$ and $c-d$ can be equal if $a+d=c+b$, we know (from more advanced knowledge) that two quotients a/b and c/d can be equal if $ad=bc$. Thus, in analogy with the integers, we create a new meaningless symbol // (which will eventually be superceded by division), and define

- **Definition** A *rational number* is an expression of the form $a//b$, where a and b are integers and b is non-zero; $a//0$ is not considered to be a rational number. Two rational numbers are considered to be equal, $a//b = c//d$, if and only if $ad = cb$. The set of all rational numbers is denoted $\mathbf{Q}$.

- Thus for instance $3//4 = 6//8 = -3//-4$, but $3//4 \neq 4//3$. Again, before we accept this definition of equality we have to verify the three axioms of equality: $x=x$; if $x=y$ then $y=x$; and if $x=y$ and $y=z$, then $x=z$. The first two are easy and are left to the reader. To verify the third, suppose that $a//b = c//d$ and $c//d = e//f$. Then $ad=bc$ and $cf=ed$. Multiplying these two together, we obtain $adcf=bced$. Applying Corollary 12 twice to cancel the *non-zero* factors c, d, we obtain $af=be$, i.e. $a//b=e//f$, as desired.

- Now we need a notion of addition, multiplication, and negation. Again, we will take advantage of our pre-existing knowledge, which tells us that $a/b+c/d$ should equal $(ad+bc)/(bd)$ and that $a/b*c/d$ should equal ac/bd, while $-(a/b)$ equals $(-a)/b$. Motivated by this foreknowledge, we define

- **Definition** If $a//b$ and $c//d$ are rational numbers, we define their sum

$$(a//b)+(c//d) := (ad+bc)//(bd)$$

their product

$$(a//b)*(c//d) := (ac)//(bd)$$

and the negation

$$-(a//b) := (-a)//b.$$

- Note that if b and d are non-zero, then bd is also non-zero, by Proposition 17, so the sum or product of a rational number remains a rational number.

- **Lemma 19.** The above operations are well-defined, in the sense that if one replaces $a//b$ with another rational number $a'//b'$ which is equal to $a//b$, then the output of the above operations remains unchanged, and similarly for $c//d$.

- **Proof** We just verify this for addition and $a//b$; the other ones are similar. Suppose $a//b = a'//b'$, so that b and b' are non-zero and $ab' = a'b$. We now show that $a//b + c//d = a'//b' + c//d$. By definition, the left-hand side is $(ad+bc)//bd$ and the right-hand side is $(a'd+b'c)//b'd$, so we have to show that

$$(ad + bc)b'd = (a'd + b'c)bd,$$

which expands to

$$ab'd^2 + bb'cd = a'bd^2 + bb'cd.$$

But since $ab' = a'b$, the claim follows. □

- We note that the rational numbers $a//1$ behave in a manner identical to the integers a:

$$(a//1)+(b//1) = (a+b)//1; \quad (a//1)\times(b//1) = (ab//1); \quad -(a//1) = (-a)//1.$$

Also, $a//1$ and $b//1$ are only equal when a and b are equal. Because of this, we will identify a with $a//1$ for each integer a: $a \equiv a//1$; the above identities then guarantee that the arithmetic of the integers is consistent with the arithmetic of the rationals. Thus just as we embedded the natural numbers inside the integers, we embed the integers inside the rational numbers. In particular, all natural numbers are rational numbers, for instance 0 is equal to $0//1$ and 1 is equal to $1//1$.

- Observe that a rational number $a//b$ is equal to $0 = 0//1$ if and only if $a \times 1 = b \times 0$, i.e. if the numerator a is equal to 0. Thus if a and b are non-zero then so is $a//b$.

- We now define a new operation on the rationals: reciprocal. If $x = a//b$ is a non-zero rational (so that $a, b \neq 0$) then we define x^{-1} to be $x^{-1} := b//a$. It is easy to check that this operation is consistent with our notion of equality: if two rational numbers $a//b$, $a'//b'$ are equal, then their reciprocals are also equal. (In contrast, an operation such as "numerator" is not well-defined: the rationals $3//4$ and $6//8$ are equal, but have unequal numerators, so we have to be careful when referring to such terms as "the numerator of x". The purpose of all these consistency checks is to ensure that we don't have to have any similar worries when doing things like adding or taking reciprocals of rational numbers). We leave the reciprocal of 0 undefined.

- We now summarize the algebraic properties of the rationals.

- **Proposition 20.** Let x, y, z be rationals. Then the following laws of algebra hold:

$$\begin{aligned} x + y &= y + x \\ (x + y) + z &= x + (y + z) \\ x + 0 = 0 + x &= x \\ x + (-x) = (-x) + x &= 0 \\ xy &= yx \\ (xy)z &= x(yz) \\ x1 = 1x &= x \\ x(y + z) &= xy + xz \\ (y + z)x &= yx + zx. \end{aligned}$$

  If x is non-zero, we also have

$$xx^{-1} = x^{-1}x = 1.$$

- The above set of ten identities have a name; they are asserting that the rationals $\mathbf{Q}$ form a *field*. This is better than being a commutative ring because of the tenth identity.

- **Proof.** The proof of this proposition is somewhat tedious: one writes $x = a//b$, $y = c//d$, $z = e//f$ for some integers a, c, e and non-zero integers b, d, f, and verifies each identity in turn using the algebra of

the integers. We shall just prove the longest one, namely $(x+y)+z = x+(y+z)$:

$$(x+y)+z = ((a//b)+(c//d))+(e//f) = ((ad+bc)//bd)+(e//f)$$
$$= (adf+bcf+bde)//bdf$$
$$x+(y+z) = (a//b)+((c//d)+(e//f)) = (a//b)+((cf+de)//df)$$
$$= (adf+bcf+bde)//bdf$$

and so one can see that $(x+y)+z$ and $x+(y+z)$ are equal. The other identities are proven in a similar fashion and are left to the reader. (As with Proposition 16, you can save some work by using some identities to prove others.) □

- We can now define the *quotient* x/y of two rational numbers x and y, *provided that* y is non-zero, by the formula

$$x/y := x \times y^{-1}.$$

Thus, for instance

$$(3//4)/(5//6) = (3//4) \times (6//5) = (18//20) = (9//10).$$

Using this formula, it is easy to see that $a/b = a//b$ for every integer a and every non-zero integer b. Thus we can now discard the // notation, and use the more customary a/b instead of $a//b$.

- The above field axioms allow us to use all the normal rules of algebra; we will now proceed to do so without further comment.

- In the previous section we organized the integers into positive, zero, and negative numbers. We now do the same for the rationals.

- **Definition** A rational number x is said to be *positive* iff we have $x = a/b$ for some positive integers a and b. It is said to be *negative* iff we have $x = -y$ for some positive rational y (i.e. $x = (-a)/b$ for some positive integers a and b).

- Thus for instance, every positive integer is a positive rational number, and every negative integer is a negative rational number, so our new definition is consistent with our old one.

- **Lemma 21.** Let x be a rational number. Then exactly one of the following three statements is true: (a) x is equal to 0. (b) x is a positive rational number. (c) x is a negative rational number.
- **Proof.** See Homework 1. □
- We now define the notion of order on rationals.
- **Definition.** Let x and y be rational numbers. We say that $x > y$ iff $x - y$ is a positive rational number, and $x < y$ iff $x - y$ is a negative rational number. We write $x \geq y$ iff either $x > y$ or $x = y$, and similarly define $x \leq y$.
- The following properties of order are easily verified:
- **Proposition 22.** Let x, y, z be rational numbers. Then the following properties hold.
- (a) Exactly one of the three statements $x = y$, $x < y$, or $x > y$ is true.
- (b) One has $x < y$ if and only if $y > x$.
- (c) If $x < y$ and $y < z$, then $x < z$.
- (d) If $x < y$, then $x + z < y + z$.
- (e) If $x < y$ and z is positive, then $xz < yz$.
- **Proof.** See Homework 1. □
- The above five properties in Proposition 22, combined with the field axioms in Proposition 20, have a name: they assert that the rationals **Q** form an *ordered field*.
- A basic property of **Q** is that they are "dense" in the following sense: given any two rationals, there is always a third between them.
- **Proposition 23.** Given any two rationals x and y such that $x < y$, there exists a third rational z such that $x < z < y$.

- **Proof.** We set $z := (x+y)/2$. Since $x < y$, and $1/2 = 1//2$ is positive, we have from Proposition 22(e) that $x/2 < y/2$. If we add $y/2$ to both sides using Proposition 22(d) we obtain $x/2+y/2 < y/2+y/2$, i.e. $z < y$. If we instead add $x/2$ to both sides we obtain $x/2+x/2 < y/2+x/2$, i.e. $x < z$. Thus $x < z < y$ as desired. □

- Despite the rationals having this denseness property, they are still incomplete; there are still an infinite number of "gaps" or "holes" between the rationals, although this denseness property does ensure that these holes are in some sense infinitely small. To plug these holes we need one last extension of our number system, from the rationals $\mathbf{Q}$ to the real numbers $\mathbf{R}$. This construction will occupy us in the next week's notes.