Notes on Number Theory, Fall 2022

Randall Holmes

October 6, 2022

Im posting these on Monday 10/3; I still need to add more to cover the lecture the previous Monday. The notes on that lecture should be complete sometime today, to be expanded to cover what I lecture today (10/3). The homework assigned for 10/3 will also appear as an update to this file.

1 Basic concepts and axioms

The mathematical system we are interested in is officially the set of integers, whose official name is $\mathbb{Z}$ (this stands for **Zahlen**, numbers, in German).

I'm going to state a definition of the set of integers as a subset of the real numbers $\mathbb{R}$, just because I would like you to know in the back of your mind that things like this can be done. This definition is not examinable, but it might in some ways be useful to understand it.

Definition (integer-closed set of reals): A set A of real numbers is said to be "integer-closed" if and only if $0 \in A$ and for every real number x, if $x \in A$ then $x + 1$ and $x - 1$ are both in A.

There are lots of integer-closed subsets of $\mathbb{R}$: $\mathbb{R}$ itself is integer-closed; the set of rationals is integer-closed; the set of all fractions with denominator 2 is integer closed (it contains all integers n and also contains $n + \frac{1}{2}$ for each integer n).

Definition of $\mathbb{Z}$: $\mathbb{Z}$ is defined as $\{x \in \mathbb{R} :$ for every set A, if A is integer closed then $x \in A\}$. It really works...think about it.

There is something odd about this definition, because it presupposes that we know what the reals are. But if you get to an advanced course

where a formal definition of the natural numbers, integers, and reals is actually given, you might get an idea of what I'm up to. Really, I just want to be able to say something more accurate than $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2\ldots\}$.

The mathematical system of the integers does not have the set of integers as its only component. It also has operations and relations for which we will state axioms. These operations and relations are familiar. The constants 0 and 1, the operations of addition, multiplication, and additive inverse, and the relation "less than" are the basics in our presentation.

First set of axioms (basic algebra): commutative laws: For any integers a, b, $a + b = b + a$ and $a \cdot b = b \cdot a$.

associative laws: For any integers a, b, c, $(a+b)+c = a+(b+c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

distributive law: For any integers a, b, c, $a(b + c) = a \cdot b + a \cdot c$.

distributive laws? (comment): We could give another distributive law $(a+b)c = ac+bc$ too, but we do not need to, since $(a+b)c = c(a+b) = ca+cb = ac+bc$ justifies the alternative distributive law in terms of the one we give and commutativity of multiplication.

It is useful to notice that matrix algebra is a familiar system (to some of you, at least) in which both forms of the distributive law have to be given, because matrix multiplication is not commutative.

identity laws: For any integer a $a + 0 = a$ and $a \cdot 1 = a$; $0 \neq 1$

additive inverse law: For any integer a, $a + (-a) = 0$.

multiplicative cancellation: For any integers a, b, c, If $c \neq 0$ and $ac = bc$, then $a = b$.

Normally, we would prove multiplicative cancellation (in the reals or rationals) using the existence of a multiplicative inverse (reciprocal) a^{-1} of each nonzero real number a such that $a \cdot a^{-1} = 1$. But this property doesn't hold in the integers. Notice that we can prove additive cancellation using the additive inverse property (if $a + c = b + c$ then $a = b$: try to demonstrate this).

Observation: All of these axioms hold in the systems of reals or rationals familiar to you. We will also study systems of modular arithmetic which

look very different from the reals or rationals (they are finite systems!) in which all of the axioms of the first set hold.

A second set of axioms rules out the modular arithmetic systems.

Second set of axioms (properties of order): $<$ is transitive: For all integers a, b, c, if $a < b$ and $b < c$, then $a < c$.

trichotomy: For any integers a, b, exactly one of the following is true: $a < b, a = b, b < a$.

additive monotonicity: For any integers a, b, c, if $a < b$ then $a + c < b + c$ (it is actually if and only if, because you can add the additive inverse to go the other way).

multiplicative monotonicity: For any integers a, b, c, if $c > 0$ and $a < b$, then $ac < bc$.

The other version, if $c < 0$ and $a < b$, then $bc < ac$, is provable from the axioms we already have.

Definitions: $a > b$ means $b < a$. $a \leq b$ means $a < b$ or $a = b$. $a \geq b$ means $b \leq a$. These relations all have similar properties to $<$, which you are already informally familiar with.

Observation: These axioms all hold in the familiar systems of reals and rationals. The axiom which separates the integers from the reals or rationals is something we have already discussed.

Axiom of Mathematical Induction: Let $P(n)$ be a statement about an integer variable n and let b be an integer. If $P(b)$ is true and for every integer $k \geq b$ such that $P(k)$ is true, it follows that $P(k + 1)$ is true, we can conclude that $P(n)$ is true for every integer $n \geq b$.

Proof structure: A proof of "for all integers $n \geq b$, $P(n)$ by math induction falls into two parts:

the basis steps shows $P(b)$

the induction step introduces an arbitrarily chosen integer k and the assumption that $P(k)$ is true [called the inductive hypothesis] and shows that $P(k + 1)$ follows.

A variation is proof by strong induction, in which the inductive hypothesis is strengthened to "for all integers m with $b \leq m \leq k$, $P(m)$.

The axiom of mathematical induction drives a wedge between the integers and the reals or rationals. For example, we can prove that there is no multiplicative inverse of 2.

Theorem: There is no integer x such that $0 < x < 1$.

Proof: We prove by mathematical induction that for every integer $x \geq 0$, either $x = 0$ or $x \geq 1$.

Basis: $0 = 0$ or $0 \geq 1$ is true because $0 = 0$.

Induction: Suppose $k = 0$ or $k \geq 1$. In either case we have $k \geq 0$. $k+1 \geq 1$ follows by additive monotonicity, so we have either $k+1 = 0$ or $k + 1 \geq 1$ because we have the latter.

Theorem: There is no integer x such that $2x = 1$.

Proof: Suppose there is such an x. We have to establish $x > 0$: clearly x is not 0 ($2 \cdot 0 = 0 \neq 1$). If $x < 0$ we would have $2x < 0$ by multiplicative monotonocity, so $1 < 0$, which is absurd. So $x > 0$. It follows by additive monotonicity that $x + x > x$, so $1 > x$. But then $0 < x < 1$, which we have just shown is not possible.

Sanity check ($1 > 0$): We have in our axioms that $1 \neq 0$. Thus we can only have $1 > 0$ (which we believe) or $1 < 0$. Suppose that $1 < 0$. Add -1 to both sides to get $0 < -1$, so $-1 > 0$. Then we can use $-1 > 0$ and multiplicative monotonicity: it follows that $(-1)(-1) > 0$, $1 > 0$, contradicting the assumption that $1 < 0$ (by an appeal to trichotomy). So we have ruled out $1 < 0$ and $1 > 0$ is the only possibility.

Again, Im appealing to common sense in the equation $(-1)(-1) = 1$. We may look into how to prove that.

This is an example of a general point. You have seen something like these axioms before, and you have been told that all of your ninth-grade algebra knowledge (say) follows from these principles. But you weren't really shown this, and proofs of "easy and obvious" things from basic sets of axioms in any area of math may turn out to be tricky.

Challenge Problem: Prove from the axioms that for any integer a, $a \cdot 0 = 0$. It is rather tricky! (Notice that we used this fact freely in our discussion above.)

2 Divisibility and a little about prime numbers

Definition (divisibility): Let a and b be integers.

We define $a|b$ (a goes into b, or equivalently b is divisible by a) as meaning "There is an integer k such that $ak = b$".

We also say "a is a factor of b", or "a is a divisor of b". We usually restrict our attention to positive divisors, but we will try always to say this explicitly.

Important Observation: Please notice that $a|b$ is not a fraction or any kind of expression: it is a sentence. And notice that writing $\frac{b}{a}$ is not a way of saying b is divisible by a: I saw this on many test papers.

What is (almost) true is that $a|b$ is equivalent to "$\frac{b}{a}$ is an integer". The exception is when a and b are both 0: $0|0$ is true but $\frac{0}{0}$ is undefined.

We will do some work on basic theorems about divisiblity, notes on which will be inserted at this point. The facts about divisibility which we will prove are fairly obvious, but writing the proofs will be useful for general thinking about how proofs are to be written.

We define a basic notion already familiar to you.

Definition (prime numbers): An integer n is prime iff $n > 1$ and the only positive divisors of n are 1 and n.

Note that 1 is not a prime.

Theorem: Each integer $n \geq 2$ is a prime or a finite product of primes.

Proof: We prove this by strong induction:

Basis ($n = 2$): 2 is a prime.

Induction step: Let k be an arbitrarily chosen integer. Suppose (ind hyp) that every m with $2 \leq m \leq k$ is a prime or a finite product of primes.

Our goal is then to show that $k + 1$ is a prime or finite product of primes.

If $k + 1$ is prime, we are done.

If $k+1$ is not prime, then $k+1 = LM$ for some $2 < L \leq M < k+1$ (it has positive divisors other than 1 and itself). But by inductive hypothesis each of L, M is either a prime or a finite product of primes, so $LM = k+1$ is a finite product of primes, and we are done.

There is a bigger result which we will prove soon: each integer ≥ 2 factors into primes in exactly one way.

The theorem which follows is ancient, and it has been seriously argued that any educated person should know it. Certainly its proof is examinable in this class.

Theorem (Euclid): There are infinitely many prime numbers.

Proof: We have at least one prime, 2. Suppose we have a list of n prime numbers $p_1, p_2 \ldots, p_n$ which contains all of the prime numbers (if there were finitely many primes we could make such a list).

Consider $P = 1 + \prod_{i=1}^{n} p_i$, one plus the product of all of the primes on our list.

By the preceding theorem, P (which is ≥ 2 because the list has 2 in it) has a prime divisor q, because it is either a prime or a product of primes.

Iff $q = p_m$ then q goes into $P = 1 + \prod_{i=0}^{n} p_i$, and $q|(P-1) = \prod_{i=0}^{n} p_i$ because it is one of the factors in the product, so $q|P - (P-1) = 1$, which is absurd. So q cannot be one of the primes we listed, so there cannot be a finite list of all the primes, which is what we wanted to prove.

3 The Division Algorithm: integer division and remainder operations

Theorem (division algorithm): For any integer a and any integer $b > 0$, there are uniquely determined integers q and r such that $a = bq + r$ and $0 \leq r < 1$.

integer division and remainder operations: Because q and r are uniquely determined, we can define a `div` b as q and a `mod` b as r. The div operation is integer division and the mod operation is remainder. These operations should actually be familiar from elementary school.

Proof of the Theorem: This falls into three sections:

induction proof of existence of q and r (not uniqueness) for $a \geq 0$:

We prove the theorem (actually, just part of it to begin with) by induction on a.

We let b be a fixed positive integer.

We prove by induction that for each $n \geq 0$ there are integers q and r such that $n = bq + r$ and $0 \leq r < 1$.

We use this result to prove the rest of the full theorem afterward.

Basis ($n = 0$): $0 = b0 + 0$ and $0 \leq 0 < b$. $q = r = 0$ works.

Induction step: Let $k \geq 0$ be chosen arbitrarily and suppose there are q_1, r_1 such that $k = bq_1 + r_1$ and $0 \leq r_1 < b$ (this is the ind hyp).

Our aim is to find q and r such that $k+1 = bq+r$ and $0 \leq r < 1$.

If $r_1 < b - 1$ then $k + 1 = bq_1 + (r_1 + 1)$ and $0 \leq r_1 + 1 < b$.

In this case let $q = q_1$ and $r = r_1 + 1$.

If $r_1 < b - 1$ is false, then $r_1 = b - 1$ (because there is no integer between 0 and 1, so there is no integer between $b - 1$ and 1). In this case, $k + 1 = q_1 b + r_1 + 1 = q_1 b + b = (q_1 + 1)b + 0$

and we can let $q = q_1 + 1$, $r = 0$.

So we have shown by math induction that for each $n \geq 0$ there are integers q and r such that $n = bq + r$ and $0 \leq r < 1$.

existence of q and r when $a < 0$: We need to deal with the case $a < 0$ in the main theorem: if $a < 0$ then $-a > 0$ and we have shown that there are q_1, r_1 such that $-a = bq_1 + r_1$ and $0 \leq r_1 < b$.

If $r_1 = 0$, then let $q = -q_1$, $r = 0$ and we have $a = -(-a) = -(bq_1) = b(-q_1) + r$, and of course $0 \leq 0 = r < b$.

If $r_1 > 0$, then let $q = -(q_1 + 1), r = b - r_1$. $bq + r = b(-q_1 + 1) + b - r_1 = -(bq_1 + r_1) = -(-a) = a$ and $0 < b - r_1 < b$ follows because $0 < r_1 < b$.

proof of uniqueness of q and r: Now we need to show that in all these cases, there can be only one such q and r.

Suppose $a = bq_1 + r_1$ and $a = bq_2 + r_2$, with $b > 0$.

We can suppose further that $r_1 \geq r_2$ (trichotomy, and pick the smaller one to be r_2).

Subtract to get $b(q_2 - q_1) = r_1 - r_2$. We have $r_1 - r_2$ nonnegative and strictly less than b.

From $0 \leq b(q_2 - q_1) < b$ we can conclude $0 \leq q_2 - q_1 < 1$ so $q_2 - q_1 = 0; q_1 = q_2$, because there is no integer strictly between 0 and 1.

So we have shown that if $a = bq_1 + r_1$ and $a = bq_2 + r_2$, with $b > 0$, it follows that $q_1 = q_2$.

Now if $bq + r_1 = bq + r_2$, it follows that $r_1 = r_2$ by adding $-bq$ to both sides of the equation.

So we are done.

4 Greatest common divisor (gcd) and the Euclidean algorithm

Definition (common divisor): Let a, b be integers. We say that d is a common divisor of a and b just in case $d|a$ and $d|b$.

Observations: If $a = b = 0$ then every integer is a common divisor of a and b, because every integer is a divisor of 0.

If $d|a$ and $a \neq 0$ then $a = kd$ for some integer k and so $|a| = k'd$ for $k' = \pm k$. Now if $d < 0$ we certainly have $d \leq |a|$, and if $d > 0$ we clearly have $d \leq k'd = |a|$. That is, $|a|$ is the largest divisor of a if $a \neq 0$.

This further implies that if a, b are not both 0 and d is a common divisor of a and b then $d \leq \texttt{max}(|a|, |b|)$.

Finally, a fact which we will prove later: any set of integers which has an upper bound has a largest element. So there is a greatest common divisor of a, b if a and b are not both zero.

Another way to see that the greatest common divisor exists is to note that an integer other than zero has only finitely many divisors, so the set of common divisors of two numbers which are not both zero is finite, and a finite set has a largest element. These statements actually require some analysis too!

Definition (greatest common divisor, gcd): For any pair of integers a, b which are not both zero, we define $\texttt{gcd}(a, b)$, the greatest common

divisor of a and b, as the largest element of the set of common divisors of a and b, which we have shown above exists.

This concept is familiar: The greatest common divisor is again a concept familiar from elementary school. When you simplify a fraction $\frac{a}{b}$, the common factor by which you divide the numerator and denominator is the gcd of a and b,

You learned a method for finding these common factors using prime factorizations. We will teach a much better method (at least, better for large numbers) which is ancient: it was known to Euclid).

Theorem (facts about the gcd):

1. $\texttt{gcd}(a, b) = \texttt{gcd}(b, a)$ Obvious by symmetry of the definition. Because of this, we can assume $a \geq b$.
2. $\texttt{gcd}(a, b) = \texttt{gcd}(|a|, |b|)$

   This is obvious: the divisors of a are the same as the divisors of $|a|$ and the same is true of b, so the common divisors of a, b make up the same set as the common divisors of $|a|, |b|$ and of course these sets have the same largest element.

   Because of this, we can assume a, b are nonnegative in gcd calculations (with possible fixes later if we have to think about the case where one of them might be negative). So our default assumption is that a is positive and $a \geq b \geq 0$.
3. $\texttt{gcd}(a, 0) = |a|$

   Common divisors of a and 0 are exactly the divisors of a, of which the largest is $|a|$. If a is positive of course this simplifies to $\texttt{gcd}(a, 0) = a$.
4. If $b > 0$, $\texttt{gcd}(a, b) = \texttt{gcd}(b, a \texttt{ mod } b)$.

   This takes a little more work. Notice that if $q = a \texttt{ div } b$ and $r = a \texttt{ mod } b$, we have $a = bq + r$ and $r = a - bq$. We will use q, r with these meanings.

   Now we argue that the set of common divisors of a and b is the same set as the set of common divisors of b and $r = a \texttt{ mod } b$.

   To show this, we show that any element of the first set belongs to the second and any element of the second set belongs to the first.

If x belongs to the set of common divisors of a and b, then $x|a$ and $x|b$. If $x|b$ then certainly $x|qb$. If $x|a$ and $x|qb$ then $x|(a-qb) = r$. So x belongs to the set of common divisors of b and $r = a \bmod b$.

If x belongs to the set of common divisors of b and $r = a \bmod b$, then $x|b$ and $x|r$. If $x|b$ then $x|bq$. If $x|bq$ and $x|r$, it follows that $x|bq + r = a$. So x also belongs to the set of common divisors of a and b.

Since these two sets are the same, they have the same largest element, so $\texttt{gcd}(a, b) = \texttt{gcd}(b, a \bmod b)$.

Notice that our default assumption that the first argument is positive and the second is nonnegative and smaller holds automatically for $\texttt{gcd}(b, a \bmod b)$.

Theorem (Euclidean algorithm): Let a, b be integers with $b > 0$.

Define a sequence D by $D_0 = a$, $D_1 = b$ and $D_{n+2} = D_n \bmod D_{n+1}$. Notice that D_{n+2} is only defined if $D_{n+1} \neq 0$.

We argue that for every a, b there is an n such that $D_{n+1} = 0$ and $D_n = \texttt{gcd}(a, b)$.

First we prove by induction that for every integer n, $\texttt{gcd}(D_n, D_{n+1}) = \texttt{gcd}(a, b)$ if D_{n+1} is defined (and so in particular if $D_{n+1} = 0$ then $D_n = \texttt{gcd}(a, b)$).

The basis is the assertion $\texttt{gcd}(D_0, D_1) = \texttt{gcd}(a, b)$ which is true by definition of D.

The induction step: If we assume $\texttt{gcd}(D_n, D_{n+1}) = \texttt{gcd}(a, b)$, then by the previous theorem $\texttt{gcd}(D_n, D_{n+1}) = \texttt{gcd}(D_{n+1}, D_n \bmod D_{n+1}) = \texttt{gcd}(D_{n+1}, D_{n+2})$ if D_{n+2} is defined.

We aren't done! We have to prove that for every a, b, there is n such that D_n is undefined (that the process stops). We prove this by induction (there is another way to prove it which we will discuss shortly). The trick, as is often the case, is to see which variable to apply induction to. We choose to do strong induction on b.

For $b = 1$, the result is clearly true: $D_0 = a, D_1 = 1, D_2 = 0$, and D_3 is undefined.

Suppose that for every a and for every b greater than one and less than k a D sequence must terminate.

The sequence which starts $D_0 = a, D_1 = k+1, D_2 = a \,\mathtt{mod}\, (k+1) \ldots$ can then be seen to terminate because the sequence beginning $D_1 = k+1, D_2 = a \,\mathtt{mod}\, (k+1) \ldots$ must terminate by ind hyp, because $a \,\mathtt{mod}\, (k+1) \leq k$ (if $a \,\mathtt{mod}\, (k+1)$ is zero, the original sequence terminates and the tail sequence isn't even a D sequence).

And this completes the argument. I'll expand these notes with sample calculations after Friday's lecture, when I will have more computational techiques for you to try out. The next homework assignment will appear as a section in these notes after the lecture on Friday.

If you don't recognize the fancy account of the Euclidean algorithm using a recursively defined sequence as the same procedure we followed at the end of Monday's lecture in an example...first, I assure you that it is. If you have notes, try computing terms of the D sequence with the numbers I used on Monday and you will hopefully see what happens.

If you don't, still fear not. The first thing I'm doing on Friday is relecturing the statement of the Euclidean algorithm in this format.