Notes on Number Theory, Fall 2022

Randall Holmes

October 12, 2022

10/12 On page 6 you will find promised theorems about divisibility. I am inserting a section on modular arithmetic (the 10/10 lecture topic) as well.

Updated 10/8. 7 pm fixed a "typo" in the extended Euclidean algorithm theorem: wrote all the quotients backward!

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.

Contents

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 $A \in \mathcal{P}(\mathbb{R})$, 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.

Theorem: If a, b, c are integers and $a|b$, then $a|(bc)$

Proof: Because $a|b$, we can choose an integer k such that $ak = b$.

To show $a|(bc)$, we need to find an integer m such that $am = bc$.

We know that $bc = (ak)c = a(kc)$, so $m := kc$ does the trick: we have $a|(bc)$ because there is an integer $m[= kc]$ such that $mm = bc$.

The point here is not that it is hard to see that it is true, but to see what a formal proof of such a statement looks like.

Theorem: If a, b, c are integers, and $c|a$ and $c|b$, then $c|(a + b)$.

Proof: Because $c|a$, we can choose an integer x such that $cx = a$.

Because $c|b$, we can choose an integer y such that $cy = b$.

We need an integer m such that $cm = a + b$ to establish the theorem.

Now $a + b = cx + cy = c(x + y)$, so if we choose $m = x + y$, we have shown that there is an integer m such that $cm = a+b$, that is, $c|(a+b)$.

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 $r = a \bmod b$.

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 $D_{n+1} = 0$, D_{n+2} is undefined, 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 $D_0 = a, D_1 = k + 1, D_2 = a \,\texttt{mod}\, (k + 1) \ldots$ terminates if $a \,\texttt{mod}\, (k+1)$ is zero. Otherwise, it can be seen to terminate because the sequence beginning $D_1 = k+1, D_2 = a \,\texttt{mod}\, (k+1) \ldots$ must terminate by ind hyp, because $a \,\texttt{mod}\, (k + 1) \leq k$.

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.

We comment that not only does this method of computation of gcd's always terminate, but it actually terminates fairly fast: the number of steps until the D sequence terminates is proportional to the logarithm of a.

We give a computational example.

Example: Compute the greatest common denominator of 925 and 851.

$\texttt{gcd}(925, 851)$

$= \texttt{gcd}(851, 925 \,\texttt{mod}\, 851)$

$= \texttt{gcd}(851, 925 - (1)851)$

$= \texttt{gcd}(851, 74)$

$= \texttt{gcd}(74, 851 \,\texttt{mod}\, 74)$

$= \texttt{gcd}(74, 851 - (11)(74))$

$= \texttt{gcd}(74, 37) = 37$ (37 goes evenly into 74).

Now we present a more impressive result, probably entirely unexpected to you.

Definition: An integer c is said to be a linear combination of integers a and b iff there are integers x and y such that $c = ax + by$.

Theorem (extended Euclidean algorithm): For any integers a, b which are not both zero, $\texttt{gcd}(a, b)$ is a linear combination of a and b, that is, there are integers x and y such that $ax + by = \texttt{gcd}(a, b)$.

Proof: We prove this by a method similar to our verification of the Euclidean algorithm, by defining additional sequences which will give us our x and y.

We assume that a, b are integers, not both zero, with $b > 0$ (notice that the result for negative b follows very easily from the result for positive b).

Define D_n as above.

Define additional sequences X and Y: $X_0 = 1; X_1 = 0; Y_0 = 0; Y_1 = 1; X_{n+2} = X_n - (D_n \texttt{div} D_{n+1})(X_{n+1}); Y_{n+2} = Y_n - (D_n \texttt{div} D_{n+1})(Y_{n+1})$.

The recurrence relations for the X and Y sequences are closely related to those for D:

$D_{n+2} = D_n \texttt{mod} D_{n+1} = D_n - (D_n \texttt{div} D_{n+1})(D_{n+1})$, which is precisely parallel in form to the definitions for X and Y.

This is actually a description of the table format I used in class, of which I'll give an example below.

We prove by strong induction on n that $D_n = aX_n + bY_n$ for every n.

The basis is direct: $D_0 = a = a1 + b0 = aX_0 + bY_0; D_0 = b = a0 + b1 = aX_1 + bY_1$.

Assume that $D_m = aX_m + bY_m$ for every $m \leq k$ and show that $D_{k+1} = aX_{k+1} + bY_{k+1}$:

If $k = 0$ we have already shown this, so assume $k \geq 1$.

$D_{k+1} = D_{k-1} - (D_{k-1} \texttt{div} D_k) D_k$ which by ind hyp

$= (aX_{k-1} + bY_{k-1}) - (D_{k-1} \texttt{div} D_k)(aX_k + bY_k)$ which by algebra

$= a(X_{k-1} - (D_{k-1} \texttt{div} D_k) X_k) + b(Y_{k-1} - (D_{k-1} \texttt{div} D_k) Y_k)$ which by the recurrence relation for X and Y

$= aX_{k+1} + bY_{k+1}$

so we have shown that every D_n is a linear combination of a and b, and we know that the second to last D_n is $\texttt{gcd}(a, b)$, so $\texttt{gcd}(a, b)$ is a linear combination of a and b.

[when I first put this up I had $k - 1$th terms and kth terms of the sequences reversed.]

I present an example of this calculation. The proof is not just an abstract argument: it directly describes how to do the calculation.

	x	y	q
925	1	0	
851	0	1	
74	1	-1	1
37	-11	12	11
0			

This calculation shows that $\texttt{gcd}(925, 851) = 37 = (-11) \cdot 925 + 12 \cdot 851$.

The first column is the D sequence, the second is the X sequence, the third is the Y sequence, and the fourth contains the integer quotient of the two entries in the first column just above that row, used in computation of the row the quotient appears in.

I will provide a spreadsheet to download on the class web page which carries out these calculations automatically. I would use this to **check** calculations, not do them in the first place, because you will not have the spreadsheet on the exam and you will need to use this computational procedure several times, not just do one demonstration.

We will demonstrate practical uses for this computation method throughout the later parts of this unit: we will use it constantly and you need to master it.

First, we present an abstract use for it in a proof:

Theorem (Euclid's lemma): Let p be a prime and let a, b be integers. If $p|ab$ then either $p|a$ or $p|b$.

Proof: Suppose p is a prime a, b are integers and $p|ab$.

If $p|a$ we are done.

Suppose p does not go into a. Then $\texttt{gcd}(p, a) = 1$ (because the only divisors of p are p and 1). Thus there are integers x and y such that $px + ay = 1$, by the extended Euclidean algorithm theorem.

Now $b = b1 = bpx + bay$. bpx is obviously divisible by p. bay is divisible by p because $p|ab$. The sum of two numbers divisible by p is divisible by p, so $p|b$, and we have shown that in either case we either have $p|a$ or $p|b$.

This theorem will be used to prove the uniqueness of prime factorizations.

Please note that this proof or something similar might appear as a test question. Be familiar with it.

5 Homework assigned 10/8/2022

1. Compute the `gcd` of each pair of numbers and present the gcd as a linear combination of the original pair of numbers.

   You can get answers (with care) using the spreadsheet, but I strongly recommend doing these (and more examples) by hand. You will need the skill of constructing these tables and reading them correctly for later activities in this unit.

   (a) 55,34

   (b) 337,216

   (c) 12076, -8976. You need to think about how to handle the fact that b is negative. The spreadsheet does **not** work correctly with negative b. Hint: I would do a calculation with positive values then fix it.

2. Write out a proof that if $d|a$ and $d|b$, then $d|(a-b)$. This is very similar to something I did in class.

3. Suppose that each of us have a large supply of 115 pound notes and a large supply of 389 pound notes. How can I pay you one pound?

4. Suppose that in my mad scientist lab, recently devastated by a monster I unwittingly created, I have a balance, but can only find a large supply of 651 gram weights and a large supply of 133 gram weights.

   Can I verify the weight of an object that is supposed to weigh 28 grams?

   What is the smallest weight I can check with these weights (remember that I can put my known weights in either pan of the balance).

6 Modular arithmetic

Let $m > 1$ be an integer. We will describe the systems of "mod m arithmetic" (one for each value of m) which are finite (though large if m is large) mathematical systems which look a lot like the integers (and some of them even like the rational numbers) in theoretically and practically interesting ways.

The objects of mod m arithmetic are written as the remainders mod m (the numbers $0, 1, \ldots, m-1$). These can be thought of as either literally those numbers, , or we can think of an element n of the system of mod m arithmetic as the set $\{x \in \mathbb{Z} : x \texttt{mod} m = n\}$. Both viewpoints have some merit.

We define an important relation.

Definition: We define a relation on integers x, y. We say that x is congruent to y mod m, which is written either $x \equiv_m y$ or $x \equiv y \texttt{mod} m$ (both usages are fairly common and I am likely to write both without thinking about it) just in case $m|(x-y)$ or, equivalently, $x \,\texttt{mod}\, m = y \,\texttt{mod}\, m$.

Theorem (verifying something I say in the definition): For any integers x, y, $m|(x-y)$ iff $x\texttt{mod}m = y\texttt{mod}m$.

We remind you that the division algorithm tells us that for any integer z, $z = m(z\texttt{div}m) + z\texttt{mod}m$ and $0 \leq z\texttt{mod}m < m$.

Suppose $m|(x-y)$. $x - y = (m(x\texttt{div}m) + x\texttt{mod}m) - (m(y\texttt{div}m) + y\texttt{mod}m) = m(x\texttt{div}m - y\texttt{div}m) + (x\texttt{mod}m - y\texttt{mod}m)$. From this we can see $x\texttt{mod}m - y\texttt{mod}m = (x-y) - m(x\texttt{div}m - y\texttt{div}m)$ is divisible by m. But $-m < x\texttt{mod}m - y\texttt{mod}m < m$, so $\texttt{xmod}m - y\texttt{mod}m$ can only be divisible by m if it is 0, establishing $x\texttt{mod}m = y\texttt{mod}m$.

Suppose that $x\texttt{mod}m = y\texttt{mod}m$. Then $x - y = (m(x\texttt{div}m) + x\texttt{mod}m) - (m(y\texttt{div}m) + y\texttt{mod}m) = m(x\texttt{div}m - y\texttt{div}m) + (x\texttt{mod}m - y\texttt{mod}m) = m(x\texttt{div}m - y\texttt{div}m)$, which is divisible by m.

Definition: A relation R on a set A is an equivalence relation if and only if it is reflexive (for all $x \in A$, xRx), symmetric (for all x, y, if xRy then yRx) and transitive (for all x, y, z, if xRy and yRz, then xRz).

Notice that equality is an equivalence relation, on any set.

Theorem: $\equiv_m$ is an equivalence relation.

Quick proof: I gave a longer one in class, which I will eventually put here. If we use the formulation of $x \equiv_m y$ as meaning $x\texttt{mod}m = y\texttt{mod}m$, this is obvious, basically because equality is an equivalence relation.

reflexive: $x \equiv_m x$ if x is an integer, because $x\texttt{mod}m = x\texttt{mod}m$.

symmetric: if $x \equiv_m y$, then $x\texttt{mod}m = y\texttt{mod}m$, so $y\texttt{mod}m = x\texttt{mod}m$, so $y \equiv_m x$.

transitive: if $x \equiv_m y$ and $y \equiv_m z$, then $x\texttt{mod}m = y\texttt{mod}m = z\texttt{mod}m$, so $x\texttt{mod}m = z\texttt{mod}m$, so $x \equiv_m z$.

The proof I gave in class uses the other equivalent form of the definition $(m|(x-y))$. It is not too much harder, and is a nice example of basic proofs about divisibility, so it will eventually be here.

As I observed above the objects of mod m arithmetic are represented by the numerals $0, \ldots m-1$ which are remainders mod m. We can think of these as literally those integers, or we can think of the object represented by n $(0 \leq n < m)$ as the set $\{x \in \mathbb{Z} : x \equiv_m n\}$. These classes are called the equivalence classes under the equivalence relation $\equiv_m$.

I will generally take the view that the objects of modular arithmetic are literally the remainders, but if so I am also taking the view informally that all the integers congruent mod m to a given remainder are in some way being identified with it in our modular arithmetic calculations.

We present the addition and multiplication tables for mod 5 arithmetic.

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Each of these tables is constructed by computing the usual operation on integers, then taking the remainder mod 5. Temporarily using $\oplus$ for modular addition and $\odot$ for modular multiplication (we will usually use the standard notation) we define $x \oplus y = (x + y)\texttt{mod}m$ and $x \odot y = (x \cdot y)\texttt{mod}m$.

You should be able to construct these tables for other small values of m.

The addition tables are very stereotyped. The multiplication tables are more interesting.

We state a theorem which justifies this procedure (and indicates why we might identify remainders in mod m arithmetic with whole equivalence classes).

Theorem: Let $x \equiv_m x'$ and $y \equiv_m y'$. It follows that $x + y \equiv_m x' + y'$ and $x \cdot y \equiv x' \cdot y'$.

Proof: Since $x \equiv_m x'$ and $y \equiv_m y'$, we have integers k, l such that $x + km = x'$, $y + lm = y'$.

Then $x' + y' = (x + km) + (y + lm) = x + y + (k + 1)m \equiv_m x + y$

and

$x' \cdot y' = (x + km)(y + lm) = xy + kmy + xlm + klm^2 = xy + (ky + lm + klm)m \equiv_m xy$

This means that as long as we are only interested in remainders mod m of outputs of the addition and multiplication operators we can in effect identify numbers with the same remainder mod m where they appear as inputs (we can collapse input numbers to their smaller remainders before carrying out addition or multiplication).

The commutative, associative, and distributive laws of algebra hold in these systems because they hold in the integers.

Order properties do not hold (you can think about how they fail).

The additive inverse property holds: the additive inverse of n in mod m arithmetic will be $m - n$ (so for example the additive inverse of 2 in mod 5 arithmetic is 5-2 = 3. $2 + 3\texttt{mod}5 = 0 \ldots$.

The multiplicative cancellation property does not hold in most systems of modular arithmetic. For example, $2 \cdot 3 \equiv_6 0 \cdot 3 \equiv_6 0$: in mod 6 arithmetic, $2 \cdot 3 = 0 \cdot 3$, $3 \neq 0$, but $2 \neq 0$ (we cannot "divide both sides by 3" to see that $2 = 0$ as we could in ordinary algebra).

But the multiplicative cancellation property *does* hold in mod 5, and in fact mod 5 arithmetic has a stronger property characteristic of the rational

numbers, not the integers: each nonzero remainder in mod 5 arithmetic has a multiplicative inverse, so we really can define division in this system (which is really unexpected).

There is a general theorem about when this is true.

Theorem: Mod m arithmetic satisfies the multiplicative cancellation property and in fact the existence of reciprocals of each nonzero remainder if and only if m is prime.

Proof: If m is not prime then there are a and b such that $0 < a \leq b < m$ and $ab = m$.

We then have in mod m arithmetic that $a0 = ab = 0$ (because $a0 \equiv_m 0 \equiv_m m = ab$) but $0 \neq b$, so multiplicative cancellation fails (and so a does not have a reciprocal: if a^{-1} such that $a^{-1}a = 1$ in mod m arithmetic existed we would have $0 = 10 = (a^{-1}a)0 = a^{-1}(a0) = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1b = b$, which is absurd.

Now suppose that m is prime. It follows that for any nonzero remainder a, we have $\texttt{gcd}(a, m) = 1$, so there are integers x, y such that $ax+my = 1$ (in the ordinary arithmetic of the integers) so $ax \equiv_m 1$, so $x]ttmodm$ is a reciprocal of a.

Further, there is only one reciprocal of a in mod m arithmetic. Suppose $ax \equiv_m ay = 1$. It follows that $m|(ax-ay)$ so $m|a(x-y)$. Now, because m is prime and m does not go into a, we must have $m|(x-y)$ by Euclid's Lemma, so $x \equiv_m y$, and there is only one inverse of a in mod m arithmetic, which we will write a^{-1}, or $a^{-1}\texttt{mod}m$. This is not to be confused with $\frac{1}{a}$, the usual reciprocal of a.

It then follows that if $ac = bc$ and $c \neq 0$, we have $(ac)c^{-1} = (bc)c^{-1}$ and so $a = b$, ini mod m arithmetic as in the arithmetic of the rationals or reals.

This shows us not only that there are reciprocals of nonzero remainders in mod m arithmetic, but also how to compute them using the extended Euclidean algorithm. Moreover, we can solve equations of the form $ax \equiv_m b$ in mod m arithmetic, if m is prime and a is not 0, by multiplying both sides by $a^{-1}\texttt{mod}m$.

Example: Find $23^{-1}\texttt{mod}137$

	x	y	q
137	1	0	
23	0	1	
22	1	-5	5
1	-1	6	1

We see that $(-1)137+(6)(23) = 1$ so $(23)(6) \equiv_{137} 1$ so $23^{-1}\texttt{mod}137 = 6$.

Example: Solve $23x \equiv_{137} 31$

Multiply both sides by the reciprocal we just found.

$x = 1x \equiv_{137} (6)(23)x \equiv_{137} (6)(31)$ for any x for which the original equation holds, so we must have $x \equiv_{137} 186 \equiv_{137} 49$ and indeed you can check that 49 is a solution.