INTRODUCTION TO NUMBER THEORY

Contents

Date: April 10, 2023.

0. Introduction

These lecture notes will be updated periodically as the semester progresses. The source code can be found on Canvas.

Anything with an asterisk denotes material beyond the scope of the course. I might also not talk about them in class in favour of more time developing examinable contents.

The exercises in these notes are *not* homework problems. They are intended for you to think about as you are reading the notes.

I will assume you know basic notations in set theory. Here is a list.

- $a \in A$: a is an element in the set A.
- $A \subseteq B$: A is a subset of B.
- $f : A \to B$: f is a function from A to B.
- $a \mapsto b$: a gets mapped to b by a function.
- $\{a \in A \,|\, P(a)\}$. the subset of A consisting of all elements such that the statement $P(a)$ holds.
- $A \cup B$: the union of A and B.
- $A \cap B$: the intersection of A and B.
- $A - B$: the set difference, equivalent to $\{a \in A \,|\, a \notin B\}$.
- $A \times B$: the Cartesian product, i.e. set of ordered pairs (a, b) where $a \in A$ and $b \in B$.
- $\#A$: the number of elements in the finite set A.

1. Foundations

In this chapter, we investigate the basic question: what are numbers?

The way we tackle this question is to answer a slightly different question: what properties do the natural numbers satisfy. We will write down some of those properties, and in the rest of the course, we will build everything up from them. This is the *axiomatic* way of looking at objects. There is a (maybe false) sense of security that everything we do is ultimately derived from some very basic hypotheses.

The list of properties we write down must be strong enough to deduce all the things we intuitively understand about the natural numbers. For example, the real numbers *must not* satisfy all of the axioms. The key player here is the *principle of mathematical induction*, which in its various forms will be an important proof technique in this course.

This chapter deviates significantly from the textbook. Most of the material is based on Sections 11–13 of P. Halmos' *Naive Set Theory*. In particular, we will freely use set theory. The last section will comment on this issue further. The material here is not typically considered number theory. It is included mainly to get you thinking about proofs.

1.1. **The natural numbers.**

Intuitively, the set of natural numbers is

$$\mathbb{N} = \{0, 1, 2, \cdots\}$$

In this course, 0 is a natural number. For each natural number n, there is one immediately after it, which we call the successor of n. Moreover, all natural numbers are built this way starting from 0.

How do we formalize this? From the above description, there is a special element 0, and there is a special operation of "taking the successor". Formally, this is a function $\mathbf{S} : \mathbb{N} \to \mathbb{N}$, so $\mathbf{S}(n)$ is the successor of n. The image in our head should be something like

$$0 \xrightarrow{\mathbf{S}} 1 = \mathbf{S}(0) \xrightarrow{\mathbf{S}} 2 = \mathbf{S}(1) = \mathbf{S}(\mathbf{S}(0)) \xrightarrow{\mathbf{S}} \cdots$$

Here, the symbol 1 is *defined* to be a shorthand for $\mathbf{S}(0)$, similarly for 2, 3, and so on. The diagram starts with 0, so 0 is not the successor of any natural number. This picture translates to the following properties

(0) 0 is a natural number.

(1) There does not exist a natural number n such that $\mathbf{S}(n) = 0$.

These seem like reasonable properties, but they cannot rule out strange behaviours: the following is perfectly acceptable so far

$$\begin{tikzcd}
0 \arrow[r, "\mathbf{S}"] & 1 = \mathbf{S}(0) \overset{??}{=} \mathbf{S}(2) \arrow[r, "\mathbf{S}"] & 2 = \mathbf{S}(1) \arrow[l, bend left, "\mathbf{S}"]
\end{tikzcd}$$

So this set of "natural numbers" is finite. We want to forbid loops, and the easiest way to say it is that $\mathbf{S}$ is injective:

(2) For all natural numbers m and n, if $\mathbf{S}(n) = \mathbf{S}(m)$, then $n = m$.

This property guarantees an infinite chain of natural numbers like our intuition. Somehow, we need to say the natural numbers is the smallest set satisfying the above properties. A natural guess would be the following property

(3?) If n is a natural number, then either $n = 0$ or $n = \mathbf{S}(m)$ for some natural number m.

This is far from enough: you can't reasonably define addition in this theory. At best, this formalizes some kind of order structure that starts with the natural numbers. Instead, we need to capture the idea that *all* natural numbers are produced as repeated successors.

Exercise. Find a set M strictly containing $\mathbb{N}$ with a function $\mathbf{S} : M \to M$ satisfying the four properties above.
Hint: $\mathbb{N} + \mathbb{Z}$*. The notation is not introduced, but think about what it could mean.*

The correct final axiom turns out to be the *principle of mathematical induction.*

(3) **Induction**: Let P be a subset of natural numbers. Suppose that
  (a) $0 \in P$.
  (b) For all natural numbers n, if $n \in P$, then $\mathbf{S}(n) \in P$.
  Then P is the set of all natural numbers.

From the hypothesis, $0 \in P$ holds, so by 3(b), $1 \in P$, so $2 \in P$, and so on, which suggests that $n \in P$ for all natural numbers n we can think of. In this way, the induction principle captures the notion that the natural numbers we intuitively know are all of the natural numbers.

From the discussion above, we have written down four properties. We now give them the status of *axioms*: these are the *only* properties we will assume about natural numbers.

Axiom (The Peano axioms)**.** The natural numbers is a set $\mathbb{N}$ with an element 0 and a function $\mathbf{S} : \mathbb{N} \to \mathbb{N}$ satisfying the following properties

(PA1) There does not exist a natural number n such that $\mathbf{S}(n) = 0$.
(PA2) For all natural numbers m and n, if $\mathbf{S}(n) = \mathbf{S}(m)$, then $n = m$.
(PA3) The induction principle holds.

Suppose we find a set $\mathbb{M}$ in the wild, pick out an element we call 0, and somehow defined an operation $\mathbf{S}$ on $\mathbb{M}$ satisfying the above three properties, then we could call $\mathbb{M}$ the set of natural numbers. Everything we develop in this course holds for this possibly exotic looking set. Such a set $\mathbb{M}$ (together with the choice of 0 and $\mathbf{S}$) is called a *model* of the axioms.

Example. We will show from the axioms that for all natural numbers n, $\mathbf{S}(n) \neq n$.

Proof. Let G be the set of $n \in \mathbb{N}$ such that $\mathbf{S}(n) \neq n$. From (PA1), $0 \in G$. Suppose $n \in G$, then $\mathbf{S}(n) \neq n$, so by (PA2), $\mathbf{S}(\mathbf{S}(n)) \neq \mathbf{S}(n)$. This is exactly the statement that $\mathbf{S}(n) \in G$. By (PA3), $G = \mathbb{N}$, so $\mathbf{S}(n) \neq n$ holds for all $n \in \mathbb{N}$. □

Clearly, this is just rephrasing a proof by induction.

Proof by induction. This holds for $n = 0$ by (PA1). Suppose for a given n, we have $\mathbf{S}(n) \neq n$. By (PA2), $\mathbf{S}(\mathbf{S}(n)) \neq \mathbf{S}(n)$, so the statement holds for $n + 1$. This concludes the inductive step, and hence the proof. □

Formally, whenever we do a proof by induction, we are implicitly defining a set G (for "good") consisting of all numbers satisfying the property and showing that $G = \mathbb{N}$ using (PA3). Therefore, proof by induction is built into the axioms.

Remark. Later, we will explain the well-ordering property, which might feel more intuitively and morally correct compared to induction. The reason we chose this axiom is because we can't even define the ordering relation yet. There are also additional complications relating to property (3?).

Remember the goal was to define $\mathbb{N}$. If there are two vastly different models for the Peano axioms, then this might not be a good enough definition. For example, if we dropped (PA2), then we could have a finite set of "natural numbers", which is clearly undesirable. Fortunately, it turns out that any two models of the three axioms are equivalent (this is not too difficult using the recursion theorem we state in the next section), so it seems like we have successfully defined the natural numbers just using a list of properties.

There is one more issue: how do we know there is a system satisfying the three axioms? Nothing we have seen rules out a possible contradiction in the axioms. Indeed, what if we try to add the following axiom:

(PA4?) There exists $n \in \mathbb{N}$ such that $\mathbf{S}(n) = n$.

This leads to a contradiction since we have proven its negation in the example. Therefore, the axioms (PA1–3), (PA4?) together does not define anything. While the contradiction is not so hard to see here, consistency is a very subtle issue. We claimed that these axioms are all that is needed to develop all of elementary number theory, so the question becomes the following: is number theory consistent?

We can attempt to resolve this issue by actually finding a model, i.e. to actually define the natural numbers. In set theory, the standard choice is

$$\mathbb{N} = \{0 = \emptyset,\ 1 = \{\emptyset\},\ 2 = \{\emptyset, \{\emptyset\}\}, \cdots\}$$

with $\mathbf{S}(A) = A \cup \{A\}$. Section 12 of Halmos contains a proof that this construction satisfy the Peano axioms. The nice thing about the axiomatic approach is that we never have to worry about this bizarre-looking definition.

However, this does not resolve the consistency issue. It just kicks the can down the road, since we have only shown that *if set theory is consistent*, then the Peano axioms are consistent. A lot of mathematics is formulated using set theory, so now the question is about the consistency of mathematics. We will discuss this point at the end of the chapter. The upshot is 1. There is no easy answer; 2. We won't worry about it.

Remark.* The standard treatment of this material in a textbook on foundations would include $+$ and $\times$ in the axioms. This removes the dependency on set theory to a large extent and allows us to restrict the type of sets allowed in the induction axiom. We will discuss this more in the final section.

1.2. **Recursive definition.**

Now that we "know" what natural numbers are, how do we work with them? As a basic point, how do we define addition?

From the way we constructed the axioms, we would like to say $n+1$ is defined as $\mathbf{S}(n)$, so $n+2$ should be $(n+1)+1 = \mathbf{S}(n+1)$. Formally, one might write down the following.

Definition. Define addition $n+m$ recursively by:

$$\begin{aligned} n + 0 &:= n \\ n + \mathbf{S}(m) &:= \mathbf{S}(n+m) \end{aligned}$$

Example.

$$\begin{aligned} 1+1 &= 1+\mathbf{S}(0) \\ &:= \mathbf{S}(1+0) \\ &:= \mathbf{S}(1) \\ &= 2 \end{aligned}$$

This is another example of a "definition" that tells you how to do something instead of what something is, so we need to prove that addition is *well-defined*, meaning there exists a unique function satisfying the two properties given in the definition. Despite the obvious nature of the result, it is hard to prove. We will state the general version of the recursion theorem now and prove it later in Section 1.5.

Theorem 1.1. *Let S be a set, $a \in S$ be an element of S, and $g : \mathbb{N} \times S \to S$ be a function. There exists a unique function $f : \mathbb{N} \to S$ such that*

$$f(0) = a, \quad f(\mathbf{S}(n)) = g(n, f(n))$$

Example. To recover addition, take $S = \mathbb{N}$, $a = n$, and $g(n, s) = \mathbf{S}(s)$.

This way of defining functions should be familiar to anybody who has done functional programming, and the underlying mathematics is in fact similar. In some sense (Curry–Howard correspondence), a proof by induction is also a recursive definition, but this is moving to a different foundation of mathematics.

We now give a few more examples of recursive definitions.

Definition. Define multiplication $n \times m$ by:

$$\begin{aligned} n \times 0 &:= 0 \\ n \times (m+1) &:= (n \times m) + 2 \end{aligned}$$

Definition. Define the factorial function $n!$ by:

$$\begin{aligned} 0! &:= 1 \\ (n+1)! &:= n! \times (n+1) \end{aligned}$$

Definition. Define the Fibonacci numbers F_n by:

$$\begin{aligned} &F_0 := 0,\ F_1 := 1 \\ &F_{n+2} := F_n + F_{n+1} \end{aligned}$$

In each case, we can use the general recursion theorem to prove that the result is well-defined. We have also used $n+1$ as a shorthand for $\mathbf{S}(n)$ to make it clearer what's going on. From now on, when we write a recursive definition, we will not comment on these aspects as long as the definition looks alright, but the recursion theorem is always implicitly present.

Exercise. How can one use Theorem 1.1 to show that the Fibonacci numbers are well-defined?

The definition we have given so far are called *primitive recursions*. There are more sophisticated recursive definitions.

Definition. The *Ackermann function* $A : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is defined recursively by

$$\begin{aligned} &A(0,n) := n+1 \\ &A(m+1,0) := A(m,1) \\ &A(m+1,n+1) := A(m, A(m+1,n)) \end{aligned}$$

Example. We want to compute $A(3,2)$. By expanding the definitions, we get

$$A(3,2) = A(2, A(3,1)) = A(2, A(2, A(3,0))) = A(2, A(2, A(2,1)))$$

So it would be helpful to know $A(2,n)$ in general. By definition,

$$A(2,0) = A(1,1),\ A(2,n+1) = A(1, A(2,n))$$

This is an instance of primitive recursion, provided we know the values of $A(1,n)$ for all n.

We can unwind further,

$$A(1,0) = A(0,1) = 2,\ A(1,n+1) = A(0, A(1,n)) = A(1,n) + 1$$

This is exactly the definition of the function $n \mapsto 2+n$, so we have $A(1,n) = 2+n$. Now, the recursion for $A(2,n)$ reads

$$A(2,0) = 3,\ A(2,n+1) = 2 + A(2,n)$$

Therefore, $A(2,1) = 2+3 = 5$, $A(2,2) = 2+5 = 7$, and so on. The table of the first 15 values can be computed by direct unfolding of the definition. They are

$$3, 5, 7, 11, 13, 17, 19, 21, 23, 27, 29, 31$$

Once we know the basic arithmetic properties of $\mathbb{N}$, it is very easy to prove by induction that $A(2,n) = 2n+3$, but we stress that this is not required.

Now, back to the computation of $A(3,2)$.

$$\begin{aligned} A(3,2) &= A(2, A(2, A(2,1))) \\ &= A(2, A(2,5)) \\ &= A(2,13) = 29 \end{aligned}$$

It's important to note that the definition unfolds in finite time to a long iterated list of successor functions, but it is a more complex unfolding than what we have seen. It's also less clear why this is well-defined. We will leave this question in suspense until we prove the recursion theorem.

1.3. **Properties of natural numbers.** We know that $2+1 = \mathbf{S}(2) = 3$, but also $1+2 = \mathbf{S}(1+1) = \mathbf{S}(2) = 3$. Coincidence?

Proposition 1.2 (Associativity of addition)**.** *If $a, b, c \in \mathbb{N}$, then*

$$(a+b)+c = a+(b+c)$$

Proof. Fix a, b and apply induction on c. The result is clear when $c = 0$. Suppose it holds for c, then

$$\begin{aligned} (a+b)+(c+1) &= \mathbf{S}((a+b)+c) \\ &= \mathbf{S}(a+(b+c)) \\ &= a + \mathbf{S}(b+c) \\ &= a + (b+(c+1)) \end{aligned}$$

so it holds for $c+1$. This completes the induction step, and hence the proof. □

Using induction, we can prove all of the following properties. Some of the proofs are surprisingly annoying.

Theorem 1.3. *For all $a, b, c \in \mathbb{N}$, the following holds*

- *Commutativity of addition:* $a + b = b + a$.
- *Associativity of addition:* $(a + b) + c = a + (b + c)$.
- *Additive identity:* $a + 0 = 0 + a = a$.
- *Commutativity of multiplication:* $a \times b = b \times a$.
- *Associativity of multiplicaiton:* $a \times (b \times c) = (a \times b) \times c$.
- *Multiplicative identity:* $a \times 1 = 1 \times a = a$.
- *Distributivity:* $a \times (b + c) = a \times b + a \times c$.
- *Additive cancellation: If* $a + b = a + c$*, then* $b = c$.
- *Multiplicative cancellation: If* $a \times b = a \times c$*, then* $b = c$.

The upshot is everything you think you knew is still true, just much harder to prove than expected. We will always use this theorem implicitly, just like what you have been doing your whole life.

We can also define an order relation on the natural numbers.

Definition. We write $a \leq b$ if there exists $c \in \mathbb{N}$ such that $a + c = b$. Such a c is unique, and we write $c = b - a$.

This definition is similar in nature to the definition $1 = \mathbf{S}(0)$, in that it is just a shorthand for a longer expression. In particular, there is no question whether "$\leq$" is well-defined. Compare this with the definition of addition, which has the flavour of telling you how to do something, but requires a hard theorem to prove that it is a valid definition.

Theorem 1.4. *For all $a, b, c \in \mathbb{N}$, the following holds*

- *Reflexivity:* $a \leq a$.
- *Anti-symmetry: if* $a \leq b$ *and* $b \leq a$*, then* $a = b$.
- *Transitivity: if* $a \leq b$ *and* $b \leq c$*, then* $a \leq c$.
- *Dichotomy: either* $a \leq b$ *or* $b \leq a$.
- *Order preservation: if* $a \leq b$*, then* $a + c \leq b + c$ *and* $ac \leq bc$.
- *Minimal element:* $0 \leq a$.

Again, none of the properties should come as a surprise, and we will be using these properties implicitly.

Exercise. Prove that if $a \leq b + 1$, then either $a \leq b$ or $a = b + 1$.

1.4. **Well-ordering property.** Now that we have an ordering on the natural numbers, we can state the following result.

Theorem 1.5 (Well-ordering property)**.** *If S is a non-empty subset of $\mathbb{N}$, then S has a minimal element, i.e. there exists $m \in S$ such that $m \leq n$ for all $n \in S$.*

Proof. We prove the following statement by induction on n:

> For any subset $S \subseteq \mathbb{N}$, if there exists an $m \leq n$ such that $m \in S$,
> then S has a minimal element.

The base case with $n = 0$ holds since we can take 0 to be the minimal element.

Suppose the statement holds for a given n. Let $T \subseteq \mathbb{N}$ be a subset satisfying the hypothesis of the statement for $n + 1$. If there exists an $m \leq n$ such that $m \in T$, then we are done by the induction hypothesis. Otherwise, for all $m \leq n$, T does not contain m. It follows that $n+1 \in T$, and it is the minimal element, as required. □

In the rest of this course, most of our proofs by induction will actually invoke this theorem instead of the usual proof by induction you have seen. We give a simple example.

Lemma 1.6 (Archimedean property). *If $a, b \in \mathbb{N}$ and $a \neq 0$, then there exists $n \in \mathbb{N}$ such that $na > b$.*

Proof. Fix a natural number $a \neq 0$, then by Homework 1.1, there exists $a' \in \mathbb{N}$ such that $a = a' + 1$, so $a \geq 1$. Suppose the lemma is false, then there exists a minimum b such that the conclusion fails. Clearly, $b \neq 0$, so $b = b' + 1$ for some $b' \in \mathbb{N}$. In particular, $b' < b$, so by minimality, we can find $n \in \mathbb{N}$ such that $na > b'$, which implies $(n+1)a > b' + a \geq b' + 1 = b$. This contradicts the choice of b. □

Equivalently, there is also have the following version of induction.

Corollary 1.7 (Well-founded induction). *Suppose $P(n)$ is a statement concerning natural numbers satisfying the following condition:*

$$\text{For any given } m\text{, if } P(m') \text{ holds for all } m' < m\text{, then } P(m) \text{ holds}$$

Then $P(n)$ is true for all natural numbers n.

Proof. Suppose it is not the case that $P(n)$ holds for all n, let m be the smallest natural number with $P(m)$ false. For all $m' < m$, the minimality of m implies that $P(m')$ is true. It follows from the property of P that $P(m)$ must also be true. This is a contradiction. □

You might have seen this corollary being stated as strong mathematical induction because it looks stronger than the induction axiom (PA3). Indeed, using this corollary, you are allowed to assume the proposition for *all* values under n in the induction step. One place where this might be useful is when the recursive definition involves multiple past values.

Lemma 1.8. *If $n \geq 3$, then $F_n > (1.1)^n$.*

Proof. This holds for $n = 3$ and $n = 4$ by direct computation. Suppose it holds for n and $n + 1$, then

$$F_{n+2} = F_{n+1} + F_n > (1.1)^{n+1} + (1.1)^n = (1.1)^n \cdot 2.1 > (1.1)^{n+2}$$

so the result holds for all $n \geq 3$. □

Exercise. Why do we need to check $n = 3$ and $n = 4$ separately in the proof?

However, it is not entirely true that well-founded induction is stronger than mathematical induction. The key point is that we need know any non-zero natural number is a successor.

Example. Consider the set $\{0, 1\} \times \mathbb{N}$ ordered by lexicographic ordering, i.e. we have two copies of $\mathbb{N}$, with every number in the second copy greater than every number in the first copy.

$$0 < 1 < 2 < \cdots < 0' < 1' < 2' < \cdots$$

If we define successor on each copy of $\mathbb{N}$ separately, the resulting set satisfies (PA1) and (PA2). Moreover, it satisfies the well-ordering property and therefore well-founded induction holds. However, it cannot satisfy (PA3) because a proof by induction will only ever see the first copy of $\mathbb{N}$. The issue is that $0'$ is not the successor of any element.

From the example, we see that the well-ordering property captures a notion that is more general than the natural numbers.

Definition. Let S be a set. A relation $\leq$ on S is a *total ordering* if for all $a, b, c \in S$, the following holds

1. Reflexivity: $a \leq a$.
2. Anti-symmetry: If $a \leq b$ and $b \leq a$, then $a = b$.
3. Transitivity: If $a \leq b$ and $b \leq c$, then $a \leq c$.
4. Totality: $a \leq b$ or $b \leq a$.

It is a *well-ordering* if in addition, the well-ordering property holds, i.e.

5. Let $A \subseteq S$. If $A \neq \emptyset$, then there exists $m \in A$ such that $m \leq a$ for all $a \in A$.

This is the exact list of properties required to perform well-founded induction.[1] As we will see in the next section, it is possible to do recursive definitions on general well-ordered sets, so in particular, the usual arithmetic operations can be defined. However, many properties given in Theorem 1.3 no longer holds.

Example. In the example $\{0, 1\} \times \mathbb{N}$ given above, $1 + 1' = 1'$, but $1' + 1 = 2'$.

The upshot is that the classical statement of induction (PA3) cannot be replaced completely with the well-ordering property. In general, well-ordered sets are classified using *ordinals*, and it is possible to axiomatize $\mathbb{N}$ as a "limit ordinal" where every element is a "successor". This feels more natural from a set theory point of view, but less so from a foundations point of view.

Exercise. If X is a well-ordered set without a maximal element, then we can define successor by

$$\mathbf{S}(x) = \min\{y \in X \mid y > x\}$$

We can also define $0 = \min(X)$. Check that this satisfies (PA1) and (PA2).

1.5. More on recursion.

We recall the statement of the recursion theorem.

Theorem. *Let S be a set, $a \in S$ be an element of S, and $g : \mathbb{N} \times S \to S$ be a function. There exists a unique function $f : \mathbb{N} \to S$ such that*

$$f(0) = a, \quad f(\mathbf{S}(n)) = g(n, f(n))$$

"Morally correct proof". Prove the following statement by well-founded induction:

> On the subset $I_n := \{x \in \mathbb{N} \mid x \leq n\}$, there exists a unique function f_n satisfying the two properties.

Fix $n \in \mathbb{N}$, and suppose the statement is known for all $n' < n$. This defines $f_n(x)$ for all $x < n$, and the recursive equation ensures that it extends to $x = n$ in a unique way. This completes the induction. Finally, the function we need is the union of all f_n. □

[1] There is a more general version without property (4) which includes structural induction.

This sketch can be made into a correct proof, and it even holds for any well-ordered sets. The reason we are not making it precise here is that this proof is circular: we needed addition to define $\leq$, and we needed the recursion theorem to define addition. This is one of the reasons that in most of the usual axiomatizations of arithmetic, addition is included in the axioms.

We still included this sketch because it shows that fundamentally, the reason recursive definitions work is the well-ordering property. Take a recursion such as

$$F_0 = 0,\ F_1 = 1,\ F_{n+1} = F_n + F_{n-1}$$

To evaluate F_n, we need to know its values at $n-1$ and $n-2$, which can further be computed using the recursion equation. Both $n-1$ and $n-2$ are *strictly smaller* than n, so this chain of evaluations must terminate by the well-ordering property.

The Ackermann function is a more sophisticated application of this idea. To evaluate $A(m,n)$, we need to know the value of $A(m-1,x)$ and $A(m,n-1)$, where x can be very large. This requires an ordering on $\mathbb{N}^2$ such that $(m-1,x) \leq (m,n)$ for an arbitrary x. This is given by the lexicographic ordering

$$(m,n) \leq (m',n') \text{ if } m < m' \text{ or } m = m',\ n \leq n'$$

It is not too hard to show that it is a well-ordering, so our "morally correct proof" of the recursion theorem immediately proves the existence of the Ackermann function. By the same idea, well-founded induction also proves Homework 1.3 in one induction:

$$\begin{aligned}
A(m+1,n+1) &= A(m, A(m+1,n)) \\
&= 2 \uparrow^{m-2} (A(m+1,n)+3) - 3 \\
&= 2 \uparrow^{m-2} \Big[(2 \uparrow^{m-1} (n+3) - 3) + 3\Big] - 3 \\
&= 2 \uparrow^{m-2} (2 \uparrow^{m-1} (n+3)) - 3 \\
&= 2 \uparrow^{m-1} (n+4) - 3
\end{aligned}$$

In this prove, we applied induction hypothesis at $(m, A(m+1,n))$ and $(m+1,n)$. Both of them precede $(m+1,n+1)$ in the lexicographic ordering. We still need the base cases $A(m,0)$ for $m \geq 2$ since those are the special cases in the recursive definition of $A(m,n)$. This is also easy.

$$\begin{aligned}
A(m,0) &= A(m-1,1) \\
&= 2 \uparrow m - 34 - 3 \\
&= 2 \uparrow m - 3(2 \uparrow m - 22) - 3 \\
&= 2 \uparrow m - 23 - 3
\end{aligned}$$

This uses the induction hypothesis at $(m-1,1)$, which precedes $(m,0)$.

We now give an actual proof of the recursion theorem from axioms (PA1–3). Instead of taking the bottom-up approach of building a function, it uses the top-down approach of "cutting down" the function.

Set theory proof. Consider the set A consisting of all subsets $F \subseteq \mathbb{N} \times S$ satisfying the following two properties

(1) $(0,a) \in F$.
(2) If $(n,x) \in F$, then $(\mathbf{S}(n), g(n,x)) \in F$.

Let $f = \bigcap_{F \in A} F$ be the intersection of all elements in A, then we will show that $f : \mathbb{N} \to S$ is a function satisfying the requirements.

The first step is to show that f is a function defined on $\mathbb{N}$, namely for all $n \in \mathbb{N}$, there exists a unique $x \in S$ such that $(n, x) \in f$. This will be shown by induction.

By the definition of A, $(0, a) \in f$. Suppose $(0, x) \in f$ where $x \neq a$, then the set $f - \{(0, x)\}$ also belongs to A: property (1) is satisfied since $x \neq a$, and (2) is satisfied by (PA1). This is a contradiction since f is supposed to be contained in all other sets of A. Therefore, if $(0, x) \in f$, then $x = a$. We have shown that the statement holds for 0. Now suppose it holds for an arbitrary n. Let $x \in \mathbb{N}$ be the natural number such that $(n, x) \in f$. It follows by condition (2) in the definition of A that $(\mathbf{S}(n), g(n, x)) \in f$. Suppose $(\mathbf{S}(n), y) \in f$, where $y \neq g(n, x)$, then the set $f' = f - \{(\mathbf{S}(n), y)\}$ is again in A: condition (1) is satisfied since $\mathbf{S}(n) \neq 0$, and condition (2) is satisfied by combining (PA2) with induction hypothesis. As before, this is a contradiction, so the statement holds for $\mathbf{S}(n)$. This completes the proof by induction.

Now that f is proven to be a function, the definition of f immediately implies its two required equations. It remains to prove uniqueness. Suppose f, f' are two functions satisfying the properties in the theorem. We will show by induction that $f(n) = f'(n)$ for all $n \in \mathbb{N}$. This holds for $n = 0$ by the hypothesis $f(0) = f'(0) = a$. Suppose it holds for a given n, so $f(n) = f'(n)$, then

$$f(\mathbf{S}(n)) = g(n, f(n)) = g(n, f'(n)) = f'(\mathbf{S}(n))$$

This concludes the induction step and hence the proof of uniqueness. □

As promised, we apply it to the Ackermann function.

Theorem 1.9. *The Ackermann function $A : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is well-defined.*

Proof. The proof is based on the concept of *currying*. Take $S = [\mathbb{N} \to \mathbb{N}]$, the set of all functions from $\mathbb{N}$ to $\mathbb{N}$, then we view the Ackermann function as a function

$$C : \mathbb{N} \to [\mathbb{N} \to \mathbb{N}],\ C(m)(n) = A(m, n)$$

Here, $C(m)$ is a function on $\mathbb{N}$, so we need to specify it by telling you what its value on each natural number is, hence the notation $C(m)(n)$.

We now define a function $R : [\mathbb{N} \to \mathbb{N}] \to [\mathbb{N} \to \mathbb{N}]$. Its input is a function f. The output is a function $g : \mathbb{N} \to \mathbb{N}$ such that

$$g(0) = f(1),\ g(n + 1) = f(g(n))$$

The existence of g is an immediate consequence of the recursion theorem. Informally, $g(n)$ is the $(n + 1)$-th iteration of f evaluated at 1.

By the definition of the Ackermann function,

$$C(0) = (n \mapsto n + 1),\ C(m + 1) = R(C(m))$$

Here, $(n \mapsto n + 1)$ is an abbreviation for the function on $\mathbb{N}$ that sends n to $n + 1$. Applying the recursion theorem with $S = [\mathbb{N} \to \mathbb{N}]$ shows that C is a well-defined function. Therefore, $A(m, n) = C(m)(n)$ is well-defined. □

This proof might be confusing, probably because it deals with higher order functions (functions on function spaces). It is helpful to sit down and unwind exactly what each formula is stating.

1.6. **Other numbers.** For this course, we still need to know about the integers $\mathbb{Z}$ and the rational numbers $\mathbb{Q}$. We will eventually define them using $\mathbb{N}$, but it is once again much more useful to know their properties. Some subsets of these properties are shared by other interesting objects, so it is beneficial to give them a name too. This section also serves as an entry point into *abstract algebra*.

Definition. A *ring* is a set R with two binary operations $+, \cdot$ and two distinguished elements $0,\ 1$, such that for all $a, b, c \in R$, the following properties hold

- $a + (b + c) = (a + b) + c$, $a + 0 = a$, $a + b = b + a$.
- There exists $-a \in R$ such that $a + (-a) = 0$.
- $a(bc) = (ab)c$, $a \cdot 1 = 1 \cdot a = a$.
- $a(b + c) = ab + ac$, $(b + c)a = ba + ca$.

It is *commutative* if in addition, $ab = ba$ for all $a, b \in R$.

Definition. An element $a \in R$ is *invertible* if there exists $a^{-1} \in R$ such that $aa^{-1} = a^{-1}a = 1$.

A commutative ring R is a *field* if $0 \neq 1$ and every non-zero element is invertible.

From experience, we know that $\mathbb{Z}$ is a commutative ring and $\mathbb{Q}$ is a field. Some other examples of rings are: the set of polynomial functions, the set of all $n \times n$ matrices, the powerset of a set (see exercise), and for any n, the set of integers modulo n (see Chapter 3).

Exercise. Let X be a set. Its powerset $\mathcal{P}(X)$ is the set of all subsets of X. This has a ring structure with

$$A + B = \{z \in X \mid z \text{ is in exactly one of } A \text{ and } B\}$$

and $A \cdot B = A \cap B$. Observe that $A + A = 0$ for all A.

The ring axioms imply many of the usual properties.

Example. If R is a ring, then $0 \cdot x = 0$ for all $x \in R$.

Formal proof. For all $x \in R$,

$$\begin{aligned} 0 \cdot x &= (0 + 0) \cdot x && \text{(Additive identity)} \\ &= 0 \cdot x + 0 \cdot x && \text{(Distributivity)} \end{aligned}$$

Add $(-0 \cdot x)$ to both sides.

$$\begin{aligned} 0 \cdot x + (-0 \cdot x) &= 0 && \text{(Additive inverse)} \\ (0 \cdot x + 0 \cdot x) + (-0 \cdot x) &= 0 \cdot x + (0 \cdot x + (-0 \cdot x)) \\ & && \text{(Associativity)} \\ &= 0 \cdot x && \text{(Additive inverse)} \end{aligned}$$

so $0 = 0 \cdot x$. □

This is a formal proof written out in full detail. It is basically a string rewrite system. In practice, we tend to skip most of the explanatory steps for greater readability.

Human proof. Let $x \in R$, then $0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x$, so we can add $-0 \cdot x$ to both sides to obtain $0 = 0 \cdot x$. □

Of course, just saying $\mathbb{Z}$ is a ring does not characterize $\mathbb{Z}$. We know that $\mathbb{Z}$ has an ordering that is preserved by addition and multiplication. These properties do not follow from the ring axioms. They are additional axioms that need to be included.

Definition. A commutative ring R is an *ordered ring* if it has a total ordering $\leq$ such that

- For all $a, b, c \in R$, if $a \leq b$, then $a + c \leq b + c$.
- For all $a, b \in R$, if $0 \leq a$ and $0 \leq b$, then $0 \leq ab$.

Once again, both $\mathbb{Z}$ and $\mathbb{Q}$ are ordered rings, but there are lot fewer natural examples. This is getting closer to completely characterizing $\mathbb{Z}$.

Exercise. Let $\mathbb{C}$ be the field of complex numbers, then $\mathbb{C}$ is not an ordered ring.

To distinguish $\mathbb{Z}$ from $\mathbb{Q}$, we want to say something like $\mathbb{Z}$ is discrete. Maybe we should include as an axiom that $(\nexists x)(0 < x < 1)$. The resulting objects are called *discretely ordered rings*. It might be hard to come up with examples of those other than $\mathbb{Z}$, but they exist.

Exercise. The set $\operatorname{Poly}(\mathbb{Z})$ of polynomial functions on $\mathbb{Z}$ with integer coefficients can be ordered by $f \leq g$ if $f(n) \leq g(n)$ for all sufficiently large n. This is a discretely ordered ring.

This is not like $\mathbb{Z}$ because the function $x \mapsto x$ is larger than any natural number. Maybe we can throw in something like the archimedean property. I can't think of any naturally occurring example of such structures other than $\mathbb{Z}$, but they exist. This is a consequence of the Löwenheim–Skolem theorem for first order logic. To characterize $\mathbb{Z}$, we need a much more powerful statement.

Definition. The integer $\mathbb{Z}$ is a totally ordered ring satisfying the following version of the well-ordering property:

> Let S be a subset of $\mathbb{Z}$. If S is non-empty and bounded below, then S has a least element.

This could have been the starting point of the chapter, and $\mathbb{N}$ is defined as the non-negative part of $\mathbb{Z}$. This is the approach followed by the textbook. We began with the Peano axioms to highlight the fundamental role of recursion and induction. Moreover, the ideas introduced will play a role in our discussion of Hilbert's 10th problem later in the course.

There are two outstanding issues.

- Does $\mathbb{Z}$ exist?
- What is $\mathbb{Q}$?

There is a nice characterization for $\mathbb{Q}$ using slightly more ring theory, but instead, we will resolve both problems by actually *defining* $\mathbb{Z}$ and $\mathbb{Q}$ using $\mathbb{N}$. Intuitively, we get integers by throwing in negative numbers, then we get the rational numbers by throwing in reciprocals of non-zero integers. In very fancy language, the two steps are taking *Grothendieck completion* and *field of fractions.* We will very briefly give the definitions, but they are not important for this course.

Definition*. On the set $\mathbb{N} \times \mathbb{N}$, define an equivalence relation $\sim$ by $(a, b) \sim (c, d)$ if $a + d = b + c$. The set of integers $\mathbb{Z}$ is the quotient set $(\mathbb{N} \times \mathbb{N})/\sim$. Arithmetic operations are defined by

$$(a,b)+(c,d)=(a+c,b+d)$$
$$(a,b)\cdot(c,d)=(ad+bc,ac+bd)$$

Intuitively, the equivalence class $[(a,b)]$ represents $a-b$.

Definition*. On the set $\mathbb{Z}\times(\mathbb{Z}-\{0\})$, define an equivalence relation $\sim$ by $(a,b)\sim(c,d)$ if $ad=bc$. The set of rational numbers $\mathbb{Q}$ is the quotient set $(\mathbb{Z}\times\mathbb{Z})/\sim$. Multiplication is defined component-wise. Arithmetic operations are defined by

$$(a,b)+(c,d)=(ad+bc,bd)$$
$$(a,b)\cdot(c,d)=(ac,bd)$$

Intuitively, the equivalence class $[(a,b)]$ represents $\frac{a}{b}$.

Remark.* The definitions contains many claim implicitly. For the first one for $\mathbb{Z}$, they include the following statements.

- The relation $\sim$ is an equivalence relation.
- It respects addition: if $(a,b)\sim(c,d)$, then $(a,b)+(n,m)\sim(c,d)+(n,m)$.
- It respects multiplication: if $(a,b)\sim(c,d)$, then $(a,b)\cdot(n,m)\sim(c,d)\cdot(n,m)$ for all $n,m\in\mathbb{N}$.

They are all easy to check, but it is important to know this step is required. When we do congruences, we will work out a similar example in detail.

1.7. ***Peano arithmetic*.** It is somehow unsatisfactory to axiomatize the natural numbers using set theory, since natural numbers feel much simpler than arbitrary sets. When we prove things about rings, sets don't have to enter the picture. On the other hand, the proof of the recursion theorem very strongly depended on statements about existence of sets beyond what the axioms covered.

There are a few systems to axiomatize the natural numbers in a purely formal way (the official term for what we have in mind is *first order logic*). A popular one is the Peano arithmetic, or PA.[2]

Axiom. The natural numbers have a constant 0, a unary operation $\mathbf{S}$, two binary operations $+$ and $\times$. They satisfy the following 6 axioms

(P1) $(\forall x)(\mathbf{S}(x)\neq 0)$.
(P2) $(\forall x,y)(\mathbf{S}(x)=\mathbf{S}(y)\to x=y)$.
(P3) $(\forall x)(x+0=x)$.
(P4) $(\forall x,y)(x+\mathbf{S}(y)=\mathbf{S}(x+y))$.
(P5) $(\forall x)(x\times 0=0)$.
(P6) $(\forall x,y)(x\times\mathbf{S}(y)=x\times y+x)$.

Plus for each "statement" $P(x)$, the sentence

$$[P(0)\wedge(\forall x)(P(x)\to P(\mathbf{S}(x)))]\to(\forall x)(P(x))$$

is an axiom.

The definition of a "statement" is technical, but essentially it consists of all sentences that can be formed using the operations we have defined and standard logical connectives (and, or, not, implies). The idea is we are not allowed to talk

[2] We have been very carefully calling our axioms the Peano Axioms.

about subsets, so we will add an instance of the induction axiom for each definable subset we can talk about.

Purely by reasoning from these axioms, we can recover a huge chunk of number theory. As a simple example, our proof of associativity still holds, since we can take $P(x)$ to be the following statement

$$(\forall a, b)((a + b) + x = a + (b + x))$$

and use the corresponding induction axiom in our proof. However, the proof of the recursion theorem does not carry through, since the definition of f involved considering all subsets of $\mathbb{N} \times \mathbb{N}$. As an example of the difficulty of recursion, think about how you would write down a sentence $E(x, y)$ using only $+$ and $\times$ which is true exactly when $2^x = y$.

The proof of the recursion theorem is based on the one by well-founded induction we sketched, but it is much more involved. The idea is to code finite sequences of natural numbers using a single number, and this coding is based on a clever trick with the Chinese remainder theorem. This requires developing a substantial portion of the first half of this course without using recursion, which we will implicitly do.

Exercise. Define the term $\mathtt{p}(x, y) := (x + y) \times (x + y) + x$, then PA proves that

For all x, y, u, v, $\mathtt{p}(x, y) = \mathtt{p}(u, v) \implies x = u, y = v$.

Therefore, the number $\mathtt{p}(x, y)$ can be used to represent the ordered pair (x, y).

In this formal theory, we absolutely needed to know addition and multiplication. Even just having addition is not enough: it is provable that you cannot define multiplication using only axioms (P1)–(P4)+induction. This weak system is called the Presburger arithmetic.

Once we have recursion (after Chinese remainder theorem), everything in this course can be formalized in PA. This seems great, so now we can talk about some results which at first seem very disturbing.

(1) Löwenheim–Skolem theorem: There are many other models of PA beyond the standard $\mathbb{N}$.
(2) Gödel's first incompleteness theorem: There are statements that can't be proven or disproven in PA.
(3) Gödel's second incompleteness theorem: If PA is consistent, then it cannot prove its own consistency.
(4) Entscheidungsproblem: There is no program to decide if a given statement can be proven in PA.
(5) Growth hierarchy: There are computable functions that PA cannot prove is well-defined.

In comparison, the Presburger arithmetic is known to be complete and decidable (but the decision algorithm is provably at least double exponential time in the length of the statement). Incidentally, this proves that you cannot define multiplication using addition alone.

We now briefly explain each of the facts:

- Fact (1) is not too bad. It is a standard property of any first-order theory. The resulting non-standard models are hard to describe (in a very precise sense, they are not computable), but they have applications. There are even uncountable models.

- Fact (2) is not too surprising: the ring axioms cannot prove or disprove $1 + 1 = 0$ because there are examples of rings satisfying this (e.g. $\mathbb{Z}/2\mathbb{Z}$) and examples of rings that don't (e.g. $\mathbb{Z}$). What is surprising is that (2) continues to hold no matter how many axioms we add to PA (with an important technical condition related to (3) and (4)).
- Facts (3) and (4) represent an essential obstacle to an idealistic foundations of mathematics. The reason for facts (2)–(4) is that PA is too powerful. Very roughly, we can reason about computation in PA, and the halting problem is undecidable. Fundamentally, the ability to prove the recursion theorem is responsible for these results.
- Fact (5) is the exact opposite problem: PA is too weak. The Goodstein function is a classical example. It outputs the length of a sequence which we can prove always terminates. However, this proof uses set theory, and it cannot be done in PA.

  The obstacle is that Goodstein's function is defined recursively on trees, but PA cannot prove that the ordering used is actually a well-ordering, so it cannot prove the function is always defined. Adding an axiom that the order is a well-ordering actually proves the consistency of PA.

As a consequence of Hilbert's 10th problem, there is a single explicit polynomial equation (in at most 53 variables) which has an integer solution if and only if set theory is inconsistent. Since most of modern mathematics is based on set theory, we just cannot prove it has no solutions. We can add it as an axiom, but then by changing one parameter in the equation, we get a new equation that can't be proven to have no solution. The point is that undecidability issues already show up at a very simple arithmetic level.

In theory, there is a chance that a famous conjecture like the twin prime conjecture is undecidable in PA, but people don't really expect that. Somehow the questions a working number theorists care about feel different from the known undecidable statements (we already have enough trouble understanding 2-variable equations). In fact, it is conjectured that much weaker axioms than PA can recover all known results in number theory (this would include Fermat's last theorem). These kinds of questions are known as reverse mathematics.

In summary, don't worry about it.

2. Divisibility and primes

The previous chapter is not typically considered a part of number theory. From this point on, we will forget about the subtleties about foundations and actually look at number theory. In particular, we will assume all the properties of the integers and rationals that you are familiar with (in sophisticated terms, $\mathbb{Z}$ is a commutative ordered ring with the well-ordering property, and $\mathbb{Q}$ is the field of fractions of $\mathbb{Z}$).

2.1. **Division.**

Definition. Let n, m be two integers, then we say n divides m (or n is a divisor of m) if there exists $d \in \mathbb{Z}$ such that $m = nd$. This is written as $n|m$. If n does not divide m, then we write $n \nmid m$.

Exercise. Everything divides 0. 1 divides everything.

Lemma 2.1. *If $a|b$ and $b \neq 0$, then $|a| \leq |b|$.*

Proof. Suppose $b = na$, then $|b| = |n|\,|a|$. Since $b \neq 0$, we have $n \neq 0$, so $|n| \geq 1$. Therefore, $|b| \geq 1 \cdot |a| = |a|$. □

The following proposition summarizes the basic properties of divisibility.

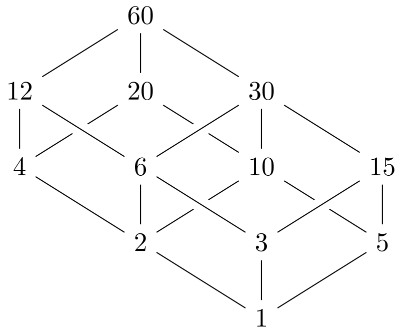
Proposition 2.2. *For all $a, b, c \in \mathbb{Z}$, we have*

- *Reflexivity: $a|a$.*
- *Anti-symmetry: If $a|b$ and $b|a$, then $a = \pm b$.*
- *Transitivity: If $a|b$ and $b|c$, then $a|c$.*
- *Addition: If $a|b$ and $a|c$, then $a|b + c$.*
- *Multiplication: If $a|b$, then $a|bc$.*

Proof. We will prove anti-symmetry. The others are easier and left as exercises.

Suppose $a|b$ and $b|a$, then there exists d and d' such that $a = bd$ and $b = ad'$. These together imply $a = add'$. By cancellation, $dd' = 1$. The previous lemma implies $|d|, |d'| \leq 1$. If $d = 0$, then $a = 0$, so $b = ad' = 0$. Similarly, if $d' = 0$, then $a = b = 0$. If d and d' are both non-zero, then $d, d' = \pm 1$, which implies $a = \pm b$. □

In particular, if we restrict to the natural numbers, then the first three properties exactly tells us that the divisibility relation is a *partial ordering.* It's beneficial to think of them as a more exotic way of ordering natural numbers. Unlike the ordinary inequality by size, it's not the case that any two elements are comparable, so for example $4 \nmid 6$ and $6 \nmid 4$. The following graph visualizes all divisors of 60.



The idea is that each line indicates the bottom element divides the top one, and there is no element between them. The graph is not something 3D, but if you squint a little it looks exactly like a cube cut in half. This is not an accident. It is in fact the fundamental theorem of arithmetic, the end goal of this chapter.

Observe that by the cancellation property, if $a|b$, then there is a *unique* integer d such that $b = ad$. We can call this d the quotient of b by a and write $d = \frac{b}{a}$. Even if $a \nmid b$, there is still some kind of approximation given by the following result.

Proposition 2.3 (Division algorithm)**.** *If $a, b \in \mathbb{Z}$ and $b \neq 0$, then there exists unique integers q and r such that $a = qb + r$ and $0 \leq r < |b|$. This number r is called the* remainder *when a is divided by b.*

Proof. We first prove uniqueness. Suppose $a = qb + r = q'b + r'$, then $(q - q')b = r' - r$. If $q = q'$, then $r = r'$. Otherwise, the left hand side has absolute value at least $|b|$, but $|r' - r| < |b|$. This is a contradiction.

For existence, consider the set $S = \{a - qb \,|\, q \in \mathbb{Z}\} \cap \mathbb{N}$. An application of Lemma 1.6 shows that $S \neq \emptyset$, so it has a least element, which we call r. By construction, we have an equation $a = qb + r$ and $r \geq 0$. Suppose $r \geq b$, then $r - |b| \in S$ and $r - |b| < r$. This contradicts the choice of r, so $r < |b|$. □

2.2. **The greatest common divisor.**

Theorem 2.4. *Let $n, m \in \mathbb{Z}$. There is a unique non-negative integer d such that*

(1) $d|n$ and $d|m$.
(2) If $a|n$ and $a|m$, then $a|d$

We say d is the greatest common divisor *of n and m, and write $d = \gcd(n, m)$ (many books may write $d = (n, m)$ instead).*

In other words, $\gcd(n, m)$ is the "largest" natural number among all common divisors of m and n, except "largest" is defined with respect to the partial order of divisibility. To drive home the point further, $\gcd(n, m)$ is the "largest" one among all natural numbers "smaller" than both n and m, ordered by divisibility instead of the usual size. In terms of order theory, $\gcd(n, m)$ is the *meet* of n and m.

Example.

(1) If $n = 6$ and $m = 10$, then the common divisors of n and m are $\{\pm 1, \pm 2\}$. The "largest" natural number in this set by divisibility is 2, so $\gcd(6, 10) = 2$. It also happens that 2 is the largest element by size.
(2) $\gcd(0, 5) = 5$, since the set of common divisors is $\{\pm 1, \pm 5\}$.
(3) $\gcd(0, 0) = 0$. Every integer divides 0, so the set of common divisors is $\mathbb{Z}$. The "largest" element of $\mathbb{Z}$ by divisibility is 0, again since all integers divide 0.

Proof of the theorem. The easy part is uniqueness: suppose d and d' both satisfy the condition, then $d|d'$ and $d'|d$. By anti-symmetry, $d = \pm d'$. But $d, d' \geq 0$, so $d = d'$. For existence, we prove something stronger. Define the sets

$$I := \{nx + my \,|\, x, y \in \mathbb{Z}\}, \quad I_+ = \{x \in n\mathbb{Z} + m\mathbb{Z} \,|\, x > 0\}$$

The set I is closed under addition, i.e. if x and y are in the set, then so is $x + y$. Moreover, it is closed under multiplication by an *arbitrary* integer. If $n = m = 0$, then we have seen that $\gcd(n, m) = 0$. Otherwise, at least one of $|n|$ and $|m|$ is positive, so $I_+ \neq \emptyset$. By the well-ordering property, I_+ has a least element, which we call d. We claim that d satisfies the two required conditions.

By the definition of d, there exists $x, y \in \mathbb{Z}$ such that $d = nx + my$. If $a|n$ and $a|m$, then $a|nx + my = d$. This proves condition (2). For condition (1), by division with remainder, $n = qd + r$, where $q, r \in \mathbb{Z}$ and $0 \leq r < d$. The properties of I shows that $r = n - qd \in I$. Since d is minimal in I_+, we must have $r = 0$, which implies $d|n$. Similarly, $d|m$, proving condition (1). □

Corollary 2.5 (Bézout's theorem)**.** *For any $n, m \in \mathbb{Z}$, there exists $x, y \in \mathbb{Z}$ such that $\gcd(n, m) = nx + my$.*

Proof. This is an immediate consequence of our proof that the GCD exists. □

Example. We have seen that $\gcd(6, 10) = 2$. By inspection, $2 = 6 \times 2 + 10 \times (-1)$. Of course, there are other solutions such as $2 = 6 \times 7 + 10 \times (-4)$.

A typical problem with proof by the well-ordering property is that the proof is non-constructive. We know the set has a least element, but nothing tells us how to identify it. Later in this chapter, we will introduce Euclid's algorithm, which allows us to efficiently compute the GCD and the integers x, y from Bézout's theorem.

Recall that in the ordering by divisibility, $\gcd(n, m)$ is the "largest" of the elements "smaller" than both n and m. We can flip this and consider the "smallest" of the elements "larger" than both n and m. In order theory, this is called the *join* of m and n. In number theory, it is called the least common multiple.

Theorem 2.6. *Let $n, m \in \mathbb{Z}$. There is a unique non-negative integer ℓ such that*

(1) $n|\ell$ and $m|\ell$.

(2) If $n|a$ and $m|a$, then $\ell|a$

We say ℓ is the least common multiple *of n and m, and write $\ell = \operatorname{lcm}(n, m)$ (some other books may write $\ell = [n, m]$ instead).*

Moreover, we have a formula

$$\operatorname{lcm}(n, m) = \begin{cases} \frac{nm}{\gcd(n,m)} & \text{If } n, m \neq 0 \\ 0 & \text{Otherwise} \end{cases}$$

Proof. If $n = 0$ or $m = 0$, then it is clear that the formula holds. Otherwise, $mn \neq 0$. Let A be the set of positive divisors of mn. The function

$$\iota : A \to A, \ d \mapsto \frac{nm}{d}$$

is an order-reversing bijection, so it exchanges the definitions of join and meet. Therefore, the join exists, and

$$\operatorname{lcm}(n, m) = \iota(\gcd(\iota(m), \iota(n))) = \frac{mn}{\gcd(n, m)}$$ □

Warning. The relation $\operatorname{lcm}(n, m) \gcd(n, m) = nm$ only holds in general for two numbers. In other words, it is not the case that $\operatorname{lcm}(S) \gcd(S) = \prod_{s \in S} s$ for all S unless $\#S = 2$.

Having proven the existence of the greatest common divisor, it's time to look at some simple properties.

Proposition 2.7.

(1) $\gcd(a, b) = \gcd(b, a)$.

(2) *For any* $n \in \mathbb{Z}$, $\gcd(a, b) = \gcd(a + nb, b)$.
(3) *If* $d|a$ *and* $d|b$, *then* $\gcd\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{\gcd(a,b)}{d}$.

Proof. For each property, there are many ways of proving it. We will illustrate three methods.

(1) From the proof of Theorem 2.4, $\gcd(a, b)$ depends only on the set $a\mathbb{Z} + b\mathbb{Z}$. However, by the commutativity of addition, $a\mathbb{Z} + b\mathbb{Z} = b\mathbb{Z} + a\mathbb{Z}$, so $\gcd(a, b) = \gcd(b, a)$.
(2) If $d|a$ and $d|b$, then $d|a+nb$. Conversely, if $d|a+nb$ and $d|b$, then $d|a$. Therefore, the pairs (a, b) and $(a + nb, b)$ share the same set of common divisors, so their GCDs must be equal.
(3) We check that $\frac{\gcd(a,b)}{d}$ satisfies the characterizing properties of $\gcd\left(\frac{a}{d}, \frac{b}{d}\right)$. First, since $\gcd(a, b)|a$, we have $\frac{\gcd(a,b)}{d} | \frac{a}{d}$, and similarly for b. This proves condition (1). Suppose $d'|\frac{a}{d}$ and $d'|\frac{b}{d}$, then $d'd|a$ and $d'd|b$. It follows that $d'd|\gcd(a, b)$, so $d'|\frac{\gcd(a,b)}{d}$, as required. □

Bézout's theorem can be viewed as a very simple Diophantine equation: find integers x and y such that $nx + my = \gcd(n, m)$. The theorem states that there are solutions. Euclid's algorithm, which we will cover later, gives a procedure to find a solution. The next question is then to find all solutions. Clearly, if (x, y) is a solution, then $(x + km, y - kn)$ is a solution for all $k \in \mathbb{Z}$, and one may wonder if this procedure produces all the solutions. Any investigation into this question will quickly run into the need for the following lemma.

Lemma 2.8 (Euclid's lemma)**.** *If* $a|bc$ *and* $\gcd(a, b) = 1$, *then* $a|c$.

Proof. By Bézout's theorem, we can find $x, y \in \mathbb{Z}$ such that $ax+by = 1$. Multiplying by c gives $acx + bcy = c$. By hypothesis, $a|bc$, so $a|acx + bcy$. Therefore, $a|c$. □

The condition $\gcd(a, b) = 1$ is clearly necessary: $4|6 \times 10$ but $4 \nmid 6$ and $4 \nmid 10$. This condition is so useful that we have a name for it.

Definition. If $\gcd(a, b) = 1$, then we say a and b are *coprime*.

In the last part of this section, we briefly indicate how the definitions of GCD and LCM generalize from a pair of numbers to an arbitrary set of numbers. The definitions are exactly as expected.

Definition. Let $S \subseteq \mathbb{Z}$, then the greatest common divisor of S, denoted by $\gcd(S)$ is the unique non-negative integer d such that

(1) $d|s$ for all $s \in S$.
(2) If $a|s$ for all $s \in S$, then $a|d$.

Similarly, we can define $\operatorname{lcm}(S)$.

For people who have taken real analysis, over the real numbers with its normal ordering, the analogue of $\gcd(S)$ is the infimum (greatest lower bound), and the analogue of $\operatorname{lcm}(S)$ is the supremum (least upper bound).

Again, uniqueness is clear, so we only need to prove existence.

Theorem 2.9. *The GCD and LCM exists for an arbitrary* S.

Proof. We only sketch the proof. For the existence of GCD, consider the set

$$\langle S\rangle = \left\{ \sum_{i=1}^{k} n_i s_i \,|\, k \in \mathbb{N}, n_i \in \mathbb{Z}, s_i \in S \right\}$$

This is the set of all *finite* integral combinations of elements of S (infinite sums do not make sense). The same proof as before shows existence and a version of Bézout's theorem. For LCM, if S is infinite or contains 0, then the only common multiple of S is 0, so $\operatorname{lcm}(S) = 0$. Otherwise, we can argue by duality using the integer $\prod_{s\in S} s$. □

Remark. The set $\langle S\rangle$ is the *ideal* generated by S, so it is the smallest set containing S that is closed under addition and closed under multiplication by an arbitrary element. The proof of the existence of GCD actually shows that any ideal of $\mathbb{Z}$ has the form $d\mathbb{Z}$ for some $d \in \mathbb{Z}$. We say that $\mathbb{Z}$ is a *principal ideal domain.* This d is the GCD of the ideal. The main logical progression in this section is

Division algorithm $\Longrightarrow$ Principal ideal domain $\Longrightarrow$ Bézout's theorem $\Longrightarrow$ GCD exists

We will later complete the implication that GCD exists $\Longrightarrow$ Fundamental theorem of arithmetic.

The terminology ideal comes from "ideal divisor". There are many rings R where $\gcd(n, m)$ can fail to exist, but we can always define the set $nR + mR$ and think of it as the set of elements divisible by a non-existent "ideal number" $\gcd(n, m)$. In modern (algebraic) number theory, the role of GCD is mostly replaced by ideals.

Definition. A set S is coprime if $\gcd(S) = 1$. It is *pairwise coprime* if $\gcd(a, b) = 1$ for all $a, b \in S$ such that $a \neq b$.

Example. The finite set $\{6, 10, 15\}$ is coprime but not pariwise coprime.

2.3. **Euclid's algorithm.**

The idea of Euclid's algorithm is simple: $\gcd(a, b) = \gcd(a + nb, b)$ for all $n \in \mathbb{Z}$, so choosing n to minimize $a + nb$ reduces the complexity of the problem. This minimizer is exactly the remainder when a is divided by b.

Notation. The remainder when a is divided by b will be denoted by $\texttt{rem}(a, b)$.

Theorem 2.10. *Let $a, b \in \mathbb{Z}$. Inductively define sequences $a_0, a_1, \cdots$ and $b_0, b_1, \cdots$ by the procedure*

$$a_0 = a,\ b_0 = |b|$$
$$a_{i+1} = b_i,\ b_{i+1} = \texttt{rem}(a_i, b_i)$$

There exists a k such that $b_k = 0$, causing the sequences to terminate, and we have $\gcd(a, b) = a_k$.

Proof. We prove by induction that $\gcd(a, b) = \gcd(a_i, b_i)$ for all $i \geq 0$. This is obvious if $i = 0$. For the inductive step, we need to show that $\gcd(n, m) = \gcd(m, \texttt{rem}(n, m))$ for all m, n. This follows from Proposition 2.7 and the definition of $\texttt{rem}$, so we are done.

Observe that by construction, $b_{i+1} < b_i$ for all i, so since $\mathbb{N}$ is well-ordered, the sequence must terminate at 0. Suppose $b_k = 0$, then $\gcd(a, b) = \gcd(a_k, b_k) = a_k$, as required. □

Example. We want to compute $\gcd(226, 710)$.

$$\begin{aligned}
226 &= 0 \cdot 710 + 226 &\qquad (a_0, b_0) &= (226, 710)\\
710 &= 3 \cdot 226 + 32 & (a_1, b_1) &= (710, 226)\\
226 &= 7 \cdot 32 + 2 & (a_2, b_2) &= (226, 32)\\
32 &= 16 \cdot 2 + 0 & (a_3, b_3) &= (32, 2)\\
&- - - & (a_4, b_4) &= (2, 0)
\end{aligned}$$

So the sequence terminates at $k = 4$, with $\gcd(226, 710) = 2$.

Going backwards,

$$\begin{aligned}
2 &= 1 \cdot 226 - 7 \cdot 32\\
&= 226 - 7 \cdot (710 - 3 \cdot 226)\\
&= 22 \cdot 226 - 7 \cdot 710
\end{aligned}$$

This is an explicit version of Bézout's theorem.

If you know programming, it is a useful exercise to write a program to compute the GCD and write it as a linear combination. We now perform a worst-case analysis of the algorithm.

Theorem 2.11. *Euclid's algorithm on input (n, m) terminates in $O(\log n)$ steps.*

Proof. If $n < m$, then the first step of Euclid's algorithm is just $n = 0 \cdot m + n$, and we have $(a_1, b_1) = (m, n)$. This adds 1 to the total length, which does not change the asymptotic behaviour. Therefore, we may assume $n \geq m$. In particular, $a_k \geq b_k$ for all k.

Suppose Euclid's algorithm terminates on the k-th step, then $(a_k, b_k) = (d, 0)$, where $d = \gcd(n, m)$. We will now prove by induction on i that $a_{k-i} \geq dF_{i+1}$ and $b_{k-i} \geq dF_i$. Suppose this is known for i, then

$$\begin{aligned}
b_{k-i-1} &= a_{k-i} \geq dF_{i+1}\\
a_{k-i-1} &= qb_{k-i-1} + b_{k-i}\\
&\geq b_{k-i-1} + b_{k-i}\\
&\geq dF_{i+1} + dF_i = dF_{i+2}
\end{aligned}$$

In the above expression, q is the quotient when a_{k-i-1} is divided by b_{k-i-1}, and $q \geq 1$ since $a_{k-i-1} \geq b_{k-i-1}$. This concludes the induction. It follows that $n \geq dF_{k+1}$ and $m \geq dF_k$.

In Lemma 1.8, we showed by induction that $F_n \geq (1.1)^n$ for all $n \geq 3$, so if Euclid's algorithm terminates on the k-th step, then $n \geq d \cdot (1.1)^{k+1} > 1.1^k$. Rearranging this gives $k < \log_{1.1} n = O(\log n)$. □

Note that an integer n has $O(\log n)$ digits, so $\log n$ is the actual size of input. Therefore, Euclid's algorithm terminates in *linear* time and not *logarithmic* time. Accounting for the fact that division takes $O((\log n)^2)$ bitwise operations (the long division method), we see that Euclid's algorithm is cubic time.

Exercise. Observe that the numbers occurring in Euclid's algorithm decrease in size rapidly. Use this to show that Euclid's algorithm only takes $O((\log n)^2)$ bitwise operations in total.

2.4. **Prime numbers and factorization.** We give two definitions that turn out to be equivalent.

Definition. A positive integer p is called *irreducible* if $p \neq 1$ and $p = ab$ with a, b positive integers implies either $a = 1$ or $b = 1$. Otherwise, it is called *composite.*

Definition. A positive integer p is *prime* if $p \neq 1$ and $p|ab$ implies $p|a$ or $p|b$.

The first definition might be closer to your intuitive notion of prime numbers: they are supposed to be building blocks of the integers. However, irreducible is a more intuitive word to describe this. The terminologies we have chosen coincide with the ones found in modern abstract algebra.

Our first goal is to show that the definitions are redundant.

Proposition 2.12. *Primes are irreducible.*

Proof. Let p be a prime. Suppose a and b are positive integers such that $p = ab$, then clearly $p|ab$, so $p|a$ or $p|b$. Without loss of generality, suppose $p|a$. On the other hand, $a|p$, so $p = a$ and $b = 1$. □

Theorem 2.13. *Irreducible numbers are prime*

Proof. Let p be an irreducible integer. Suppose a is an integer such that $p \nmid a$, then we claim that $\gcd(p, a) = 1$. Indeed, the only possible divisors of p are ± 1 and $\pm p$. Since $p \nmid a$, the only common divisors are ± 1. Now suppose $p|ab$ and $p \nmid a$, then by Euclid's lemma (Lemma 2.8), $p|b$. □

The reason we introduced both definitions is that they play different roles in our proof of unique factorization. In more general rings, it is no longer true the irreducible elements are primes, and this is why unique factorization fails. A key insight that started modern algebraic number theory is that the situation can be salvaged by talking about ideals instead of elements, and in that setting primes play a central role.

The main theorem of this section is the following, which says that every positive integer n is a product of primes in an essentially unique way.

Theorem 2.14 (Fundamental theorem of arithmetic)**.** *Every positive integer n greater than 1 is a product of primes. This expression is unique up to reordering the factors.*

Proof. Existence: we prove this by induction on n. If n is a prime, then we are done. Otherwise, we can write $n = ab$, where $a, b > 1$. By induction hypothesis, each of a and b is a product of primes, so the same holds for n.

Uniqueness: we again prove this by induction on n. Suppose $n = p_1 \cdots p_k = q_1 \cdots q_l$, where all p_i and q_j are primes. Since $p_1 | q_1 \cdots q_l$ and p_1 is prime, there exists j such that $p_1 | q_j$. But q_j is prime, and hence irreducible, so $p_1 = q_j$. After dividing by p_1, we are left with two factorizations of $\frac{n}{p_1}$. By induction hypothesis, they are the same after reordering. It follows that the original two factorizations of n are also equivalent. □

As an immediate corollary, we can make the following definition.

Definition. Let p be a prime and $n \geq 1$, then $v_p(n)$ is the number of times p appears in the prime factorization of n.

Corollary 2.15. *Any $n \geq 1$ can be written as*

$$n = \prod_p p^{v_p(n)}$$

where the product is over all primes, and all but finitely many terms are 1.

In this form the content of the fundamental theorem of arithmetic is that the exponents $v_p(n)$ are well-defined. We now give a list of properties.

Proposition 2.16. *Let p be a prime, then for all $m, n \geq 1$,*

(1) $v_p(n)$ *is the largest integer* $k \geq 0$ *such that* $p^k | n$.
(2) $v_p(mn) = v_p(m) + v_p(n)$.
(3) $v_p(m+n) \geq \min(v_p(m), v_p(n))$, *with equality if* $v_p(m) \neq v_p(n)$.
(4) $v_p(\gcd(m,n)) = \min(v_p(m), v_p(n))$, $v_p(\operatorname{lcm}(m,n)) = \max(v_p(m), v_p(n))$

Proof. Both (1) and (2) are immediate consequences of the corollary. Parts (3) and (4) are on the homework. □

We can extend the function $v_p(\cdot)$ to the rational numbers: for $x = \pm\frac{a}{b}$ with $a, b \geq 1$, we can set $v_p(x) = v_p(a) - v_p(b)$. Part (2) of the proposition shows that this is independent of the way we choose a and b. Moreover, since everything divides 0, it is reasonable to say $v_p(0) = \infty$. This gives a function $v_p : \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ for each prime p.

To make things look more familiar, let's define

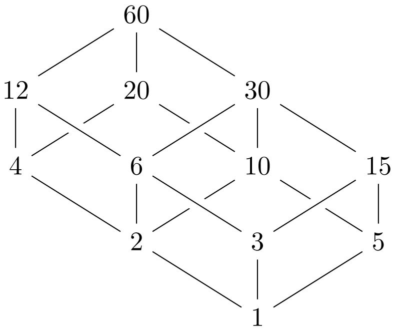
$$|x|_p := p^{-v_p(x)}$$

and define $|0|_p = 0$. It's easy to see that parts (2) and (3) of the proposition translates to the following three properties: for all $x, y \in \mathbb{Q}$,

(1) $|x|_p \geq 0$, with equality exactly when $x = 0$.
(2) $|xy|_p = |x|_p |y|_p$.
(3) $|x+y|_p \leq \max(|x|_p, |y|_p)$, with equality if $|x|_p \neq |y|_p$.

The third property is called the *ultra-triangle inequality*, since it is stronger than the classical triangle inequality $|x+y| \leq |x| + |y|$. One can think of $|\cdot|_p$ as a new way of measuring distance on $\mathbb{Q}$. In this bizarre distance, large powers of p are "close to 0", so something like $1 + p + p^2 + \cdots$ converges to $\frac{1}{1-p}$.

Finally, we go back to the diagram from before



Corollary 2.17. *Let $\mathbb{N}^\omega$ be the set of infinite sequences of natural numbers which are 0 at all but finitely many places. It has a partial order of component-wise comparison, i.e.* $(a_i)_{i=0}^\infty \leq (b_i)_{i=0}^\infty$ *if* $a_i \leq b_i$ *for all* $i \in \mathbb{N}$.

There is an order-preserving bijection from $\mathbb{Z}_{>0}$ ordered by divisibility to $\mathbb{N}^\omega$.

Proof. Let $p_0, p_1, p_2, \cdots$ be the list of all primes. We will prove in the next section that there are infinitely many primes. Define a map

$$v : \mathbb{Z}_{>0} \to \mathbb{N}^\omega, \ v(n) = (v_{p_i}(n))_{i=0}^\infty$$

This is a bijection by the fundamental theorem of arithmetic, and it's clear this preserves the ordering. □

In the above graph, we are taking a piece of $\mathbb{Z}_{>0}$, namely those below $60 = 2^2 \times 3 \times 5$. The proof shows that this poset should look like a product of three chains $(0 \to 1 \to 2) \times (0 \to 1) \times (0 \to 1)$. This exactly corresponds to the three directions that you think you see.

2.5. **Some applications.**

We give some miscellaneous collection of results that follow from the fundamental theorem of arithmetic. The theme is that a lot of questions about the integers can be studied one prime at a time. This principle will play an important role in many aspects of this course.

2.5.1. *Square-free numbers.*

Definition. A positive integer n is *square-free* if there does not exist an integer $d > 1$ such that $d^2|n$.

Example. Among the first 10 positive integers, the square free numbers are 1, 2, 3, 5, 6, 7, 10.

Lemma 2.18. *A positive integer n is square-free if and only if it is a product of distinct primes.*

Proof. Let $n = \prod_p p^{e_p}$ be its prime factorization. If $e_p \geq 2$ for some p, then $p^2|n$, so n is not square-free. Conversely, if n is not square-free, then there exists an integer $d > 1$ such that $d^2|n$. There must exist a prime p such that $p|d$, so $p^2|n$, forcing e_p to be at least two. □

Theorem 2.19. *Every positive integer n is the product of a square and a square-free integer in a unique way.*

Proof. Observe that if $n = r^2 s$ with s square-free, then for all prime p, we have $v_p(n) = 2v_p(r) + v_p(s)$, and $v_p(s) \leq 1$ by the lemma. The existence and uniqueness of $v_p(r)$ and $v_p(s)$ follow from the division algorithm (Proposition 2.3). □

2.5.2. *Irrationality of certain numbers.* It is a classical result that $\sqrt{2}$ is irrational. Technically we haven't defined the real numbers, so the real statement should be that the equation $x^2 - 2 = 0$ has no solutions in $\mathbb{Q}$. This is a special case of a much more general theorem.

Theorem 2.20. *Let $a_1, \cdots, a_d \in \mathbb{Z}$. Suppose $x \in \mathbb{Q}$ satisfies the equation*

$$x^d + a_1 x^{d-1} + \cdots + a_{d-1} x + a_d = 0$$

then $x \in \mathbb{Z}$.

Proof. Suppose $x = \frac{a}{b}$, then by reducing the fraction, we may assume $\gcd(a, b) = 1$. Multiplying the equation by b^d gives the following equation

$$a^d + a_1 a^{d-1} b + \cdots + a_{d-1} a b^{d-1} + a_d b^d = 0$$

If p is a prime and $p|b$, then this equation implies $p|a^d$, so Euclid's lemma gives $p|a$. This contradicts $\gcd(a, b) = 1$. Therefore, b is not divisible by any primes, which implies $b = \pm 1$. It follows that $x \in \mathbb{Z}$. □

A polynomial of whose highest coefficient is 1 is called *monic*. The above theorem gives us an algorithm to find rational solutions of monic polynomials. Indeed, since a polynomial tends to $\pm\infty$ as $x \to \infty$, there is an effective bound on the sizes of the solutions. Within the bound, there are only finitely many integers.

Example. The equation $x^3 - x - 1 = 0$ has no solution in $\mathbb{Q}$. Indeed, by calculus, the function $x^3 - x - 1$ is increasing when $x > 1$ and $x < -1$. Moreover, it is positive at $x = 2$ and negative at $x = -1$, so the only real solutions are in the interval $(-1, 2)$. Neither 0 nor 1 is a solution, so it has no integer solution. By the theorem, it has no rational solution.

2.5.3. *The Riemann zeta function.* We now do some analysis. Let s be a real number, then we can define

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

It is a popular fact that $\zeta(2) = \frac{\pi^2}{6}$.

Euler observed that $\zeta(s)$ has an infinite product expression by the fundamental theorem of arithmetic:

$$\begin{aligned}\zeta(s) &= \sum_{n=1}^{\infty} \left(\prod_p p^{-v_p(n)s} \right) \\ &= \sum_{(k_p) \in \mathbb{N}^\omega} \left(\prod_p p^{-k_p s} \right) \\ &= \prod_p \sum_{k=0}^{\infty} p^{-ks} \\ &= \prod_p \frac{1}{1 - p^{-s}}\end{aligned}$$

Incidentally, this proves there are infinitely many primes: $\zeta(1)$ is not defined since it is the harmonic series, so we must have an infinite product.

This infinite product expression shows that the analytic behaviour $\zeta(s)$ is closely related to the distribution of primes. In the next section, we will sketch a more refined analysis of this behaviour.

Exercise. By taking logarithm of the Euler product and using the Taylor expansion for $\log(1 - x)$, show that $\sum_p \frac{1}{p}$ diverges.

2.6. ***Distribution of primes*.** The fundamental theorem of arithmetic shows that the multiplicative theory of $\mathbb{Z}$ reduces to studying primes. This section will state some results towards their distributions. To begin with, we are obligated to include the classical proof of infinitude of primes.

Theorem 2.21. *There are infinitely many primes.*

Proof. Suppose not. Let P be the product of all primes, and let $N = P+1$. If p is a prime, then $p|P$, so $p \nmid N$. This contradicts the existence of prime factorization. □

Given this, it is natural to study the following function.

Definition. The prime-counting function $\pi(x)$ is the number of primes less than or equal to x.

One simple way to compute and even generate all primes up to x is the *sieve of Eratosthenes*. The algorithm proceeds as follows: first write down all integers up to x excluding 1. At each step, take the smallest number not crossed out and cross out all of its multiples. After you have reached $\sqrt{x}$, the remaining numbers are primes. The reason is if n is composite, then it is divisible by a prime less than or equal to $\sqrt{n}$. For example

$$\begin{gathered}2, 3, 4, 5, 6, 7, 8, 9, 10\\ 2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, 9, \cancel{1}0\\ 2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{1}0\end{gathered}$$

The primes under 10 are 2, 3, 5, 7. This is probably the fastest algorithm which lists out all primes below a certain bound.

On the theoretical side, the study of $\pi(n)$ and its asymptotic behaviour turns out to be a very deep problem. Famously, we have

Theorem 2.22 (Prime number theorem)**.** $\pi(x) \sim \frac{x}{\log x}$*, in the sense that*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1$$

A better estimate is actually given by

$$\pi(x) \sim \mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}$$

This function is the one that shows up naturally in the proof. We will give a glimpse of this proof later, but for now, a natural question is how large can the error $\pi(x) - \mathrm{Li}(x)$ get. Since the main term is $O(x^{1-\epsilon})$, we might hope that the error term involves a power saving. This is not known, but we have the following major open conjecture.

Conjecture (Riemann Hypothesis)**.**

$$\pi(x) - \mathrm{Li}(x) = O\big(x^{\frac{1}{2}+\epsilon}\big)$$

We now explain how the prime number theorem is related to the analytic behaviour of the Riemann zeta function. Recall the Euler infinite product expression

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

Take the logarithmic derivative, we get

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{\log p}{p^s - 1}$$

After rearranging the right hand side, we obtain the identity

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^\infty \psi(t)t^{-s-1}dt$$

where $\psi(x) = \sum_{p\leq x}\left\lfloor\frac{\log x}{\log p}\right\rfloor \log p \approx \pi(x)\log x$, with an error term that can be controlled. We hope that $\psi(x)\approx x$. If this were exact, the right hand side evaluates to $1+\frac{1}{s-1}$. It turns out that near $s=1$,

$$\zeta(s) = \frac{1}{s-1} + \gamma + O(s-1)$$

where $\gamma = 0.5772\dots$ is the *Euler–Mascheroni constant*. This formula is a refinement of the statement that the harmonic series diverges. Even though the constant terms don't agree, the order of growth does, so we can hope this goes somewhere.

To push the proof through, we actually need to allow s to take values in $\mathbb{C}$. This was the starting point of Riemann's work and the reason it is now called the Riemann zeta function. Once we defined $\zeta(s)$ as a function on $\mathbb{C}$, results in Fourier analysis gives us what we want if we can show that $\zeta(s)$ has no zero on the line $\mathrm{Re}(s) = 1$. Hadamard and de la Vallée Poussin independently proved this fact, thereby giving the first proof of the prime number theorem.

It's actually better: Fourier inversion gives an explicit formula for $\pi(x)$ in terms of the zeroes of $\zeta(s)$. Any information on the zeroes of the Riemann zeta function gives information on the error term in $\pi(x) - \mathrm{Li}(x)$. A power saving error term would require us to know $\zeta(s)$ has no zero for $\mathrm{Re}(s) > 1-\delta$, for some $\delta > 0$. This is still not known, but Riemann conjectured something much stronger.

Conjecture (Riemann Hypothesis)**.** *The only zeroes of* $\zeta(s)$ *with* $0 \leq \mathrm{Re}(s) \leq 1$ *occurs when* $\mathrm{Re}(s) = \frac{1}{2}$.

3. Congruence

This chapter introduces the important notion of congruence. In a very precise sense that we will somewhat explain, congruence enables us to study the "local" behaviour of $\mathbb{Z}$. Often, problems are easier to understand locally, and we will introduce tools to aid that.

3.1. Basic properties.

Definition. Let m be an integer. We say that integers a and b are *congruent modulo* m if $m|a-b$. This is written as $a \equiv b \pmod{m}$.

Example.

(1) All even numbers are congruent to each other modulo 2.
(2) All integers are congruent modulo 1.
(3) $a \equiv b \pmod{0}$ implies $a = b$.
(4) $a \equiv b \pmod{m}$ is equivalent to $a \equiv b \pmod{-m}$, so usually, people only talk about congruences modulo a positive integer.

The first properties of congruence is that it is an *equivalence relation.*

Proposition 3.1. *Let $m \in \mathbb{Z}$. For all $a, b, c \in \mathbb{Z}$, the following holds*

– *Reflexivity: $a \equiv a \pmod{m}$.*
– *Symmetry: If $a \equiv b \pmod{m}$, then $b \equiv a \pmod{m}$.*
– *Transitivity: If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.*

Proof. All of them are straightforward and on the homework. □

One thing you can do with an equivalence relation is to take its equivalence classes. Recall that this means treating two elements as equal if they are equivalent.

Definition. The set of equivalence classes modulo m is denoted by $\mathbb{Z}/m\mathbb{Z}$.

Remark. Some books write $\mathbb{Z}_m$ for this set. They are wrong.

In this chapter, we will use $[n]_m$ to denote the equivalence class of n. This defines a function

$$\texttt{red} : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}, \quad n \mapsto [n]_m$$

which we call *reduction* modulo m. Other common notations are $n \pmod{m}$, or just $\bar{n}$ when m is clear from context. It is also very common to abuse notation and conflate an integer $n \in \mathbb{Z}$ with its congruence class $[n]_m \in \mathbb{Z}/m\mathbb{Z}$. We will phase in this notation gradually. One reason we can get away with it is that reduction map respects arithmetic operations.

Let's begin by trying to define the two operations in the obvious way.

$$[a]_m + [b]_m := [a+b]_m, \quad [a]_m[b]_m := [ab]_m$$

This may look like an obvious definition, but it is hiding the fact that multiple integers a may be responsible for the same congruence class. If $[a] = [a']$, then we need to know $[a+b] = [a'+b]$. Otherwise, the definition is not *well-defined.* In our case, this check is done by the following proposition.

Proposition 3.2. *Let $m \in \mathbb{Z}$. For all $a, b, c, d \in \mathbb{Z}$ such that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, we have*

$$a + c \equiv b + d \pmod{m}, \quad ac \equiv bd \pmod{m}$$

Proof. Observe that

$$(a+c)-(b+d)=(a-b)+(c-d), \quad ac-bd=(a-b)c+b(c-d)$$

Since $m|a-b$ and $m|c-d$, the claim follows from basic properties of division. □

Example. The definition

$$[a]_m^{[b]_m} \overset{?}{:=} [a^b]_m$$

is not well-defined. Suppose we are working in $\mathbb{Z}/3\mathbb{Z}$, then by this definition, $[2]_3^{[1]_3} = [2]_3$ and $[2]_3^{[4]_3} = [16]_3 = [1]_3$. This is a problem since $[1]_3 = [4]_3$ and we have just assigned two different values to $[2]_3^{[1]_3}$.

Remark. We briefly explain the notation $\mathbb{Z}/m\mathbb{Z}$. Recall that $\mathbb{Z}$ is a ring, and $m\mathbb{Z}$ is the ideal $\{md \mid d \in \mathbb{Z}\}$. In general, given a ring R and an ideal I, we can form an equivalence relation

$$x \sim y \iff x-y \in I$$

which in our case is exactly the condition of congruent modulo m. The proof of the previous proposition carries over in this abstract setting to show that R/I is actually a ring. This is the *quotient ring* construction.

The set $\mathbb{Z}/m\mathbb{Z}$ is very simple to describe.

Proposition 3.3. *If $m \geq 1$, then the set $\mathbb{Z}/m\mathbb{Z}$ is finite of size m.*

Proof. Observe that $n \equiv r \pmod{m}$ if and only if there exists $q \in \mathbb{Z}$ such that $n = qm + r$. Therefore, by the division algorithm (Proposition 2.3), there exists a unique $r \in \{0, 1, \cdots, m-1\}$ such that $n \equiv r \pmod{m}$. This set is then a set of representatives of the equivalence classes, proving the claim. □

Suppose we have a polynomial P with integer coefficients, and suppose the equation $P(x) = 0$ holds for some $x \in \mathbb{Z}$. Let $m \in \mathbb{Z}$ be arbitrary, then we also have $P([x]_m) = [0]_m$. Take the contrapositive, we see that if $P([x]_m) \neq [0]_m$ for all $[x]_m \in \mathbb{Z}/m\mathbb{Z}$, then $P(x) = 0$ has no integer solution. Moreover, it's a *finite* computation to check if $P([x]_m)$ can be $[0]_m$ since $\mathbb{Z}/m\mathbb{Z}$ is finite. Therefore, we have found a computable necessary condition for $P(x) = 0$ to have an integer solution.

Example. There does not exist integers x, y such that $y^2 = x^3 + 2x - 1$.

Proof. Consider the equation in $\mathbb{Z}/3\mathbb{Z}$. It is easy to verify that $[y]^2 = [0]_3$ or $[1]_3$ for all $[y] \in \mathbb{Z}/3\mathbb{Z}$. Checking the three possibilities for $[x]_3$ shows that the right hand side is always equal to $[2]_3$. Therefore, the equation has no solution in $\mathbb{Z}/3\mathbb{Z}$, and so no solution in $\mathbb{Z}$. □

This technique is called "reduction mod m". Note that m is arbitrary, and as soon as you find an m such that $P(x) = 0$ has no solution in $\mathbb{Z}/m\mathbb{Z}$, you are done. This is called a *local obstruction* to solutions. In this chapter, we will study the ring $\mathbb{Z}/m\mathbb{Z}$ closely, which will help in identifying local obstructions.

Example. The equation $x^2 + y^2 + z^2 = 7$ has no rational solution. To see this, after clearing out denominators, we need to show that $x^2 + y^2 + z^2 = 7d^2$ has no non-trivial integer solution (the trivial solution is where all variables are 0).

Reduction mod m can't directly work since it has a trivial solution. However, reduction mod 8 shows that all x, y, z, d must be even. Therefore, if we have a non-trivial solution, then dividing all of them by 2 gives a solution with a smaller

$|x|+|y|+|z|$. By the well-ordering property, this process can't continue forever, so we have no non-trivial solution. This is a simple instance of the technique of infinite descent.

Unfortunately, as the next example shows, having solutions in $\mathbb{Z}/m\mathbb{Z}$ for all positive m does not imply having solutions in $\mathbb{Z}$. This is an instance of the failure of the *local-global principle.* The extent of this failure and questions surrounding it is an active area of research.

Example. The equation $x^2+y^2+7z^2=3$ clearly has no integer solutions, but we will later show that it has a solution modulo m for all positive integer m.

3.2. Chinese remainder theorem.

Recall that any positive integer is a product of prime powers. Continuing the theme that primes are the multiplicative building blocks, it is natural to ask how this interacts with congruences.

Theorem 3.4 (Chinese remainder theorem)**.** *Let $m_1,\cdots,m_n$ be pairwise coprime positive integers. Let $m=\prod_{i=1}^n m_i$, then the function*

$$\texttt{red}:\mathbb{Z}/m\mathbb{Z}\to\mathbb{Z}/m_1\mathbb{Z}\times\cdots\times\mathbb{Z}/m_n\mathbb{Z},\ [a]_m\mapsto([a]_{m_1},\cdots,[a]_{m_n})$$

is a bijection.

Proof. Both sides have the same number of elements, so we only need to show `red` is injective. Suppose $\texttt{red}([a]_m)=\texttt{red}([b]_m)$. Let $c=b-a$, then

$$\texttt{red}([c]_m)=([b-a]_{m_i})_{i=1}^n=([0]_{m_i})_{i=1}^n$$

Equivalently, $m_i|c$ for all i. By the definition of the least common multiple, $\operatorname{lcm}(m_1,\cdots,m_n)|c$. On the other hand, Homework 4.5 shows that $\operatorname{lcm}(m_1,\cdots,m_n)=m$. Therefore, $[c]_m=[0]_m$, which implies $[a]_m=[b]_m$. This proves injectivity. □

Remark. The function `red` actually preserves addition and multiplication, so it is a *ring isomorphism.* Vaguely speaking, two objects are isomorphic if they are "essentially the same thing". In this context, it means any arithmetic statement about integers modulo m is equivalent to a system of statements about integers modulo m_i, for $i=1,\cdots,n$.

Corollary 3.5. *Let $m_1,\cdots,m_n$ be pairwise coprime positive integers, then for any integers $a_1,\cdots,a_n$, the system of equations*

$$\begin{cases} x\equiv a_1 \pmod{m_1}\\ \vdots\\ x\equiv a_n \pmod{m_n}\end{cases}$$

has a unique solution modulo $m=m_1\cdots m_n$.

Alternative phrasing: there exists an $a\in\mathbb{Z}$ such that the system is equivalent to a single equation $x\equiv a \pmod{m}$.

Proof. The solution x is the pre-image of $(a_1,\cdots,a_n)$ under `red`.[3] □

[3] Here, we are using the abuse of notation that treats for example x as simultaneously an integer and a congruence class modulo m. We can call it "the" solution since x is viewed as an element in $\mathbb{Z}/m\mathbb{Z}$, which is exactly what "unique modulo m" means.

The proof of the theorem was very short, but it does not tell us how to find x since it is based on a counting argument. For practical applications, we need a constructive proof. We will give it in the case $n = 2$. In general, it can be handled by induction: combine the first two equation using the $n = 2$ case into a single congruence equation modulo m_1m_2, then combining this with the third equation to obtain a congruence equation modulo $m_1m_2m_3$, and so on.

By letting $(a_1, a_2) = (1, 0), (0, 1)$ in the corollary, we get the following two systems of congruence equations

$$\begin{cases} e_1 \equiv 1 \pmod{m_1} \\ e_1 \equiv 0 \pmod{m_2} \end{cases} \qquad \begin{cases} e_2 \equiv 0 \pmod{m_2} \\ e_2 \equiv 1 \pmod{m_2} \end{cases}$$

Suppose we have found e_1, e_2 satisfying these properties, then we can actually solve the equation for a general pair (a_1, a_2), since by the basic properties of congruence, $a = a_1e_1 + a_2e_2$ satisfies

$$a \equiv a_1 \pmod{m_1}, \quad a \equiv a_2 \pmod{m_2}$$

In some sense, (e_1, e_2) acts as some kind of "basis" for $\mathbb{Z}/m_1m_2\mathbb{Z}$, and it corresponds to the "standard basis" $((1, 0), (0, 1))$ of $\mathbb{Z}/m_1\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z}$.

We now solve the two basic systems of congruences in one step. Bézout's theorem gives us integers u_1, u_2 such that $m_1u_1 + m_2u_2 = 1$. Reducing this equation modulo m_1, we see that $m_2u_2 \equiv 1 \pmod{m_1}$. Moreover, $m_2u_2 \equiv 0 \pmod{m_2}$, so we can take $e_1 = m_2u_2$. Similarly, we can take $e_2 = m_1u_1$. Moreover, both of them can be computed efficiently from Euclid's algorithm.

Example. Solve the system of equations

$$\begin{cases} x \equiv 1 \pmod{2} \\ x \equiv 2 \pmod{3} \\ x \equiv 3 \pmod{5} \end{cases}$$

Solution. By inspection, the first two equations have the solution $x \equiv 5 \pmod{6}$ (6 is small enough that it is faster to just guess).

Now, to solve the system $x \equiv 5 \pmod{6}$ and $x \equiv 3 \pmod{5}$, we run Euclid's algorithm for the pair $(6, 5)$ and see that $6 \times 1 + 5 \times (-1) = 1$. Therefore, $e_1 = -5$ and $e_2 = 6$ in the notation of the discussion above, so a solution is $x = 5e_1 + 3e_2 = -25 + 18 = -7$. The general solution is all x such that $x \equiv -7 \pmod{30}$.

There is a canonical way of writing a positive integer m as a product of pariwise coprime integers, namely we can take $m = p_1^{f_1} \cdots p_k^{f_k}$, its prime factorization. The Chinese remainder theorem tells us that studying a congruence modulo m is equivalent to studying congruences modulo $p_i^{f_i}$ for each i.

Example. Observe that if p is a prime and $n \geq 1$, then $x^2 \equiv x \pmod{p^n}$ only has the two obvious solutions $x \equiv 0, 1 \pmod{p^n}$.[4] Indeed, if $p^n | x^2 - x$, then $p|x$ or $p|x - 1$, but not both. Therefore, $p^n|x$ or $p^n|x - 1$ by Euclid's lemma.

How many solutions does $x^2 \equiv x \pmod{m}$ has for a general odd integer m?

[4] Technically, we should write $x = [0]_{p^n}, [1]_{p^n}$, but this is another standard abuse of notation where we treat a statement $x \equiv 0 \pmod{p^n}$ as an equivalence class $[0]_{p^n}$.

Answer. Suppose $m = p_1^{f_1} \cdots p_k^{f_k}$ is its prime factorization, then the equation has exactly 2^k solutions. The reason is that $x^2 \equiv x \pmod{m}$ is equivalent to a system of k congruences modulo prime powers by the Chinese remainder theorem. Each congruence has 2 solutions, and we can choose either of them independently for each equation. The Chinese remainder theorem again allows us to combine each of these choices into a single solution modulo m.

As a simple example, take $m = 100 = 4 \times 25$. In addition to the obvious solutions $x \equiv 0, 1 \pmod{100}$ corresponding to the choices $(0, 0)$ and $(1, 1)$, we also have solutions to the following two systems

$$\begin{cases} x \equiv 0 \pmod{4} \\ x \equiv 1 \pmod{25} \end{cases} \qquad \begin{cases} x \equiv 1 \pmod{4} \\ x \equiv 0 \pmod{25} \end{cases}$$

The algorithm described above gives us additional solutions $x \equiv 25, 76 \pmod{100}$. In very elementary terms, these are the two-digit integers whose square has the same final two digits.

As a consequence of this kind of observation, in our discussion of local obstructions, we may restrict m to being only prime powers without losing anything. Now the situation is actually local, in a sense we still have not yet explained.

3.3. **Interlude: polynomials.** To study polynomial equations, we need to define polynomials. Some of their properties will also be needed later in the course. By definition, polynomials are finite sequences of elements. *They are not functions.*

Definition. Let S be a set. Let X be a formal symbol. The set $S[X]$ consists of formal expressions of the form

$$c_0 + c_1 X + \cdots + c_n X^n$$

where $n \in \mathbb{N}$ and $c_0, c_1, \cdots, c_n \in S$. Elements of this set are called *polynomials over S*. The least k such that $c_k \neq 0$ is called the *degree* of the polynomial. Two polynomials are equal if and only if they have the same degree and list of coefficients.

If S is a ring, then the sum of two polynomials is defined by adding their coefficients term by term. If $\sum_{k=0}^{n} a_k X^k$ and $\sum_{k=0}^{m} b_m X^k$ are polynomials, then their product is the polynomial $\sum_{k=0}^{n+m} c_k X^k$, where

$$c_k = \sum_{\substack{i+j=k \\ 0 \leq i \leq n \\ 0 \leq j \leq m}} a_i b_j$$

Informally, they are multiplied in the usual way. With these operations, the set $S[X]$ becomes a ring, which we call the polynomial ring of S.

Let $f(X)$ be a polynomial over S, and let $x \in S$, then we define the *evaluation* of $f(X)$ at x to be

$$f(x) := c_0 + c_1 x + \cdots + c_n x^n \in S$$

This defines a function from S to S.

Example. Let $S = \mathbb{Z}/3\mathbb{Z}$. Let

$$f(X) = [0]_3 + [-1]_3 X + [0]_3 X^2 + [1]_3 X^3$$

then f is a polynomial over $\mathbb{Z}/3\mathbb{Z}$ of degree 3. It is standard to abbreviate this to just $f(X) = X^3 - X$.

This defines a function $\mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$. It is to compute that

$$f([0]_3) = [0]_3, \quad f([1]_3) = [0]_3, \quad f([2]_3) = [0]_3$$

so the function is identically 0.

On the other hand, $g(X) = [0]_3$ is also a polynomial over $\mathbb{Z}/3\mathbb{Z}$. It does not have a well-defined degree since all coefficients are zero. The function $g(X)$ defines is also identically 0. However, $f(X) \neq g(X)$ since they have different degrees.

A polynomial is a different object from the function it defines.

Definition. Let R be a ring. Let $f(X) \in R[X]$ be a polynomial over R. A *solution* to the equation $f(X) = 0$ is an element $x \in R$ such that $f(x) = 0$.

Example. The polynomial equation $X^3 - X = 0$ has three solutions in $\mathbb{Z}/3\mathbb{Z}$. If we interpret this as an equation over $\mathbb{Z}$, then it still has three solutions: $x = 0, \pm 1$.

We can also interpret the equation in $\mathbb{Z}/2\mathbb{Z}$, but then it only has two solutions. Over $\mathbb{Z}/10\mathbb{Z}$, it has 6 solutions: $0, 1, 4, 5, 6, 9$.

Here is the formal version of the reduction mod m strategy.

Definition. Let $f(X) = c_0 + \cdots + c_n X^n$ be a polynomial over $\mathbb{Z}$. Let $m \geq 1$. The polynomial over $\mathbb{Z}/m\mathbb{Z}$ given by

$$\bar{f}(X) = [c_0]_m + [c_1]_m X + \cdots + [c_n]_m X^n$$

is called the reduction of $f(X)$ modulo m.

Lemma 3.6. *In the notation of the definition, for all $a \in \mathbb{Z}$, we have*

$$[f(a)]_m = \bar{f}([a]_m)$$

In particular, if $\bar{f}(X) = 0$ has no solution, then $f(X) = 0$ has no solution.

Proof. This is an immediate consequence of the definition of addition and multiplication in $\mathbb{Z}/m\mathbb{Z}$. □

*Remark**. We have a *commutative diagram*

$$\begin{tikzcd}
\mathbb{Z}[X] \arrow[r, "\mathtt{ev}_a"] \arrow[d] & \mathbb{Z} \arrow[d] \\
(\mathbb{Z}/m\mathbb{Z})[X] \arrow[r, "\mathtt{ev}_{[a]_m}"] & \mathbb{Z}/m\mathbb{Z}
\end{tikzcd}$$

There is a general version with $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ replaced by an arbitrary "structure-preserving map" $R \to S$ between two rings. These are called *ring homomorphisms.*

Polynomials also have a division algorithm. In fact, there is a deep analogy between the ring $k[X]$ for a finite field k and the ring $\mathbb{Z}$. While polynomial division is not hard to prove in general, we will only need a very small piece of it.

Lemma 3.7. *Let R be a ring, let $f(X) \in R[X]$. Suppose $\alpha \in R$ satisfies $f(\alpha) = 0$, then there exists $g(X) \in R[X]$ such that $f(X) = (X - \alpha)g(X)$.*

Proof. Suppose $f(X) = \sum_{k=0}^{n} a_k X^k$. For $k = 1, \cdots, n$, define

$$b_{n-k} = \sum_{i=0}^{k-1} \alpha^i a_{n-k+i+1}$$

Let $g(X) = \sum_{k=0}^{n-1} b_k X^k$, then a computation shows that $f(X) = (X - \alpha)g(X)$. □

Theorem 3.8. *If k is a field, then a polynomial of degree d in $k[X]$ can have at most d roots.*

Proof. We prove this by induction on d. The base case when $d = 0$ is clear. Suppose the theorem is known for a given d, let $f(X)$ be a polynomial of degree $d + 1$. If $f(X)$ has no root, then we are done. Otherwise, let α be a root. By the lemma, we can find $g(X)$ of degree d such that $f(X) = (X - \alpha)g(X)$. Suppose β is any root of $f(X)$, then $(\beta - \alpha)g(\beta) = 0$. Since k is a field, this implies $\beta = \alpha$ or $g(\beta) = 0$. By the induction hypothesis, $g(\beta) = 0$ occurs for at most d possible choices of β. Therefore, $f(X)$ has at most $d + 1$ roots. □

Finally, we do some calculus.

Definition. Let R be a ring, and let $f(X) = \sum_{k=0}^{n} c_k X^k$ be a polynomial over R. Its *derivative* is the polynomial

$$f'(X) = \sum_{k=1}^{n} kc_k X^{k-1}$$

Lemma 3.9 (1st order Taylor expansion)**.** *Let R be a ring and let $f(X)$ be a polynomial over R, then for all $x, \varepsilon \in R$, there exists $z \in R$ such that*

$$f(x + \varepsilon) = f(x) + f'(x)\varepsilon + z\varepsilon^2$$

Proof. Suppose $f(X) = \sum_{k=0}^{n} c_k X^k$. Applying binomial expansion to each term of $f(x + \varepsilon)$, we get

$$f(x + \varepsilon) = \sum_{k=1}^{n} c_k(x^k + kx^{k-1}\varepsilon + z_k\varepsilon^2) + c_0$$

where $z_k \in R$ accounts for the remaining terms in the expansion. Therefore,

$$f(x + \varepsilon) - (f(x) + f'(x)\varepsilon) = \varepsilon^2 \sum_{k=1}^{n} c_k z_k$$

Taking $z = \sum_{k=0}^{n} c_k z_k \in R$ gives the required formula. □

3.4. **Congruence equations.** To effectively use the technique of reduction mod m, we need to understand solutions to equations in $\mathbb{Z}/m\mathbb{Z}$. By the Chinese remainder theorem, we only need to consider the case when m is a prime power. The upshot of this section is that in a lot of the cases, we only need to consider the case when m is a prime.

3.4.1. *Linear equations.* This is the simplest case. Let a, b, m be integers, $m \geq 1$. We want to find all solutions to the equation

$$ax \equiv b \pmod{m}$$

Example.

(1) The equation $2x \equiv 3 \pmod{5}$ has a unique solution $x \equiv 4 \pmod{5}$. The integer $x = -1$ is also a solution, but 4 and -1 are the same in $\mathbb{Z}/5\mathbb{Z}$.
(2) The equation $2x \equiv 1 \pmod{4}$ has no solution.
(3) The equation $3x \equiv 6 \pmod{9}$ has exactly 3 solutions: $x \equiv 2, 5, 8 \pmod{9}$.

By the definition of congruence, we need to find integers x such that there exists an integer y such that $ax - b = ym$. Equivalently, we need to solve the 2-variable linear equation

$$ax - my = b$$

for $x, y \in \mathbb{Z}$. This is exactly what you studied in the homework. We summarize the results here.

Theorem 3.10. *Let $d = \gcd(a, m)$. If $d \nmid b$, then the equation $ax \equiv b \pmod{m}$ has no solution. Otherwise, it has exactly d solutions in $\mathbb{Z}/m\mathbb{Z}$. Moreover, if x_0 is a solution, then all solutions have the form $x_0 + k\frac{m}{d}$ for $k \in \mathbb{Z}$.*

The following two special cases are very important

Corollary 3.11. *An integer a is invertible in $\mathbb{Z}/m\mathbb{Z}$ if and only if $\gcd(a, m) = 1$.*

Corollary 3.12. *Let p be a prime, then the ring $\mathbb{Z}/p\mathbb{Z}$ is a field.*

Proof. Since p is a prime, $\gcd(p, a) = 1$ or p, and it equals to p exactly when $p|a$, which is equivalent to $a \equiv 0 \pmod{p}$. Therefore, by the previous corollary, every non-zero element is invertible. □

Notation. We will sometimes write $\mathbb{F}_p$ for $\mathbb{Z}/p\mathbb{Z}$.

3.4.2. **p-adic numbers**. Before moving on to higher degree equations, we take an interlude to introduce an important structure. Its purpose is to package together information modulo all powers of a single prime.

> *The real numbers is like the sun, the p-adic numbers are like the stars. Normal people sleep at night, so they only see the sun. But if you stay up all night, you see all the infinite brilliance of the stars.*
>
> Kazuya Kato

Let p be a prime, then the rings $\mathbb{Z}/p^n\mathbb{Z}$ for all n fit together into an infinite sequence of rings

$$\cdots \to \mathbb{Z}/p^3 \to \mathbb{Z}/p^2\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$$

where each function is the reduction map. More visually, if you expand a positive integer N in base p, then reduction modulo p^n truncates N and only keep the final n digits. As we go further left, we recover more information about N:

$$\begin{tikzcd}
\cdots \arrow[r] & \mathbb{Z}/2^4\mathbb{Z} \arrow[r] & \mathbb{Z}/2^3\mathbb{Z} \arrow[r] & \mathbb{Z}/2^2\mathbb{Z} \arrow[r] & \mathbb{Z}/2\mathbb{Z} \\
\cdots \arrow[r] & 1101_2 \arrow[u, phantom, sloped, "\in"] \arrow[r, mapsto] & 101_2 \arrow[u, phantom, sloped, "\in"] \arrow[r, mapsto] & 1_2 \arrow[u, phantom, sloped, "\in"] \arrow[r, mapsto] & 1_2 \arrow[u, phantom, sloped, "\in"]
\end{tikzcd}$$

Since N is a natural number, eventually (when $p^n > N$), no new information is obtained by going to higher powers of p. This is like a terminating decimal.

On the other hand, we can consider an infinite sequence like

$$\cdots \mapsto 1111_2 \mapsto 111_2 \mapsto 11_2 \mapsto 1_2$$

This is like a repeating decimal. What kind of object is this? We give two ways of getting to the answer.

(1) At each step, we can use a different representative. The sequence can be re-written as

$$\cdots \mapsto [-1]_{2^4} \mapsto [-1]_{2^3} \mapsto [-1]_{2^2} \mapsto [-1]_2$$

so it's natural to claim this sequence should be called -1.

(2) Write out the binary expansion:

$$\underbrace{1\cdots 1_2}_{n} = \sum_{i=0}^{n-1} 2^i = 2^n - 1$$

If we want to treat it as an infinite decimal, then we should treat large powers of 2 as being small, so $\lim_{n\to\infty}(2^n - 1) = -1$.

Definition. A p-adic integer is a compatible sequence of elements in the chain

$$\cdots \to \mathbb{Z}/p^3\mathbb{Z} \to \mathbb{Z}/p^2\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$$

The set of all p-adic integers is denoted by $\mathbb{Z}_p$. It carries a ring structure by component-wise operation.

By using base p expansions, they can be thought of as infinite integers in base p. For example, we can write -1 as

$$-1 = \cdots 1111_2 = \sum_{i=0}^{\infty} 2^i$$

To add two infinite decimals, you start with the most significant bits and move to the right until you reached the required precision. This is also how one deals with the p-adic numbers, except the situation is flipped: the right most digit is the most significant, and we move to the left.

Example. Let $p = 5$, and let $x = 24031_5$. To compute x^2, do the usual multiplication process in base 5.

$$x^2 = 1244300011_5 = 1244300000_5 + 6$$

The first number is 0 in the five most significant bits, so you should think of it as being small. Therefore, $x^2 \approx 6$ to five bits of accuracy, so $x \approx \sqrt{6}$. We can write $\sqrt{6} = \cdots 24031_5$. There is no pattern. It's like saying $\sqrt{6} = 2.4494\cdots$ in $\mathbb{R}$.

This is not entirely correct. There are two square roots of 6 in $\mathbb{Z}_5$. For real numbers, the convention is that $\sqrt{6}$ denotes the positive one, but over the p-adic numbers, there is no notion of ordering, so the notation $\sqrt{6}$ is ambiguous. It is always preferable to say "let α be a square root of 6" than to write $\alpha = \sqrt{6}$.

Exercise. How do you do long division in $\mathbb{Z}_p$?

The real numbers have a measure of distance: the distance between x and y is $|x - y|$. We have earlier introduced a different function

$$|x|_p = p^{-v_p(x)}$$

which satisfies $|p^n|_p \to 0$ as $n \to \infty$. This is the basis for saying that the series $\sum_{i=0}^{\infty} 2^i$ converges in the 2-adic numbers: we are just measuring sizes in a strange way. Using this metric, we can do calculus over the p-adic numbers.

> *One day I woke up, take 3 steps out of my house and found myself close to it. I go home and take 9 steps, and now I'm even closer. I go to the doctor and say "Doctor, I live in a 3-adic world!", but the doctor just looks at me and says "No, you're drunk".*
>
> Kazuya Kato, allegedly

3.4.3. *Hensel's lemma.* It's hard to solve polynomial equations exactly, but as you learned in calculus, you can easily approximate the solutions using Newton iteration. We can do that in the p-adic world.

Theorem 3.13 (Hensel's lemma, nice version)**.** *Let $f(X)$ be a polynomial over $\mathbb{Z}_p$. Suppose $a \in \mathbb{Z}_p$ satisfies*

$$|f(a)|_p < |f'(a)|_p^2$$

then there exists a unique $\alpha \in \mathbb{Z}_p$ such that $f(\alpha) = 0$ and $|\alpha - a|_p < |f'(a)|_p$.

Proof. We perform Newton iteration starting from a. Define a sequence $(x_n)_{n\in\mathbb{N}}$ recursively by

$$x_0 = a, \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Let $\delta = |f(a)|_p / |f'(a)|_p^2 < 1$, then the idea is to show the estimates

$$\begin{aligned} |x_{n+1} - x_n|_p &\le |f'(a)|_p \, \delta^{2^n} \\ |f(x_n)|_p &\le |f'(a)|_p^2 \, \delta^{2^n} \end{aligned}$$

by induction. Let $\alpha = \lim_{n\to\infty} x_n$, then $f(\alpha) = 0$. The estimates are proven using the usual Taylor expansion argument by applying Lemma 3.9.

To prove uniqueness, suppose $|\beta - a|_p < |f'(a)|_p$, then the ultratriangle inequality implies $|\beta - \alpha|_p < |f'(a)|_p$. Moreover, Taylor expansion gives

$$f(\beta) = (\beta - \alpha)f'(\alpha) + O((\beta - \alpha)^2)$$

The norm of the second term is strictly smaller than the norm of the first term, so by the equality case of ultratriangle inequality, $|f(\beta)|_p = |(\beta - \alpha)f'(\alpha)|_p \neq 0$ unless $\alpha = \beta$. □

From our elementary point of view, this is an algorithm to solve a polynomial equation $f(X) \equiv 0 \pmod{p^n}$ for an arbitrary n, provided that we have a sufficiently good initial guess.

Example. Suppose we want to find x such that $x^2 \equiv 6 \pmod{5^4}$. This is equivalent to computing $\sqrt{6}$ in $\mathbb{Z}_5$ to four digits of precision.

Modulo 5 (i.e. to one digit precision), we have $6 = 1$, so a starting guess can be $x_0 = 1$. Let $f(X) = x^2 - 6$, then

$$|f(1)|_5 = 5^{-1}, \quad |f'(1)|_5 = |2|_5 = 1$$

so the condition of Hensel's lemma is satisfied, and we can perform Newton iteration. In the first step,

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1 - \frac{-5}{2} = 1 + \frac{5}{2} = \cdots 222231_5 = 22231_5 + O(5^5)$$

The last step is like rounding to five decimal places. The second step is

$$x_2 = x_1 - \frac{x_1^2 - 6}{2x_1} = 22231_5 - 23200_5 + O(5^5) = 44031_5 + O(5^5)$$

By the estimates given in the proof, this should be accurate to four digits, which agrees with our earlier example that $\sqrt{6} = \cdots 24031_5$. If we had chosen $x_0 = -1$, we would have gotten the other square root of 6.

All of these computations are easy to do on a computer, but they are annoying for a human. In a simple but typical case, we can elaborate on the proof to produce an algorithm that's doable on paper for small numbers.

Theorem 3.14 (Hensel's lemma, elementary version)**.** *Let $f(X)$ be a polynomial over $\mathbb{Z}$. Suppose a is an integer such that*

$$f(a) \equiv 0 \pmod{p}, \quad f'(a) \not\equiv 0 \pmod{p}$$

then for all $n \geq 1$, there exists a unique integer x_n modulo p^n such that

$$f(x_n) \equiv 0 \pmod{p^n}, \quad x_n \equiv a \pmod{p}$$

Proof. We prove the result by induction on n using a technique known as "lifting". The case $n = 1$ is immediate. Suppose this is known for a given n. By the uniqueness part, any solution x_{n+1} must satisfy $x_{n+1} \equiv x_n \pmod{p^n}$. Modulo p^{n+1}, there are p possibilities for x_{n+1}

$$x_{n+1} = x_n + p^n h, \; h \in \{0, 1, \cdots, p-1\}$$

We need to prove that there is a unique h such that $f(x_{n+1}) \equiv 0 \pmod{p^{n+1}}$.

By Taylor expansion (Lemma 3.9),

$$f(x_n + p^n h) = f(x_n) + p^n h f'(x_n) + \xi p^{2n}$$

for some integer ξ. By induction hypothesis, $f(x_n) \equiv 0 \pmod{p^n}$, so there exists $d \in \mathbb{Z}$ such that $f(x_n) = dp^n$. Therefore,

$$f(x_n + p^n h) = p^n(d + hf'(x_n)) + p^{2n}\xi$$

We need this to be a multiple of p^{n+1}, so we need $d + hf'(x_n) \equiv 0 \pmod{p}$. Since $x_n \equiv a \pmod{p}$, we also have $f'(x_n) \equiv f'(a) \pmod{p}$, which is non-zero by hypothesis. Therefore, this congruence modulo p has a unique solution for h. □

Example. In our example from before with $p = 5$ and $f(X) = X^2 - 6$, we can take $a = 1$, then the two hypotheses are satisfied.

In the first step,

$$f(a + hp) = (a + hp)^2 - 6 = a^2 + 2ahp + h^2p^2 - 6 = 5(2h - 1) + h^2p^2$$

so we need to solve $2h - 1 \equiv 0 \pmod{5}$, which has the solution $h \equiv 3 \pmod{5}$. It follows that $1 + 3 \times 5 = 16$ is a solution to $x^2 \equiv 6 \pmod{5^2}$. Equivalently, $31_5 \approx \sqrt{6}$ to two digits of accuracy.

To continue this algorithm further, each step requires solving the equation

$$f'(a)h + d \equiv 0 \pmod{p}$$

for some d. If u is an inverse of $f'(a)$, meaning $f'(a)u \equiv 1 \pmod{p}$, then the solution is $h \equiv -ud \pmod{p}$. The inverse u can be found by applying Euclid's algorithm to the pair $(f'(a), p)$. After pre-computing u, each additional digit only requires one evaluation of the function and one additional multiplication modulo p to compute.

3.4.4. *Local obstructions.* The elementary version of Hensel's lemma tells us that an equation $f(X) \equiv 0 \pmod{p^n}$ has a solution for all n when $f(X) \equiv 0 \pmod{p}$ has a solution, subject to certain non-degeneracy conditions. This extra condition can be eliminated using the full version of the lemma. The upshot is that congruence equations modulo a prime power p^n can be reduced to congruence equations modulo p or a small power.

In this section, we prove that the equation $x^2 + y^2 + 7z^2 = 3$ has no local obstructions. On the surface, this is still an infinite computation, since we still need to check at all primes. However, we can use a trick: observe that

$$(x, y, z) = \left(\frac{1}{2}, 1, \frac{1}{2}\right)$$

satisfies $x^2 + y^2 + 7z^2 = 3$. Therefore, the equation has a solution in *all* rings where $\frac{1}{2}$ is defined, i.e. where 2 is invertible. This immediately allows us to conclude that there are no local obstructions modulo any odd integer.

The remaining possibilities are powers of 2. This requires the strong form of Hensel's lemma. Let $y = 0$ and $z = 3$, then $x^2 = -60$. We will show that -15 has a square root in $\mathbb{Z}_2$. Let α be one such root, then $(2\alpha, 0, 3)$ is a solution to the equation in $\mathbb{Z}_2$, so its truncation modulo 2^n is a solution in $\mathbb{Z}/2^n\mathbb{Z}$.

Let $f(X) = X^2 + 15$, then modulo 8, $f(X) \equiv X^2 - 1$. An initial guess could be $x_0 = 1$. We compute

$$|f(1)|_2 = |16|_2 = 2^{-4}, \quad |f'(1)|_2 = |2|_2 = 2^{-1}$$

so the hypothesis of Hensel's lemma is satisfied. Note that the weak version fails since $2|f'(1)$.

Remark. We make some remarks about the general strategy. The following statements should contain some technical hypotheses to rule out degenerate cases, but we will not give the details.

(1) There are general results that says a system of polynomial equations in sufficiently many variables always has solutions modulo p for any sufficiently large prime p, where the bound is effectively computable.
(2) For any polynomial $f(X)$, there are only finitely many primes p such that the system
$$\begin{cases} f(X) \equiv 0 \pmod{p} \\ f'(X) \equiv 0 \pmod{p} \end{cases}$$
has a common solution. In fact, they are all divisors of the *discriminant* of $f(X)$, which is a computable integer attached to $f(X)$.
(3) Combining the two results, we see that for all but finitely many primes, the elementary form of Hensel's lemma guarantees there is no local obstruction modulo powers of that prime.
(4) Let p be one of the remaining finitely many primes, then we can search for solutions in $\mathbb{Z}/p^n\mathbb{Z}$ in the naïve way for small values of n. Either there is no solution, in which case we found a local obstruction, or at some point the strong form of Hensel's lemma applies.

Item (1) is probably the deepest result, but we bypassed it here by a trick with rational solutions. There are plenty of cases where there is not even a rational solution. In general, the bound on (1) depends on the *geometry* of the equations in a serious way. This was only understood in the last century.

3.5. **Special congruences.** This section will continue with properties of $\mathbb{Z}/m\mathbb{Z}$. Instead of trying to solve equations, we will look at some identities. Fermat's little theorem in particular will become very useful.

3.5.1. *Wilson's theorem.*

Theorem 3.15. *Let p be prime, then $(p-1)! \equiv -1 \pmod{p}$.*

Proof. The result is clear if $p = 2$. We now assume $p > 2$. Let $\mathbb{F}_p^\times$ denote the set of non-zero congruence classes in $\mathbb{Z}/p\mathbb{Z}$. By Corollary 3.12, taking the inverse is a bijection on $\mathbb{F}_p^\times$. Suppose $x \in \mathbb{F}_p^\times$ satisfies $x = x^{-1}$, then $x^2 = 1$, so by Homework 6.2, $x = \pm 1$. Therefore, the set $\mathbb{F}_p^\times$ can be partitioned into $\frac{p-1}{2} - 1$ pairs $\{a, a^{-1}\}$ and two singletons $\{1\}$, $\{-1\}$.

By definition, $[(p-1)!]_p = \prod_{x \in \mathbb{F}_p^\times} x$. Grouping this product by the partition described above shows that $[(p-1)!]_p = 1^{\frac{p-1}{2}-1} \cdot 1 \cdots (-1) = -1$, as required. □

We cannot think of any application of this theorem.

3.5.2. *Fermat's little theorem.*

Theorem 3.16. *Let p be prime, then for all $a \in \mathbb{Z}$ such that $\gcd(a, p) = 1$,*

$$a^{p-1} \equiv 1 \pmod{p}$$

Proof. Again let $\mathbb{F}_p^\times$ be the set of non-zero congruence classes in $\mathbb{Z}/p\mathbb{Z}$. Consider the function

$$f : \mathbb{F}_p^\times \to \mathbb{F}_p^\times, \; x \mapsto ax$$

Since $\gcd(a, p) = 1$, this is a bijection. Therefore,

$$\prod_{x \in \mathbb{F}_p^\times} x = \prod_{x \in \mathbb{F}_p^\times} f(x) = a^{p-1} \prod_{x \in \mathbb{F}_p^\times} x$$

Again since $\mathbb{F}_p$ is a field, we can cancel out the product to get $a^{p-1} \equiv 1 \pmod{p}$. □

The following theorem is an immediate corollary, but it is so important we are calling it a theorem.

Theorem 3.17. *Let p be a prime, then for all $a \in \mathbb{Z}$,*

$$a^p \equiv a \pmod{p}$$

Proof. Either $a \equiv 0 \pmod{p}$, in which case the result is trivial, or $a \not\equiv 0 \pmod{p}$, in which case the result follows from Fermat's little theorem. □

Consider the polynomial $f(X) = X^p - X$ on $\mathbb{Z}/p\mathbb{Z}$. The above theorem tells us that $f(x) = 0$ for all $x \in \mathbb{Z}/p\mathbb{Z}$. On the other hand, we can find another polynomial of degree p satisfying this property, namely

$$g(X) = \prod_{k=0}^{p-1} (X - k) = X(X-1)\cdots(X-(p-1))$$

Using Lemma 3.7 repeatedly, it is easy to show that $f(X) = g(X)$ as polynomials. In particular, their linear terms agree. The linear term of $f(X)$ is $-X$, whereas the linear term of $g(X)$ is $X(-1)(-2)\cdots(-(p-1)) = (-1)^{p-1}(p-1)!X$, so we immediately recover Wilson's theorem.

3.5.3. *Euler's theorem.* There are many Euler's theorems. The one we are talking about is a generalization of Fermat's little theorem. To state it, we need to define a new function.

Definition. Let $n \in \mathbb{Z}$, and let $(\mathbb{Z}/n\mathbb{Z})^\times$ denote the set of invertible elements in the ring $\mathbb{Z}/n\mathbb{Z}$. The *Euler φ-function* is defined by

$$\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times$$

By Corollary 3.11, we have the following alternative description for $n \geq 1$.

$$\varphi(n) = \#\{1 \leq a \leq n \mid \gcd(a, n) = 1\}$$

Technically, our definition gives $\varphi(0) = 2$, but nobody writes that.

Theorem 3.18 (Euler's theorem)**.** *Let $n \geq 1$ be an integer. Let a be an integer coprime to n, then*

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

Proof. Replace all $\mathbb{F}_p^\times$ by $(\mathbb{Z}/n\mathbb{Z})^\times$ in the proof of Fermat's little theorem. □

Remark. Let S be a finite set with an operation $\cdot$ satisfying the following properties

1. For all $a, b, c \in S$, we have $(ab)c = a(bc)$.
2. There exists $1 \in S$ such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in S$.
3. For all $a \in S$, there exists $a^{-1} \in S$ such that $aa^{-1} = a^{-1}a = 1$.
4. For all $a, b \in S$, we have $ab = ba$.

then the proof of Fermat's little theorem applied to S shows that $a^{\#S} = 1$ for all $a \in S$.

A structure satisfying items (1)–(3) is a *group.* If it satisfies all four, it is an *abelian group.* The statement $a^{\#S} = 1$ for all $a \in S$ is a special case of *Lagrange's theorem*, which applies to all groups, not just the abelian ones. The proof in the general case is not hard but very different, and we will see aspects of it later.

We now give some properties of Euler's φ-function.

Proposition 3.19.

1. *Let m, n be positive integers. If* $\gcd(m, n) = 1$*, then* $\varphi(mn) = \varphi(m)\varphi(n)$.
2. *Let p be a prime and $d \geq 1$, then $\varphi(p^d) = p^{d-1}(p-1)$.*
3. *For a general n,*

$$\varphi(n) = n \prod_{p \mid n} \left(1 - \frac{1}{p}\right)$$

where the product is taken over all prime divisors of n.

Proof.

1. By the Chinese remainder theorem, there is a bijection

$$\mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$

which preserves multiplication. It follows that the bijection restricts to a bijection on invertible elements

$$(\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$$

Counting elements on both sides shows that $\varphi(mn) = \varphi(m)\varphi(n)$.

(2) We use the alternative description in terms of coprime integers. The positive divisors of p^n are exactly $1, p, \cdots, p^n$, so $\gcd(a, p^n) \neq 1$ if and only if $p|a$. There are p^{n-1} integers between 1 and p^n which are multiples of p, so $\varphi(p^n) = p^n - p^{n-1} = p^{n-1}(p-1)$.

(3) This follows from (1) and (2) by unique factorization. □

The next property will be used when we prove the existence of primitive roots.

Theorem 3.20. *Let $n \geq 1$, then*

$$\sum_{d|n} \varphi(d) = n$$

where the sum is taken over all positive divisors of n.

Proof. Let $A = \{1, 2, \cdots, n\}$, then $\#A = n$. For each positive divisor d of n, let $A_d = \{a \in A \mid \gcd(a, n) = d\}$, then the sets A_d partition A, so

$$n = \sum_{d|n} \#A_d$$

Fix a d and suppose $a \in A_d$, then we can write $a = da'$ for a unique $a' \in \mathbb{Z}$. It satisfies the properties that $a' \leq \frac{n}{d}$ and $\gcd\left(a', \frac{n}{d}\right) = \frac{1}{d}\gcd(a, n) = 1$. Therefore, $\#A_d = \varphi\left(\frac{n}{d}\right)$. This proves the result by making the change of variable $d' = \frac{n}{d}$. □

3.6. **Primality testing.** We take a short break to talk about something fun. The question we pose here is

Given a positive integer N, how can we decide if N is prime?

This is a very important problem in modern cryptography.

The immediate thought is to check if $d|N$ for all $d < N$. This is an $O(N)$-algorithm. A moment's reflection shows that we only need to test all $d < \sqrt{N}$ (if $N = ab$, then one of a or b is at most $\sqrt{N}$). This cuts the complexity down to $O(\sqrt{N})$. Both algorithms actually factorizes N if N is not prime, so prime factorization can also be done in $O(\sqrt{N})$-time. These are both exponential time algorithms. In practice, $N \sim 10^{1000}$. The world's computing power can probably do 10^{20} operations per second, so $O(\sqrt{N})$-time would take 10^{480} seconds. In comparison, the age of the universe is about 4×10^{17} seconds.

The fun fact is we can decide if a number is prime without factoring it. One simple observation is the following: if $2^{N-1} \not\equiv 1 \pmod{N}$, then N cannot be prime. At first, this also looks like $O(N)$ operations, but there is a fast way to exponentiate: exponentiation by repeated squaring.

Example. To compute a^6, we can multiply a by itself 6 times, or we can do it in 3 multiplications

$$a^6 = (a^2)^2 \cdot a^2$$

In words: first compute $b = a^2$, then compute $c = b^2$, then compute bc.

The nice thing about modular arithmetic is that all numbers are bounded by N, so multiplication is always $O((\log N)^2)$. The number of multiplications required is the number of binary digits of the exponent, so we can compute $2^{N-1} \pmod{N}$ in $O((\log N)^3)$ time. This is polynomial time.

So it seems like we are done: given N, decide if $2^{N-1} \equiv 1 \pmod{N}$. If not, N is not a prime. Unfortunately, the congruence can still hold if N is composite. Those are called *Fermat pseudoprimes*.

Example. The smallest Fermat pseudoprime is $341 = 11 \times 31$. We can check this quickly on a computer by fast modular exponentiation. Alternatively, since we know the factorization, we can check it by the Chinese remainder theorem.

We need to show that $2^{340} \equiv 1 \pmod{11}$ and $2^{340} \equiv 1 \pmod{31}$. The first congruence is immediate from Fermat's little theorem. The second congruence follows from the observation that $2^5 = 32 \equiv 1 \pmod{31}$.

There is nothing special about 2, so maybe some other number a will work. In the case $N = 341$, we have

$$3^{N-1} \equiv 56 \not\equiv 1 \pmod{N}$$

so N is not a prime, and we say 3 is a witness. Of course, if you happen to choose an a which is not coprime to N, then the congruence will not hold, so a is a witness. The chance of that happening is about $\frac{1}{N}$, so hoping to hit one such a randomly is no better than trial division. One might hope that there are always a large number of witnesses, so a random selection will work, but there are very extreme examples.

Definition. An integer N is a Carmichael number if it is composite, and for all $a \in \mathbb{Z}$ with $\gcd(a, N) = 1$, we have

$$a^{N-1} \equiv 1 \pmod{N}$$

The smallest Carmichael number is $561 = 3 \times 11 \times 17$. The fact that it is a Carmichael number follows easily from Fermat's little theorem and the Chinese remainder theorem. There are infinitely many Carmichael numbers, but they are very rare.

There is a further piece of test we can use: if p is a prime, then $x^2 \equiv 1 \pmod{p}$ implies $x \equiv \pm 1 \pmod{p}$. Since we are computing a^{N-1} by repeated squaring, we might as well check if at any point, we have a counterexample to this fact. This is the Miller–Rabin test:

```
1. Divide $N - 1$ by 2 repeatedly until we find $N - 1 = 2^e \beta$, where $\beta$ is odd.
2. Choose a random $a$ strictly between 1 and $N$.
3. Compute $x = a^\beta \pmod{N}$ by fast modular exponentiation.
4. Repeat $e$ times: set $y = x^2 \pmod{p}$.
     If $y = 1$ but $x \neq \pm 1$, then return $N$ is composite.
     Otherwise, set $x = y$.
5. If $y \not\equiv 1 \pmod{N}$, then return $N$ is composite. Otherwise, go to step 2.
```

Example. Take $N = 561$, then $N - 1 = 2^4 \times 35$. Let $a = 2$, then

$$\begin{aligned} 2^{35} &\equiv 263 \pmod{561} \\ 2^{2\times 35} &\equiv 166 \pmod{561} \\ 2^{4\times 35} &\equiv 67 \pmod{561} \\ 2^{8\times 35} &\equiv 1 \pmod{561} \end{aligned}$$

At this point, we are done: $67^2 \equiv 1 \pmod{561}$, but $67 \not\equiv \pm 1 \pmod{561}$. This proves 561 is not a prime. Moreover, $\gcd(67 + 1, N) = 17$ is a non-trivial divisor of N, which allows us to further factor N.

In the general algorithm, the steps 3 to 5 take polynomial time to complete, and they are actually quite efficient. Moreover, there are no longer any analogues of

Carmichael numbers in this case. In fact, the situation is much better. We state without proof the following theorem.

Theorem 3.21. *If N is composite, then at least $\frac{3}{4}$ of integers between 1 and N are witnesses for the Miller–Rabin test. Therefore, the probability of a false positive is at most $\left(\frac{1}{4}\right)^k$ if we choose k bases randomly.*

The Miller–Rabin test is therefore probabilistically correct, and it is very widely used in this form. But we cannot claim it as a deterministic polynomial time algorithm, since we don't yet know how the bad bases are distributed. This is conjecturally remedied by the following theorem.

Theorem 3.22. *Assuming the* generalized Riemann hypothesis*, there exists a witness within the first $2(\log N)^2$ integers.*

Therefore, the deterministic Miller–Rabin test would test all those integers. This can be done in polynomial time. Unfortunately, it depends on a very hard conjecture. Without GRH, the best upper bound is still a power of N, which makes it an exponential time algorithm. In the next chapter, we will briefly sketch how something like the Riemann hypothesis got involved here.

The story does not end here. In 2001, Agrawal–Kayal–Saxena announced an unconditional deterministic algorithm that decides if N is prime in $O((\log N)^{12})$ time. A revised version with an $O((\log N)^{10.5})$ time is published in Annals in 2004. The exponents are too large for practical uses, but this finally resolves a long-standing theoretical question.

3.7. **Primitive roots.** We saw that $a^{p-1} = 1$ in $\mathbb{F}_p^\times$. Is this the best possible? It is not the best possible if we were working with a composite modulus: $a^4 = 1$ in $(\mathbb{Z}/15\mathbb{Z})^\times$, but $\varphi(15) = 8$.

Definition. Suppose $\gcd(a, m) = 1$. The *(multiplicative) order* of a modulo m is the least positive integer d such that $a^d \equiv 1 \pmod{m}$.

If a has order $\varphi(m)$ modulo m, then we say a is a *primitive root* for m.

Example. Take $m = 31$, then 2 has order 5, and 3 is a primitive root (need to check $3^n \not\equiv 1 \pmod{31}$ for all n such that $1 \le n < 30$).

The main theorem of this section is the following.

Theorem 3.23. *A positive integer m has a primitive root if and only if $m = 2, 4, p^k, 2p^k$, where p is an odd prime and $k \ge 1$. In these cases, there are exactly $\varphi(\varphi(m))$ primitive roots.*

This will be proven over three subsections using a variety of techniques. The first subsection proves the primitive roots exist if m is prime. The second subsection proves it for prime powers. The third subsection cleans up the loose ends by showing no other integers have primitive roots.

We begin with some general results.

Lemma 3.24. *If d is the multiplicative order of a modulo m, and $a^n \equiv 1 \pmod{m}$, then $d|n$. In particular, $d|\varphi(m)$.*

Proof. Write $n = dq + r$ where $0 \le r < d$ and $q, r \in \mathbb{Z}$, then

$$a^r = a^{n-dq} = a^n(a^{-d})^q = 1$$

By the minimality of d, we must have $r = 0$. □

Remark. (1) In the proof, we wrote down expressions of the form a^n, where n can be a negative integer. Formally, we define a^{-1} to be the inverse of a, i.e. the unique element b such that $ab \equiv 1 \pmod{m}$. We then set $a^{-n} = (a^{-1})^n$ for all $n \in \mathbb{N}$. It is easy to check that this has the expected properties, namely $a^{m+n} = a^m a^n$ for all $m, n \in \mathbb{Z}$. We will consistently use this notation.

(2) In fact, if $\gcd(a, m) = 1$, then the set

$$S := \{n \mid a^n \equiv 1 \pmod{m}\}$$

is an ideal in $\mathbb{Z}$, and the order of a is the unique positive generator of this ideal. The proof given above is exactly the proof from Chapter 2 that every ideal is principal, cf. Theorem 2.4 and the remark after Theorem 2.9.

Lemma 3.25. *Let x be an invertible element of $\mathbb{Z}/m\mathbb{Z}$, then x is a primitive root if and only if the set*

$$\langle x \rangle := \{x^n \mid n \in \mathbb{Z}\}$$

is equal to $(\mathbb{Z}/m\mathbb{Z})^\times$.

Proof. By Euler's theorem, $x^{\varphi(m)} = 1$, so

$$\langle x \rangle = \{x^n \mid 0 \leq n < \varphi(m)\}$$

Since $\langle x \rangle \subseteq (\mathbb{Z}/m\mathbb{Z})^\times$, we have equality if and only if $\#\langle x \rangle = \varphi(m)$, which is equivalent to $x^i \neq x^j$ whenever $0 \leq i, j \leq n-1$, $i \neq j$. Taking $i = 0$ shows that x is a primitive root. Conversely, if x is a primitive root and $x^i = x^j$, then $x^{j-i} = 1$. But $0 \leq j - i < n$, so we must have $j - i = 0$. □

Lemma 3.26. *If m has a primitive root, then there are exactly $\varphi(\varphi(m))$ primitive roots for m.*

Proof. By hypothesis, m has a primitive root which we will call g. Let $n = \varphi(m)$, then the previous lemma implies that the function

$$\exp : \mathbb{Z}/n\mathbb{Z} \to (\mathbb{Z}/m\mathbb{Z})^\times, \ k \mapsto g^k$$

is a bijection. Moreover, we have $\exp(a + b) = \exp(a)\exp(b)$. The preimage of $\langle \exp(a) \rangle$ is $\{ka \mid 0 \leq k < n - 1\}$. By our results on linear congruences (Theorem 3.10), this set is the entire $\mathbb{Z}/n\mathbb{Z}$ exactly when $\gcd(a, n) = 1$. Therefore, there are exactly $\varphi(n)$ primitive roots. □

Remark. The previous three lemmas hold for arbitrary groups (defined in the remark after Theorem 3.18). In group theory terms, having a primitive group is equivalent to saying $(\mathbb{Z}/m\mathbb{Z})^\times$ is *cyclic*, that is to say generated by a single element. All cyclic groups of the same size are isomorphic (meaning they are essentially the same), so the multiplicative structure on $(\mathbb{Z}/m\mathbb{Z})^\times$ is equivalent to the additive structure on $\mathbb{Z}/\varphi(m)\mathbb{Z}$. This is exactly what was written out in the previous lemma.

3.7.1. *Prime modulus.* The lemma implies the multiplicative order of all elements of $\mathbb{F}_p^\times$ is a divisor of $p - 1$. Let $d | p - 1$, and let $\psi(d)$ be the number of elements whose order is d. We will show that $\psi(d) = \varphi(d)$. A primitive root is an element whose order is $p - 1$, so this implies there are exactly $\varphi(p - 1)$ primitive roots.

Since every element has a multiplicative order, we have

$$\sum_{d|n} \psi(d) = n$$

Theorem 3.20 states that

$$\sum_{d|n} \varphi(d) = n$$

It remains to prove that $\psi(d) \leq \varphi(d)$ for all d.

Fix a divisor d. If $\psi(d) = 0$, then we are done. Otherwise, let $x \in \mathbb{F}_p^\times$ be an element of order d, then the set $S = \{1, x, \cdots, x^d\}$ has size d, and each member satisfies the equation $X^d = 1$. By Theorem 3.8, these are all of the solutions. It follows that every element of order d is in the set. Given an integer k, if $\ell = \gcd(k, d) > 1$, then

$$(x^k)^{\frac{d}{\ell}} = x^{\frac{k}{\ell} d} = 1$$

so x^k has order strictly less than d. Therefore, the only elements of S that can have order d are those of the form x^k with $\gcd(k, d) = 1$. This proves the assertion $\psi(d) \leq \varphi(d)$, and hence the theorem.

3.7.2. *Prime power modulus.* We will give the usual proof that p^n has a primitive root for all odd primes p. Afterwards, we will explain a simpler proof based on the p-adic logarithm.

Proposition 3.27. *Let g be a primitive root modulo p, then among the integers*

$$\{g + kp \,|\, 0 \leq k < p\}$$

all except for exactly one of them is a primitive root modulo p^2.

Proof. Suppose g has order d modulo p^2. In particular, $g^d \equiv 1 \pmod p$, so by the lemma, $p-1|d$. On the other hand, $d|\varphi(p^2) = p(p-1)$, so $d = p-1$ or $d = p(p-1)$. The same reasoning applies to $g + kp$ for all k. By the binomial theorem

$$(g + kp)^{p-1} = g^{p-1} + (p-1)g^{p-2}kp \equiv 1 + (p-1)g^{p-2}p \pmod{p^2}$$

Therefore, $g + kp$ is a primitive root of p^2 if and only if

$$g^{p-1} + (p-1)g^{p-2}kp \not\equiv 1 \pmod{p}^2$$

Observe that $g^{p-1} \equiv 1 \pmod p$. The above congruence is equivalent to

$$(p-1)g^{p-2}k \not\equiv -\frac{g^{p-1} - 1}{p} \pmod p$$

This holds for all residue classes of $k \pmod p$ except for exactly one. □

Proposition 3.28. *If g is a primitive root modulo p^2, then it is a primitive root modulo p^n for all $n \geq 2$.*

Proof. We will prove that for all $n \geq 2$,

$$g^{\varphi(p^{n-1})} = 1 + p^{n-1}h, \text{ for some } h \in \mathbb{Z},\ p \nmid h$$

Let d be the multiplicative order of g modulo p^n, then given this fact, $\varphi(p^{n-1}) = p^{n-2}(p-1)$ is not a multiple of d. But $d|\varphi(p^n) = p^{n-1}(p-1)$. It follows that $d = \varphi(p^n)$, as required.

This is the lifting the exponent lemma from Homework 7.4. We will include a different write-up of the induction proof here since the statement is important. The base case of the statement follows by hypothesis. Suppose it holds for n, we need to show it for $n+1$. By induction hypothesis, write

$$g^{p^{n-2}(p-1)} = 1 + p^{n-1}h$$

where $h \in \mathbb{Z}$ and $p \nmid h$. Raise this to the p-th power and use binomial expansion.

$$\begin{aligned} g^{p^{n-1}(p-1)} &= (1 + p^{n-1}h)^p \\ &= 1 + p^n h + \binom{p}{2} p^{2(n-1)} h^2 + zp^{3(n-1)} \\ &= 1 + p^n \Big(h + \frac{p-1}{2} p^{n-1} h^2 + zp^{2n-3} \Big) \end{aligned}$$

for some $z \in \mathbb{Z}$ which accounts for the higher order terms. Recall that $n \geq 2$ and p is odd, so the final two terms in the bracket are divisible by p. Since $p \nmid h$, the whole expression in the bracket is not divisible by h. This concludes the proof. □

The proofs might appear unmotivated. The first proposition especially seemed bizarre at first sight. To explain it, we talk about what the theorems mean for the p-adic numbers. The essential idea is that

$$\mathbb{Z}_p^\times = \mu_{p-1} \times (1 + p\mathbb{Z}_p)$$

Here, μ_{p-1} means all solutions to $x^{p-1} = 1$ in $\mathbb{Z}_p$. There are in fact exactly $p - 1$ solutions, obtained by lifting every element of $\mathbb{F}_p^\times$ using Hensel's lemma. The set $1 + p\mathbb{Z}_p$ is a "disc" of radius p^{-1} around 1. The decomposition means that any element $x \in \mathbb{Z}_p^\times$ can be written uniquely as $\zeta(1+z)$, where $\zeta^{p-1} = 1$ and $v_p(z) \geq 1$.

For $x \in \mathbb{Z}_p^\times$, we can construct the subset $\langle x \rangle = \{x^k \mid k \in \mathbb{Z}\}$. This is the same type of set we constructed before, except now it is likely that x has infinite order, in which case this set is infinite. Suppose any element of $\mathbb{Z}_p^\times$ can be approximated arbitrarily well by elements of $\langle x \rangle$, then the image of $\langle x \rangle$ in $\mathbb{Z}/p^n\mathbb{Z}$ is the full set $(\mathbb{Z}/p^n\mathbb{Z})^\times$, since taking image in $\mathbb{Z}/p^n\mathbb{Z}$ is like approximating with n digits of precision. In this case, we say $\mathbb{Z}_p^\times$ is *topologically cyclic*.

We know $\mathbb{Z}_p$ is topologically cyclic with generator 1. This is just saying every p-adic number can be approximated arbitrarily well by an integer, which was exactly accomplished by the base-p expansion. The issue is the operation here is addition. The operation on $\mathbb{Z}_p^\times$ is multiplication.

From calculus, you know that if we define $\log(1 + z) = \sum_{n=1}^\infty (-1)^{n+1} \frac{z^n}{n}$, then

$$\log(xy) = \log x + \log y$$

This maps the positive real numbers to $\mathbb{R}$, and changes multiplication to addition. Over the p-adic number, the series converges too, provided $|z|_p < 1$. Therefore, we can use the p-adic logarithm to convert $1 + p\mathbb{Z}_p$ to $\mathbb{Z}_p$. By carefully keeping track of things, we can therefore show that any number of the form $1 + pz$ with $p \nmid z$ generates $1 + p\mathbb{Z}_p$, for example $1 + p$.

A primitive root for p gives rise to a generator ζ for μ_{p-1}. The two parts don't interact much by the Chinese remainder theorem, so $\zeta \cdot (1+p)$ is a generator for $\mathbb{Z}_p^\times$, and its truncation to n digits is a primitive root for $\mathbb{Z}/p^n\mathbb{Z}$. If we had started with a primitive $g \in \mathbb{Z}$, then g/ζ is an element of $1+p\mathbb{Z}_p$. If we were unlucky, then this is an element of $1+p^2\mathbb{Z}_p$, so it does not generate $\mathbb{Z}_p^\times$. In that case, $g+p \in 1+p+p^2\mathbb{Z}_p$, so it is a generator.

This nicely explains why p^2 is a special case. Moreover, in the case $p = 2$, the same analysis shows that $1 + 2^2\mathbb{Z}_2$ is topologically cyclic, but $1 + 2\mathbb{Z}_2$ is not. This is why in the next subsection, the number 5 appears.

3.7.3. *Odds and ends.* This subsection will finish up the proof of the theorem. We start with a lemma which refines Euler's theorem.

Lemma 3.29. *Let* $m = \prod_{i=1}^{n} p_i^{f_i}$ *be its prime factorization. Let*

$$\psi(m) = \operatorname{lcm}(\varphi(p_i^{f_i})|_{1 \leq i \leq n})$$

then $a^{\psi(m)} \equiv 1 \pmod{m}$ *whenever* $\gcd(a, m) = 1$.

Proof. By construction and Euler's theorem, $a^{\psi(m)} \equiv 1 \pmod{p_i^{f_i}}$ for all i. The claim now follows from the Chinese remainder theorem. □

Corollary 3.30. *If* m *is not of the form* p^n *or* $2p^n$*, where* p *is any prime, then* m *has no primitive root.*

Proof. Observe that if $\psi(m) < \varphi(m)$, then m has no primitive root. This happens in particular when for some $i \neq j$, $\varphi(p_i^{f_i})$ and $\varphi(p_j^{f_j})$ are not coprime.

If m is divisible by two odd primes, then the factor corresponding to them share a common factor of 2. If m is divisible by 4 and an odd prime, then the same holds. The only remaining cases are the exceptions listed in the corollary. □

Now we complete everything. The case $m = 2p^n$ for an odd prime p is easy: any odd integer which is a primitive root for p^n is also a primitive root for $2p^n$ by the Chinese remainder theorem. It remains to consider the case $m = 2^n$. By direct computation, $2, 4$ have primitive roots. However, $a^2 \equiv 1 \pmod{8}$ for all odd a, so 8 has no primitive root. Finally, if 2^n has a primitive root for $n \geq 3$, then its image in $\mathbb{Z}/8\mathbb{Z}$ would also be a primitive root by applying Lemma 3.25. This completes the proof of the theorem.

Exercise. We can refine the result for powers of 2. Show that 5 has order $\frac{1}{2}\varphi(2^n)$ modulo 2^n, and -1 is not in $\langle 5 \rangle$.

4. Quadratic reciprocity law

The goal of this entire chapter is to answer what seems like a very basic question: when does the equation $x^2 \equiv a \pmod{m}$ have a solution? Of course, this is immediately reduced to the case when m is a prime power by the Chinese remainder theorem. Hensel's lemma shows that we can lift a solution modulo p to a solution modulo any of its power, provided $p \neq 2$. For powers of 2, the strong version says this has a solution if and only if $a \equiv 1 \pmod{8}$.

Therefore, everything reduces down to the question: if p is an odd prime, and $a \in \mathbb{Z}$, when does $x^2 \equiv a \pmod{p}$ have a solution? The answer is given by the law of quadratic reciprocity, an amazing and deep result that is the start of a long story in modern number theory.

4.1. Legendre symbol.

Definition. Let p be an odd prime. Let a be coprime to p. We say a is a *quadratic residue* modulo p if the equation $x^2 \equiv a \pmod{p}$ has a solution. Otherwise, we say a is a *quadratic non-residue.*

Definition. Let p be an odd prime and $a \in \mathbb{Z}$. The *Legendre symbol* is defined by

$$\left(\frac{a}{p}\right) := \begin{cases} +1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic non-residue modulo } p \\ 0 & \text{if } p|a \end{cases}$$

Lemma 4.1. *The number of solutions to the equation $x^2 \equiv a \pmod{p}$ is $1 + \left(\frac{a}{p}\right)$.*

Proof. The only thing that needs proving is the case a is a quadratic residue. In that case, if x is a square root, then so is $-x$. They are distinct since $p \neq 2$. Moreover, Theorem 3.8 implies that there are at most two solutions. □

Clearly, the Legendre symbol only depends on the image of a in $\mathbb{F}_p$, so we can view it as a function from $\mathbb{F}_p$ to $\{-1, 0, 1\}$. When restricted to $\mathbb{F}_p^\times$, it gives a function $\mathbb{F}_p^\times \to \{\pm 1\}$. To understand the function in greater depth, we use the fact that p has a primitive root.

Proposition 4.2. *Let p be an odd prime.*

(1) *There are exactly $\frac{p-1}{2}$ residues and non-residues.*
(2) *Euler's criterion: if $a \in \mathbb{F}_p^\times$, then*

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}$$

(3) *The Legendre symbol is a* character, *meaning for all $a, b \in \mathbb{F}_p^\times$*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$

Proof. Let g be a primitive root for p, then all elements of $\mathbb{F}_p^\times$ are of the form g^k for $k \in \mathbb{Z}/(p-1)\mathbb{Z}$. The quadratic residues have the form g^{2k} for an integer k. Since $2|p-1$, elements of the form g^{2k+1} are quadratic non-residues. This proves item (1). If $a \in \mathbb{F}_p^\times$, then we know that $a^{\frac{p-1}{2}} = \pm 1$ since its square is 1. Moreover, since g is a primitive root

$$(g^{2k})^{\frac{p-1}{2}} = g^{k-1} = 1, \quad (g^{2k+1})^{\frac{p-1}{2}} = g^{\frac{p-1}{2}} \neq 1$$

which proves (2). Item (3) is an immediate consequence of (2). □

Exercise. Prove the above without using primitive roots.

This proposition already has some immediate applications.

Corollary 4.3. *Let p be an odd prime, then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} +1 & p \equiv 1 \pmod 4 \\ -1 & p \equiv 3 \pmod 4 \end{cases}$$

Proposition 4.4. *Suppose $n \geq 1$ is exactly divisible by a prime q such that $q \equiv 3 \pmod 4$, then n is not a sum of two squares.*

Proof. Suppose $n = x^2 + y^2$. If $q | x$, then $q | y$, so $q^2 | n$. This contradicts the hypothesis, so $q \nmid x$. Reduce the equation modulo q, we get $x^2 + y^2 \equiv 0 \pmod q$. Since $q \nmid x$, there exists $z \in \mathbb{Z}$ such that $xz \equiv y \pmod q$. But then $z^2 \equiv -1 \pmod q$, contradicting the corollary. □

Remark. A converse which we will not prove is Fermat's theorem on two squares: every prime congruent to 1 modulo 4 is a sum of two squares. In general, there is a formula to count how many ways are there to represent an integer n as a sum of two squares.

4.1.1. **Generalized Riemann Hypothesis**. We can modify the definition of the Riemann zeta function by inserting signs.

Definition. The *Dirichlet L-function* attached to the Legendre symbol is

$$L(s, \chi_p) = \sum_{n=1}^{\infty} \left(\frac{n}{p}\right) n^{-s}$$

Example. If $p = 3$, then

$$L(s, \chi_3) = 1 - \frac{1}{2^s} + \frac{1}{4^s} - \frac{1}{5^s} + \frac{1}{7^s} - \frac{1}{9^s} + \cdots$$

In particular,

$$L(1, \chi_3) = \sum_{n=0}^{\infty} \frac{1}{(3n+1)(3n+2)} = \frac{\pi}{3\sqrt{3}}$$

If $p = 5$, then

$$\begin{aligned} L(1, \chi_5) &= \sum_{n=0}^{\infty} \left(\frac{1}{5n+1} - \frac{1}{5n+2} - \frac{1}{5n+3} + \frac{1}{5n+4} \right) \\ &= 10 \sum_{n=0}^{\infty} \frac{2n+1}{(5n+1)(5n+2)(5n+3)(5n+4)} \\ &= \frac{2}{\sqrt{5}} \log \frac{1+\sqrt{5}}{2} \end{aligned}$$

The two explicit values are not obvious, and they actually carry information. In the case $p = 3$, it is related to the failure of unique factorization in certain generalizations of $\mathbb{Z}$. In the case $p = 5$, it is related to solutions to the Pell equation.

Using the fundamental theorem of arithmetic and the multiplicative property of the Legendre symbol, the same proof before shows that $L(s,\chi_p)$ also has an infinite product expression.

$$L(s,\chi_p)=\prod_{\ell\neq p}\left(1-\left(\frac{\ell}{p}\right)p^{-s}\right)^{-1}$$

Here, the product is over all primes ℓ not equal to p. It is a general fact that $L(1,\chi_p)\neq 0$, though this is very difficult to prove. This is the key input in the following refinement of the prime number theorem:

Theorem 4.5 (Dirichlet's theorem on primes in arithmetic progression)**.** *Suppose* $\gcd(a,q)=1$. *Let* $\pi(x,a;q)$ *be the number of primes less than* x *that are congruent to* a *modulo* q*, then*

$$\pi(x,a;q)\sim\frac{1}{\varphi(q)}\frac{x}{\log x}$$

In other words, the primes are asymptotically evenly distributed among the possible congruence classes.

To obtain a good error term in this theorem, it is natural we need something analogous to the Riemann hypothesis.

Conjecture (Generalized Riemann Hypothesis)**.** *The only zeroes of* $L(s,\chi_p)$ *with* $0\leq \mathrm{Re}(s)\leq 1$ *lie on the line* $\mathrm{Re}(s)=\frac{1}{2}$.

We consider another application. One may be interested in the question of the least quadratic non-residue, that is the least positive integer a such that $\left(\frac{a}{p}\right)=-1$. Suppose we can prove something like

$$\sum_{a\leq x}\left(\frac{a}{p}\right)\overset{?}{\leq}\sqrt{x}\log p$$

for all x. If $\left(\frac{a}{p}\right)=1$ for all $a\leq x$, then the sum is exactly x. The inequality then shows that $x\leq(\log p)^2$, proving that the least quadratic non-residue is at most equal to $(\log p)^2$. Therefore, the question of least quadratic non-residue is closely related to the question of estimating so-called character sums. Such a square-root upper bound is a sign that $\left(\frac{a}{p}\right)$ is "randomly" ±1.

Now take the definition of the Dirichlet L-function and let $s=0$, then we get

$$L(0,\chi_p)\overset{?}{=}\sum_{n=1}^{\infty}\left(\frac{n}{p}\right)$$

this doesn't make sense unless you are Euler. But by techniques from complex analysis, there is a way to "take a limit $\lim_{s\to 0}$" and obtain a sensible answer. By throwing in some cut-off functions, we might hope to relate $L(s,\chi_p)$ to the character sum. The issue is all the complex analysis magic depends on knowing the roots of $L(s,\chi_p)$ for $0\leq\mathrm{Re}(s)\leq 1$. Assuming the GRH, the above strategy can be carried out with some care. Without the GRH, the best upper bound we know for the least quadratic non-residue is $O\big(p^{\frac{1}{4\sqrt{e}}+\varepsilon}\big)$.

Exercise. Find an elementary argument to show that the least quadratic non-residue is at most $\sqrt{p}+1$.

Almost identical discussions apply to the Miller–Rabin witness problem. It is elementary but slightly annoying to prove that all non-witnesses to the Miller–Rabin algorithm lie in a proper subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ (where by subgroup we just mean closed under multiplication). In the least quadratic non-residue case, the subgroup is the set of residues. In the Miller–Rabin case, we need to find the least witness. The usual version of GRH is more general than the one we stated and covers the relevant characters appearing in the Miller–Rabin case.

4.2. **Quadratic reciprocity.** We now turn to the question of computing the Legendre symbol. Here, we see a miracle.

Theorem 4.6 (Quadratic reciprocity)**.** *Let p, q be odd primes, then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$$

Somehow, the equations $x^2 \equiv p \pmod{q}$ and $x^2 \equiv q \pmod{p}$ are related. Gauss noticed and proved this by induction when he was 19. It was a very long proof with 8 cases each split into many sub-cases. He also introduced the congruence notation that we are using. In modern language, this proof translates to a computation of some really sophisticated object known as $K_2(\mathbb{Q})$. I believe this connection was first observed by Tate.

Unsatisfied with this, Gauss proceeded to publish five other completely different proofs. The proof in the textbook is Eisenstein's modification of Gauss' third published proof. We will give a different proof using Gauss sums in the next section, which was apparently the final proof Gauss discovered. This proof also generalizes to higher powers.

Philosophically, we have said that by the Chinese remainder theorem, different primes act more or less independently. What quadratic reciprocity says is that it's not completely independent. Investigations into version of quadratic reciprocity for higher powers eventually led to what's now called *class field theory*, which clarifies a small part of these relations. Further investigations beyond class field theory are part of the *Langlands program*, a very active area of research.

Back to actual course content, we need two supplementary results for the missing cases. We have already proven one of them.

Theorem 4.7 (Supplements)**.** *Let p be an odd prime, then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} +1 & p \equiv 1 \pmod{4} \\ -1 & p \equiv 3 \pmod{4} \end{cases}$$

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} +1 & p \equiv 1, 7 \pmod{8} \\ -1 & p \equiv 3, 5 \pmod{8} \end{cases}$$

We will prove both of these theorems in the next section. The rest of this section is used to show how they can be applied to effectively compute the Legendre symbol.

If we can factor a, then quadratic reciprocity can be used to reduce the computation of $\left(\frac{a}{p}\right)$ to something with a smaller base, similar to Euclid's algorithm. Unfortunately, factoring is hard. Instead, we introduce an intermediate device.

Definition. Let m, n be integers and suppose n is positive and odd. Let $n = \prod_{i=1}^{k} p_i^{f_i}$ be its prime factorization. Then the *Jacobi symbol* $\left(\frac{m}{n}\right)$ is defined by

$$\left(\frac{m}{n}\right) := \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{f_i}$$

Remark. Just knowing $\left(\frac{m}{n}\right) = 1$ does not implies m is a quadratic residue modulo n. Deciding quadratic residue in general is as hard as factoring.

By definition, the Jacobi symbol agrees with the Legendre symbol when the base is a prime. It is also completely multiplicative in the numerator. Moreover, using quadratic reciprocity, it is possible to show the following theorem.

Theorem 4.8. *Let m, n be positive odd integers, then*

$$\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = (-1)^{\frac{m-1}{2}\frac{n-1}{2}}$$
$$\left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}$$
$$\left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}$$

In other words, the statements of quadratic reciprocity are unchanged in the more general setting.

Proof. This is messy but not difficult. We prove the final statement as an example.

Let $n = \prod_{i=1}^{k} p_i$ be a factorization n into a product of (not necessarily distinct) primes, then

$$\left(\frac{2}{n}\right) = \prod_{i=1}^{k} \left(\frac{2}{p_i}\right) = (-1)^{\sum_{i=1}^{k} \frac{p_i^2-1}{8}}$$

so we just need to show $\sum_{i=1}^{k}(p_i^2 - 1) \equiv n^2 - 1 \pmod{16}$. This will be done by induction on k. For the induction step, we consider what happens when n is replaced by np and prove that

(1) If $p \equiv 1, 7 \pmod 8$, then both sides do not change.
(2) If $p \equiv 3, 5 \pmod 8$, then both sides change by 8.

The left hand side part is immediate. For the right hand side, if $p \equiv 1, 7 \pmod 8$, then $p^2 \equiv 1 \pmod{16}$, otherwise $p^2 \equiv 9 \pmod{16}$. Therefore, in the first case, there is no change. In the second case, we are changing $n^2 - 1$ to $9n^2 - 1$. The difference is $8n^2$, which is congruent to 8 modulo 16 since n is odd. □

Using these facts, it is clear how to compute Jacobi symbols in general using an algorithm exactly like Euclid's algorithm. In particular, this allows us to efficiently compute Legendre symbols, which are actually useful.

Example. We compute the Legendre symbol $\left(\frac{66}{101}\right)$.

$$\left(\frac{66}{101}\right) = \left(\frac{2}{101}\right)\left(\frac{33}{101}\right) = -\left(\frac{33}{101}\right)$$

Now we use quadratic reciprocity,

$$-\left(\frac{33}{101}\right) = -\left(\frac{101}{33}\right) = -\left(\frac{2}{33}\right) = -1$$

Therefore, 66 is not a quadratic residue modulo 101.

We could have done the computation using the formula $a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}$ using fast modular exponentiation. Recall that Euclid's algorithm uses $O((\log p)^2)$ bit operations, and the algorithm using Jacobi symbols has the same computations. On the other hand, fast modular exponentiation requires $\log p$ multiplications of integers of size p, so it takes $O((\log p)^3)$ bit operations.

4.3. **Proofs.** As we alluded to before, there are many proof of quadratic reciprocity. We will give a proof based on Gauss sums, following Lang's *Algebraic Number Theory.* The idea is to "invent" a square root of p and check it is actually real using some criterion based on Theorem 3.17.

4.3.1. *Supplement.* We start with the supplementary statement

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$$

This proof uses the same idea as the main proof, but it involves less technicalities. Here is the idea: by Euler's criterion, we need to prove that

$$\sqrt{2}^{p-1} \equiv (-1)^{\frac{p^2-1}{8}} \pmod{p}$$

for some integer d. Here, the left hand side is a rational number since $p-1$ is even. If we multiply both sides by $\sqrt{2}$, then we need to show that

$$\sqrt{2}^{p} \equiv (-1)^{\frac{p^2-1}{8}}\sqrt{2} \pmod{p}$$

We will make sense of this statement later. For now, the right hand side depends on $p \pmod 8$, so we need to bring in some object on the left hand side that also obviously depends on $p \pmod 8$.

Let $\zeta = \exp\left(\frac{2\pi i}{8}\right) = \frac{1+i}{\sqrt{2}}$ be an 8th root of unity, then $\zeta + \zeta^{-1} = \sqrt{2}$. By the binomial theorem

$$(\zeta + \zeta^{-1})^p = \zeta^p + \zeta^{-p} + \sum_{k=1}^{p-1} \binom{p}{k} \zeta^k \zeta^{-(p-k)}$$

If $p \equiv \pm 1 \pmod 8$, then the first two terms sum to $\sqrt{2}$. Otherwise, they sum to $-\sqrt{2}$. Therefore,

$$\sqrt{2}^p = (-1)^{\frac{p^2-1}{8}}\sqrt{2} + \sum_{k=1}^{p-1} \binom{p}{k} \zeta^k \zeta^{-(p-k)}$$

This looks really promising. In fact, observe that $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is always an integer. The numerator is divisible by p, but if $1 \leq k \leq p-1$, neither term in the denominator is divisible by p (since p is a prime). Therefore, $p | \binom{p}{k}$, and

$$\sqrt{2}^p = (-1)^{\frac{p^2-1}{8}}\sqrt{2} + pz, \quad z = \sum_{k=1}^{p-1} \frac{1}{p}\binom{p}{k} \zeta^{2k-p}$$

We can now rearrange this expression and square it to get

$$\left(2^{p-1} - (-1)^{\frac{p^2-1}{8}}\right)^2 \cdot 2 = p^2 z^2$$

It follows that $x = z^2$ is a rational number. If we can show that $x \in \mathbb{Z}$, then p divides the left hand side. Since p is an odd prime, we must have the term in the bracket is divisible by p. This is exactly what Euler's criterion requires us to prove.

Observe that $\zeta^4 = -1$, so we can write x as a linear combination of $1, \zeta, \zeta^2, \zeta^3$, and moreover the coefficients are integers. Therefore,

$$x = a + b\Big(\frac{1+i}{\sqrt{2}}\Big) + ci + d\Big(\frac{-1+i}{\sqrt{2}}\Big) = \Big(a - \frac{b-d}{\sqrt{2}}\Big) + \Big(c + \frac{b+d}{\sqrt{2}}\Big)i$$

where $a, b, c, d \in \mathbb{Z}$. For this to be in $\mathbb{Q}$, we must have $c + \frac{b+d}{\sqrt{2}} = 0$ and $a - \frac{b-d}{\sqrt{2}} \in \mathbb{Q}$. Rearranging the two equations shows that $b\sqrt{2}, d\sqrt{2} \in \mathbb{Q}$, which implies $b = d = 0$. The first equation then shows that $c = 0$. Therefore, the expression is just equal to a, which is an integer.

4.3.2. *Main part of proof.* It is useful to slightly reformulate the statement. Given an odd prime p, define

$$p^* := \Big(\frac{-1}{p}\Big)p = \begin{cases} p & p \equiv 1 \pmod 4 \\ -p & p \equiv 3 \pmod 4 \end{cases}$$

then we can check by case work that

$$\Big(\frac{p^*}{q}\Big) = \begin{cases} \big(\frac{p}{q}\big) & p \equiv 1 \pmod 4 \text{ or } q \equiv 1 \pmod 4 \\ -\big(\frac{p}{q}\big) & p, q \equiv 3 \pmod 4 \end{cases} = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\Big(\frac{p}{q}\Big)$$

Therefore, the law of quadratic reciprocity is equivalent to

$$\Big(\frac{p^*}{q}\Big) = \Big(\frac{q}{p}\Big)$$

We now try to invent a square root of p^*.

Theorem 4.9. *Let* $\zeta = \exp\big(\frac{2\pi i}{p}\big)$. *Define*

$$S = \sum_{a=0}^{p-1}\Big(\frac{a}{p}\Big)\zeta^a$$

then $S^2 = p^*$.

Proof. First note that the terms of the sum depend only on $a \pmod p$, so we can write the sum as a sum over $\mathbb{F}_p$. In particular, we will freely use the arithmetic operations over $\mathbb{F}_p$. Now expand the product,

$$S^2 = \sum_{a,b \in \mathbb{F}_p}\Big(\frac{ab}{p}\Big)\zeta^{a+b}$$

Let $c = a + b$, and first sum over c, then the sum rearranges to

$$S^2 = \sum_{c \in \mathbb{F}_p}\zeta^c \sum_{a \in \mathbb{F}_p}\Big(\frac{a(c-a)}{p}\Big)$$

For the inner sum, if $a = 0$, then the Legendre symbol is 0, so it doesn't contribute. If $a \neq 0$, then

$$\Big(\frac{a(c-a)}{p}\Big) = \Big(\frac{a^2(ca^{-1}-1)}{p}\Big) = \Big(\frac{ca^{-1}-1}{p}\Big)$$

If $c = 0$, then this is always equal to $\left(\frac{-1}{p}\right)$. Otherwise, as a runs over all elements in $\mathbb{F}_p^\times$, the element ca^{-1} also runs over all elements in $\mathbb{F}_p^\times$. In this case,

$$\sum_{a \in \mathbb{F}_p} \left(\frac{a(c-a)}{p}\right) = \sum_{a' \in \mathbb{F}_p^\times} \left(\frac{a'-1}{p}\right) = \sum_{a'' \in \mathbb{F}_p} \left(\frac{a''}{p}\right) - \left(\frac{-1}{p}\right)$$

For the final sum, recall that exactly half of the elements of $\mathbb{F}_p^\times$ are quadratic residues, so the sum is 0.

Finally, plugging everything into our expression gives

$$\begin{aligned} S^2 &= (p-1)\left(\frac{-1}{p}\right) + \sum_{c \in \mathbb{F}_p^\times} \zeta^c \left(-\left(\frac{-1}{p}\right)\right) \\ &= \left(\frac{-1}{p}\right)\left(p - 1 - \sum_{c \in \mathbb{F}_p^\times} \zeta^c\right) \\ &= \left(\frac{-1}{p}\right)\left(p - \sum_{c=0}^{p-1} \zeta^c\right) \end{aligned}$$

Using the geometric series formula, the final sum is 0, so $S^2 = \left(\frac{-1}{p}\right)p$. □

Remark. The Gauss sum is the discrete Fourier transformation of the Legendre symbol character, so it is of independent interest. The value of S^2 is an easy consequence of the Plancherel formula, so this theorem is not as mysterious as it first appears.

Following the proof of the supplementary case, the next step is to compute S^q and observe that many terms are divisible by q. It will be helpful to introduce a notation for the type of numbers we are dealing with.

Definition. Let $\mathbb{Z}[\zeta]$ denote the set of complex numbers of the form $\sum_{k=0}^{p-1} a_k \zeta^k$, where $a_0, \cdots, a_{p-1} \in \mathbb{Z}$.

In particular $S \in \mathbb{Z}[\zeta]$, and $\mathbb{Z}[\zeta]$ forms a ring. We can now define congruence modulo q in the same way. Instead of introducing the notion formally, we will write out everything in equation forms.

Lemma 4.10. *There exists $z \in \mathbb{Z}[\zeta]$ such that*

$$S^q = \left(\frac{q}{p}\right) S + qz$$

Proof. The multinomial expansion formula is

$$(x_1 + \cdots + x_n)^q = \sum_{d_1 + \cdots + d_n = q} \frac{q!}{d_1! \cdots d_n!} x_1^{d_1} \cdots x_n^{d_n}$$

The coefficients are all integers. Observe that if $d < q$, then $q \nmid d!$ since q is prime. Therefore, if $d_k < q$ for all $k = 1, \cdots, n$, then the coefficient for that term is a multiple of q. The only other terms are $x_1^q + \cdots + x_n^q$. We have shown that

$$(x_1 + \cdots + x_n)^q = x_1^q + \cdots + x_n^q + qR$$

where R is an integer combination of products of the various x_k.

We apply this to S^q. First we compute the sum of the q-th power of each term.

$$\begin{aligned}\sum_{a\in\mathbb{F}_p}\left(\frac{a}{p}\right)^q\zeta^{aq} &= \sum_{a'\in\mathbb{F}_p}\left(\frac{a'q^{-1}}{p}\right)\zeta^{a'}\\ &= \left(\frac{q^{-1}}{p}\right)\sum_{a'\in\mathbb{F}_p}\left(\frac{a'}{p}\right)\zeta^{a'}\\ &= \left(\frac{q}{p}\right)S\end{aligned}$$

The lemma follows from our discussion on multinomial expansion. □

Remark. Again, this is less mysterious than it might first appear. The map $x\mapsto x^q$ is called the *Frobenius*. In a ring where $q=0$, this map preserves addition by the exact proof we have given. Moreover, Theorem 3.17 tells us that $\mathbb{F}_q$ is exactly the fixed points of the Frobenius. In a very precise sense, we constructed an *extension* of $\mathbb{F}_q$ containing a square root of p^*, and now we are using the Frobenius to check if this square root is in $\mathbb{F}_q$ or not.

Writing out Euler's criterion without the congruence notation gives

$$(p^*)^{\frac{q-1}{2}} = \left(\frac{p^*}{q}\right) + qn$$

for some $n\in\mathbb{Z}$. Since $S^2=p^*$, we have

$$S^q = (S^2)^{\frac{q-1}{2}}S = \left(\frac{p^*}{q}\right)S + qz'$$

where $z'\in\mathbb{Z}[\zeta]$. By comparing the this equation with the previous lemma, we get

$$\left(\left(\frac{p^*}{q}\right)-\left(\frac{q}{p}\right)\right)S\in q\mathbb{Z}[\zeta]$$

Let $\delta\in\{0,\pm2\}$ be the difference of the two Legendre symbols. Squaring both sides of the equation gives

$$\delta^2p^* = q^2x$$

where $x\in\mathbb{Z}[\zeta]$. Since everything else is rational, we also have $x\in\mathbb{Q}$. What remains is to prove that $\mathbb{Z}[\zeta]\cap\mathbb{Q}=\mathbb{Z}$. Given this, the above equality implies $q|\delta^2p^*$. But q is an odd prime distinct from p and $\delta^2\in\{0,\pm4\}$, so the only way this divisibility can hold is if $\delta=0$, which proves quadratic reciprocity.

In the supplementary case, we proved it by hand, but now it is necessary to be systematic. This fact is actually a part of some general results in algebraic number theory, completely independent of what we have discussed so far. It is completely proven in the next subsection, but the proof is not too important for this course.

4.3.3. **Algebraic number theory*.* Our proof is more or less the standard one. Let $x\in\mathbb{Z}[\zeta]$, then we will find an integer N and integer coefficients $a_1,\cdots,a_N$ such that

$$x^N + a_1x^{N-1} + \cdots + a_{N-1}x + a_N = 0$$

It follows from Theorem 2.20 that if $x\in\mathbb{Q}$, then $x\in\mathbb{Z}$.

Let $x=\sum_{k=0}^{p-1}c_k\zeta^k$, where c_k are integers. Define the polynomial

$$f(X_0,\cdots,X_{p-1}) = \sum_{k=0}^{p-1}c_kX_k$$

then $x = f(\zeta^0, \zeta, \cdots, \zeta^{p-1})$. Further, define

$$\begin{aligned} g(X, X_0, \cdots, X_{p-1}) &= \prod_{\sigma} (X - f(X_{\sigma(0)}, \cdots, f_{\sigma(p-1)})) \\ &= X^N + g_1(X_0, \cdots, X_{p-1})X^{N-1} + \cdots + g_N(X_0, \cdots, X_{p-1}) \end{aligned}$$

Here, the product runs over all permutations σ of the set $\{0, 1, \cdots, p-1\}$, so $N = p!$. Each coefficient is a polynomial in p variables whose coefficients all lie in $\mathbb{Z}$.

The key point is that by construction, each polynomial g_k is *symmetric*, namely for all k and all permutations σ,

$$g_k(X_0, \cdots, X_{p-1}) = g_k(X_{\sigma(0)}, \cdots, f_{\sigma(p-1)})$$

Therefore, it is a polynomial in the elementary symmetric polynomials. We can also define elementary symmetric polynomials $e_1, \cdots, e_p$ using the equation

$$\prod_{k=0}^{p-1} (X - X_k) = X^{p-1} + e_1(X_0, \cdots, X_{p-1})X^{p-2} + \cdots + e_p(X_0, \cdots, X_{p-1})$$

For example, $e_p(X_0, \cdots, X_{p-1}) = (-1)^p X_0 \cdots X_{p-1}$ and $e_1 = -(X_0 + \cdots + X_{p-1})$.[5] It is not too difficult to prove by induction on degree that any symmetric polynomial can be written as a polynomial in the elementary symmetric polynomials. For example,

$$X_0^3 + X_1^3 = (X_0 + X_1)^3 - 3X_0X_1(X_0 + X_1) = -e_1^3 - 3e_3e_1$$

Now, we substitute in $X_k = \zeta^k$. Observe that

$$\prod_{k=0}^{p-1} (X - \zeta^k) = X^p - 1$$

Therefore, all the elementary symmetric polynomials evaluate to integers. It follows that all $g_k(\zeta^0, \cdots, \zeta^{p-1})$ are integers. The polynomial $g(X, \zeta^0, \cdots, \zeta^{p-1})$ satisfies the required conditions.

[5] Usually, the definition does not include the sign.

5. Diophantine equations

Diophantine equations are polynomial equations with integer coefficients. For example, each of the following is a Diophantine equation.

$$x^2 + y^2 = z^2 \tag{1}$$

$$y^2 = x^3 + 2x - 1 \tag{2}$$

$$y^2 = x^3 - 2x \tag{3}$$

$$x^3 + y^3 + z^3 = 42 \tag{4}$$

One typically asks for solutions either in integers or in the rational numbers. This makes the question difficult in general.

(1) Equation (1) is the classical question about Pythagorean triples. We will see that it has infinitely many integer solutions, and they can all be written down in very simple formulae. The problems of integer and rational solutions are equivalent since multiplying each variable by the same rational number results in the same equation.

(2) Equation (2) defines an elliptic curve. We have shown by reduction mod 3 that it has no integer solution. It turns out this equation also has no rational solutions, but this fact is much harder.

(3) Equation (3) is also an elliptic curve. It has the obvious integer solutions $(x, y) = (-1, \pm 1), (0, 0), (2, \pm 2)$. Less obviously, $(x, y) = (338, \pm 6214)$ are also solutions. These turn out to be all integer solutions.

    There are infinitely many rational solutions, examples include

$$\Big(\frac{9}{4}, \pm\frac{21}{8}\Big),\ \Big(\frac{12769}{7056}, \pm\frac{900271}{592704}\Big), \cdots$$

    There is no explicit formula like in the case of Pythagorean triples, but there is an easy algorithmic process to generate all points.

(4) There are algebraic identities that represent every integer as a sum of three rational cubes. The case of integers is much harder. Booker and Sutherland discovered in 2019 that 42 is a sum of three integer cubes, with the minimal solution being

$$\begin{aligned} 42 = {} & (-80538738812075974)^3 \\ & + 80435758145817515^3 + 12602123297335631^3 \end{aligned}$$

    This was done by a very clever computer search.

5.1. *Hilbert's 10th problem*.

In the early 20th century, Hilbert posed 23 questions which he considered important. The 10th problem is to find an "algorithmic process" to determine if a Diophantine equation has integer solutions. Of course, you can just start testing all integers to see if it is a solution. If there is a solution, then this terminates. The problem is that this does not terminate if there is no solution, but you don't know that at any point in the process.

This might remind you of the halting problem: given a program, determine if it terminates. Again, you can run it until it terminates, but you never know when to stop if it doesn't halt. The fundamental result of computation theory is that no algorithm exists to decide if a program halts. Maybe not surprisingly, Hilbert's 10th problem has no solution. What is surprising is that the proof shows that Diophantine equation simulates all computations.

Theorem 5.1 (Davis–Putnam–Robinson–Matiyasevich)**.** *Let P be a program with input on $\mathbb{N}$. There exists a Diophantine equation depending on a natural number parameter n which has an integer solution if and only if $P(n)$ terminates.*

In particular, there is no algorithm to decide if a Diophantine equation has integer solutions.

Here is a consequence: there is a program which lists out all possible texts and check if it is a proof that PA implies $0 = 1$. This program terminates if and only if PA is inconsistent. Therefore,

Corollary 5.2. *There is a Diophantine equation which has an integer solution if and only if* PA *is inconsistent.*

We now briefly talk about the proof. There are three steps:

(1) Davis: any computable subset of $\mathbb{N}$ can be described by an expression of the form

$$(\exists z)(\forall y \leq z)(\exists x)(P(a, x, y, z) = 0)$$

where $P(a, x, y, z)$ is a polynomial with integer coefficients. This is a consequence of the proof that PA can define recursion, based on the Gödel β-numbers introduced on Homework 6.4. For example, the following long sentence with free variables x and y

$$(\exists a, N)\big(\beta(a, N, 0) = 1 \wedge \beta(a, N, x) = y \wedge (\forall i < x)(\beta(a, N, i + 1) = 2\beta(a, N, i))\big)$$

is equivalent to $y = 2^x$. Here, we are using the integer N to code the finite sequence $(1, 2, \cdots, 2^x)$. The relation $\beta(a, N, i) = b$ can be further unwound into something Diophantine.

We now need to eliminate the middle $(\forall y \leq z)$.

(2) Davis–Putnam–Robinson: The bounded universal quantifier can be removed if we allow exponents. This is done using a more sophisticated application of the Gödel numbering idea. To bound the size of parameters required, we need exponentiation.

(3) Robinson+Matiyasevich: Robinson observed that as long as we can use Diophantine equations to represent a function that's roughly exponential, this is enough.

Matiyasevich then wrote down a system of 8 equations, 12 variables, and 2 parameters (x, n) such that the system has a solution if and only if $x = F_n$, the n-th Fibonacci number. This uses entirely elementary properties of Fibonacci numbers and Pell equation (a topic we unfortunately do not have time to cover), but it also came basically out of nowhere.

I gave a talk on the proof, and I have uploaded the notes to Canvas. It contains a more expanded version of this proof, as well as the actual definition of computability.

5.2. **Fermat's last theorem.**

Famously, Fermat's last theorem is the statement that if $n \geq 3$, then

$$x^n + y^z = z^n$$

has no positive integer solution. Also famously, this was finally proven by Andrew Wiles (and Richard Taylor) in 1994, more than 300 years after Fermat claimed a proof that he didn't write down. We will talk about the general proof after we introduced elliptic curves. This section will completely solve the cases $n = 2$. It will also reduce the cases $n = 3$ and $n = 4$ to questions about *elliptic curve*, which we have general methods for studying.

5.2.1. $n = 2$. We now study the equation

$$X^2 + Y^2 = Z^2$$

which are known as Pythagorean triples since they are side lengths to a right angled triangle. The solutions to this equation were already known to Euclid.

We begin with a simple reduction. If $Z = 0$, then the only solution is $(0, 0, 0)$. Otherwise, we can divide by Z^2 to get

$$x^2 + y^2 = 1$$

where $x = \frac{X}{Z}, y = \frac{Y}{Z} \in \mathbb{Q}$. Conversely, if we can find rational solutions to this equation, then multiplying through by the common denominator gives an integer solution to $X^2 + Y^2 = Z^2$.

This is a circle if x and y are allowed to be real numbers. Nowadays, it is known that the geometry of the equation is deeply connected with its rational solutions. In our case, the connection is very explicit. Let $\mathcal{O} = (1, 0)$, then there is a bijection

$$\{\text{Points on the circle}\} - \{\mathcal{O}\} \longleftrightarrow \{\text{Points on the } x\text{-axis}\}$$

given by connecting $\mathcal{O}$ with a point on the circle and taking intersection with the x-axis. This procedure is illustrated in Figure 5.4.1.

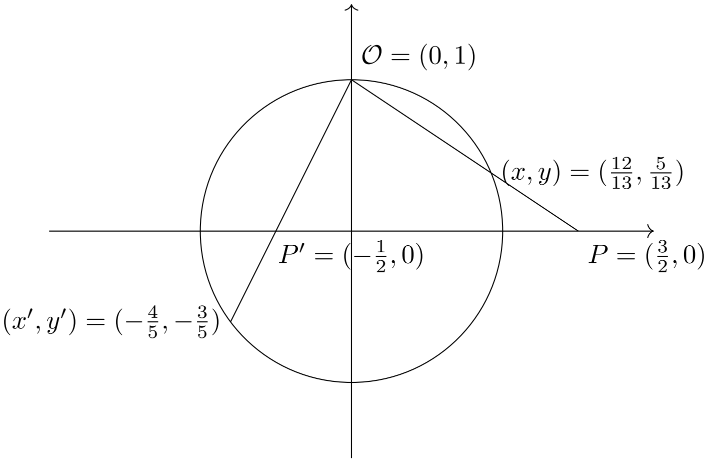

$\mathcal{O} = (0, 1)$
$(x, y) = (\frac{12}{13}, \frac{5}{13})$
$P' = (-\frac{1}{2}, 0)$
$P = (\frac{3}{2}, 0)$
$(x', y') = (-\frac{4}{5}, -\frac{3}{5})$

Figure 1. Construction of rational points

By writing down formulae for this in coordinates, we will see that this preserves rational points, so we get a bijection.

$$\{\text{Rational points on the circle}\} - \{\mathcal{O}\} \longleftrightarrow \{\text{Rational points on the } x\text{-axis}\}$$

Finally, the rational points on the x-axis are just given by the rational numbers, so we have a method to produce all rational points on the circle. Moreover, we can complete the bijection by saying $\mathcal{O}$ should be sent to "infinity", and "infinity" should be a rational point.

$$\{\text{Rational points on the circle}\} \longleftrightarrow \{\text{Rational points on the } \textit{projective line}\}$$

Here, we can think of a projective line as the usual line plus a point at infinity. In this form, the observation works for any quadratic equation with at least one rational solution, and it yields a complete parametrization of all rational solutions. We will work it out in detail in this case.

Theorem 5.3. *All rational solutions of the equation $x^2+y^2=1$ have the form*

$$\Big(\frac{2t}{1+t^2}, -\frac{1-t^2}{1+t^2}\Big)$$

for a unique $t \in \mathbb{Q} \cup \{\infty\}$.

Proof. The line connecting $\mathcal{O}=(0,1)$ and a point $(t,0)$ has the equation $y = 1-\frac{x}{t}$, so its intersection with the circle can be obtained by solving the equation

$$x^2 + \Big(1-\frac{x}{t}\Big)^2 = 1$$

By rearranging, this simplifies to

$$\frac{1+t^2}{t^2}x^2 - \frac{2}{t}x = 0$$

which has the solutions $x=0$ (as expected) and $x=\frac{2t}{1+t^2}$. It follows that

$$y = 1-\frac{x}{t} = -\frac{1-t^2}{1+t^2}$$

If $t \in \mathbb{Q}$, then both expressions are rational. If $t=\infty$, then by taking limit, we obtain $(x,y)=(0,1)$. Conversely, if (x,y) is a rational point on the circle, then $t=\frac{x}{1-y} \in \mathbb{Q}\cup\{\infty\}$ is its parameter. □

To get integer solutions to the classical Pythagorean triples problem, we just need to multiple the resulting fractions by their common denominators.

Corollary 5.4. *Let X,Y,Z be integers such that $X^2+Y^2=Z^2$, then there exists $m,n\in\mathbb{Z}$ and $d\in\mathbb{Q}$*

$$(X,Y,Z) = (2dmn, d(m^2-n^2), d(m^2+n^2))$$

Proof. Let (X,Y,Z) be an integer solution, and let t be the parameter attached to the pair $\big(\frac{X}{Z},\frac{Y}{Z}\big)$ as in the theorem. Writing $t=\frac{m}{n}$ for $m,n\in\mathbb{Z}$ shows that

$$\frac{X}{Z} = \frac{2mn}{m^2+n^2}, \quad \frac{Y}{Z} = \frac{m^2-n^2}{m^2+n^2}$$

Therefore, (X,Y,Z) is proportional to $(2mn, m^2-n^2, m^2+n^2)$. □

There is an unfortunate issue that even if $\gcd(m,n)=1$, the resulting triple $(2mn, m^2-n^2, m^2+n^2)$ may not be coprime. You will show on the homework that the GCD of this triple is at most 2, and it is equal to 2 exactly when m and n are both odd. In this case, we must allow d to be a half-integer to recover all integer solutions. Indeed, the parameters for $(3,4,5)$ is

$$d=\frac{1}{2},\ m=3,\ n=1$$

In classical formulations of this problem, the variables X and Y are treated as being interchangeable, so $(3,4,5)$ and $(4,3,5)$ are treated as the same solution. It happens that allowing this ambiguity gives a nicer uniqueness statement.

Corollary 5.5 (Euclid)**.** *If two integers squares add up to another integer squares, then there exists a unique pair* (m, n) *with* $\gcd(m, n) = 1$ *and* m, n *not both odd, as well as a unique multiplier* $d \in \mathbb{Z}$*, such that the three integers are*

$$(2dmn, d(m^2 - n^2), d(m^2 + n^2))$$

The key point of the proof is that if m and n are both odd, then using the pair $\left(\frac{m+n}{2}, \frac{m-n}{2}\right)$ gives the same solution with X and Y interchanged, and you can prove that this pair has GCD 1, and they are not both odd.

5.2.2. $n = 3$ *and* $n = 4$*.* The case $n = 4$ was already known to Fermat using the method of infinite descent. The case $n = 3$ was later resolved by Euler using the same method. From a modern point of view, both proofs are instances of a general method for finding rational points on an *elliptic curve*, which we will introduce later. For now, we simply re-arrange the equations in the appropriate forms.

We begin with $n = 3$. As before, it remains to show that $x^3 + y^3 = 1$ has no rational solutions except for the two obvious ones $\{(1, 0), (0, 1)\}$. Consider the change of variable

$$u = \frac{12}{x+y}, \quad v = 36\frac{x-y}{x+y}$$

then a brief calculation shows that

$$v^2 = u^3 - 432$$

Therefore, we have obtained a map

$$\{\text{Rational solutions of } x^3 + y^3 = 1\} \longrightarrow \{\text{Rational solutions of } v^2 = u^3 - 432\}$$

If $u \neq 0$, then we can invert the change of variable by

$$x = \frac{1}{2}\left(\frac{12}{u} + \frac{v}{3u}\right), \quad y = \frac{1}{2}\left(\frac{12}{u} - \frac{v}{3u}\right)$$

But since $u = 0$ does not lead to a rational solution of $v^2 = u^3 - 432$, this is a bijective correspondence. The trivial rational points become $\{(12, \pm 36)\}$, so we need to show there are no other solutions.

In the case $n = 4$, the curve is not itself an elliptic curve, but we can actually prove something stronger, namely

$$x^2 + y^4 = z^4$$

has no rational solutions with $xyz \neq 0$. In this case, consider the change of variable

$$u = \frac{z^2}{y^2}, \quad v = \frac{xz}{y^3}$$

Another brief calculation shows that

$$v^2 = u^3 - u$$

This new equation does have rational solutions: three obvious ones are $(\pm 1, 0)$ and $(0, 0)$. However, they all satisfy $v = 0$, which corresponds to a trivial solution with x or z equal to 0. Therefore, we need to show that $v^2 = u^3 - u$ have no other rational solutions.

Remark. Geometrically, we are doing a bijective coordinate substitution for $n = 3$, and for $n = 4$, we are actually considering a *covering map.* The optimal way to state them involves points at infinity.

As a basic example, consider the equations $x^2 + y^2 = 1$ and $uv = 1$. The change of variables

$$u = \frac{1-y}{x}, \quad v = \frac{1+y}{x}$$

gives a bijection between their solutions, except we are missing the two points $(x, y) \neq (0, \pm 1)$. The idea is that $uv = 1$ has two asymptotes, and those should give two points at infinity corresponding to the two missing solutions. After adding those points, the two solutions sets are actually in bijection. Geometrically, the two equations describe the same shape in different coordinates.

For the equation $x^3 + y^3 = 1$, there are three points at infinity. In projective coordinates, they are $[-\zeta : 1 : 0]$, where $\zeta \in \{1, \frac{-1\pm\sqrt{-3}}{2}\}$. We will not define this, but the main point is one of them is a rational point, and two of them involves $\sqrt{-3}$. On the equation $v^2 = u^3 - 432$, recall that we are missing two points $(\pm\sqrt{-432}, 0)$ from the bijection. This didn't matter before since they are not rational, but geometrically they correspond to the two non-rational points at infinity. Note that $\sqrt{-432} = 12\sqrt{-3}$, so the "field of definition" agrees. Finally, as we will see later, an elliptic curve has a point at infinity $[0 : 1 : 0]$. This corresponds to the rational point at infinity $[-1 : 1 : 0]$ on $x^3 + y^3 = 1$. By adding in the "extra" points, we obtain a bijective change of coordinates.

We now consider the equation $x^4 + y^4 = 1$ instead of $x^2 + y^4 = z^4$ to avoid a lot of additional complications. The transformations

$$u = \frac{z^2}{y^2}, \quad v = \frac{x^2 z}{y^3}$$

satisfy $v^2 = u^3 - u$, as before. In projective coordinate, this becomes

$$[u : v : w] = [y : x^2 : y^3]$$

The (x, y)-equation has four points at infinity $[\zeta_4 : i : 0]$, where $\zeta_4 \in \{\pm 1, \pm i\}$. They are all sent to the single point $(u, v) = (0, 0)$. The four (complex) points with $y = 0$ gets sent to the unique point at infinity on the elliptic curve $v^2 = u^3 - u$. This is the generic behaviour: each solution (u, v) comes from *four* solutions (x, y). There are two exceptions: $(u, v) = (\pm 1, 0)$. Each of them only comes from two (x, y)-solutions. These are the *ramified points.* Geometrically, this is analogous to a tangent line. In the end, we have a map from solutions of $x^4 + y^4 = 1$ to solutions of $v^2 = u^3 - u$ which is generically 4-to-1.

We will see later that the equations $v^2 = u^3 - 432$ and $v^2 = u^3 - u$ are both examples of elliptic curves, so the above discussions reduces to showing that certain elliptic curves have no rational points. It should be noted that these reductions are fundamentally different from the way elliptic curves appeared in Andrew Wiles' proof of Fermat's last theorem. The same strategy also works for $n = 7$, though the substitutions are much more complicated now.

5.3. **Congruent number problem.** We say a positive integer N is a congruent number if it is the area of a right-angled triangle with rational side lengths. Since $(3, 4, 5)$ is a right-angled triangle, its area 6 is a congruent number. The

Pythagorean triple $(9, 40, 41)$ has area $180 = 6^2 \times 5$, so 5 is a congruent number using the triangle $\left(\frac{3}{2}, \frac{20}{3}, \frac{41}{6}\right)$.

On the other hand, Fermat showed that 1 is not a congruent number using his method of infinite descent. This proof is written out in the textbook, but we will see that it is closely related to the $n = 4$ case of Fermat's last theorem. It seems difficult to determine if a given number is a congruent number. For example, it turns out that 157 is a congruent number, but the "simplest" triangle has the two right angled sides equal to

$$\frac{6803298487826435051217540}{411340519227716149383203}, \frac{411340519227716149383203}{21666555693714761309610}$$

The goal of this section is to reformulate the problem in terms of elliptic curve and explain the state of the art on this problem. The next section will introduce a tiny portion of the ideas used in the proof.

By the parametrization of Pythagorean triples, the area of a right angled triangle with rational side lengths is

$$\frac{1}{2}(2dt)(d(t^2 - 1)) = d^2(t^3 - t)$$

Therefore, we need to solve the equation

$$N = d^2(t^3 - t)$$

for $d, t \in \mathbb{Q}$. We further massage the equation by setting $t = \frac{x}{N}$ and $d = \frac{N^2}{y}$, then the equation simplifies to

$$y^2 = x^3 - N^2x$$

This equation has the trivial solutions $(0, 0)$ and $(\pm N, 0)$. They do not correspond to triangles. Our deduction shows that N is a congruent number exactly when this equation has additional solutions. In fact, using the addition law we will cover, if there is any additional solution, there are infinitely many solutions.

5.4. **Elliptic curves.** We now introduce elliptic curves and give some of their properties. In particular, we will sketch a (occasionally practical) algorithm that can be used to compute all of its rational points (in some sense). Using this modern language, we will prove that $y^2 = x^3 - x$ has no rational solutions other than the three obvious ones, which implies that 1 is not a congruent number and Fermat's last theorem for $n = 4$ holds.

For our purpose, an elliptic curve is the equation

$$y^2 = x^3 + ax + b$$

where $a, b \in \mathbb{Z}$, and $4a^3 + 27b^2 \neq 0$. This second condition is to ensure the cubic polynomial has no repeated roots. The two equations we have seen already $y^2 = x^3 - N^2x$ and $y^2 = x^3 - 432$ are both elliptic curves. Moreover, any cubic equation in two variables can be transformed into this form by a linear change of variables.

We will be mainly interested in its rational solutions. If E is an elliptic curve, we let $E(\mathbb{Q})$ denote the set of its rational points (plus a point at infinity, to be explained later). The key fact is that $E(\mathbb{Q})$ actually forms an abelian group.

5.4.1. *Group law.* We begin by introducing a strange operation.

Definition. Let $P, Q \in E(\mathbb{Q})$ be two rational points. If ℓ is the line passing through P and Q, then $E \cap \ell$ contains a third point $R = (x, y) \in E(\mathbb{Q})$. Define the *sum* of P and Q to be $P + Q = (x, -y)$.

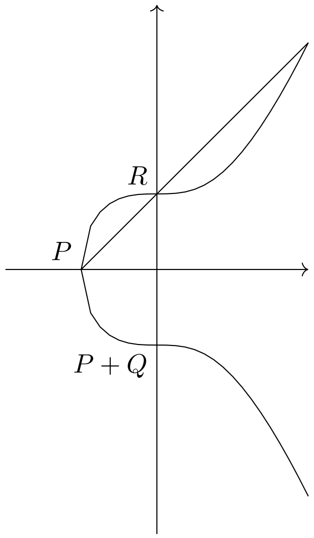


Figure 2. Group law

This definition of course runs into a few special cases. For example, if $P = Q$, then we need to take the tangent line to E at P. If two points have the same x-coordinates, then the line connecting them is the vertical line, which does not appear to intersect E at a third point. We introduce a "point of infinity" $\mathcal{O}$ which we think of as lying on every vertical line. Formally, the definition would suggest we set $P + \mathcal{O} = P$. Moreover, if $P = (x, y)$ and $Q = (x, -y)$, then $P + Q = \mathcal{O}$.

Theorem 5.6. *The set $E(\mathbb{Q})$ is forms a group under this operation. In particular, associativity holds*

$$P + (Q + R) = (P + Q) + R$$

for all $P, Q, R \in E(\mathbb{Q})$.

For once, associativity is the hardest to prove. It is possible to do it by writing down the formulae for adding two points, but the nice way involves more advanced knowledge of the geometry.

Example. For the curve $E : y^2 = x^3 - Nx$, we will compute the doubling formula. Given a point (a, b), the tangent line through it can be found by calculus to be

$$y = b + \frac{3a^2 - N}{2b}(x - a)$$

Substituting this into the equation for E gives a cubic equation in x

$$x^3 - Nx - \left(b + \frac{3a^2 - N}{2b}(x - a)\right)^2 = 0$$

The sum of the three roots of this equation is $\left(\frac{3a^2-N}{2b}\right)^2$. By construction, a is a solution with multiplicity 2, so the third solution must be

$$a' = \left(\frac{3a^2 - N}{2b}\right)^2 - 2a$$

Substituting this into the equation for the line gives the y-coordinate. After simplifying, the result is $2(a, b) = (a', b')$, with

$$a' = \frac{a^4 + 2Na^2 + N^2}{4b^2}, \quad b' = \frac{a^6 - 5Na^4 - 5N^2a^2 + N^3}{8b^3}$$

In particular, if $b = 0$, then $2(a, 0) = \mathcal{O}$.

Further specialize to $N = 25$, then we are looking at the congruent number problem for 5. The solution we have given corresponds to the point $P = (45, 300)$. Using the formula above, we get

$$2P = \left(\frac{1681}{144}, \frac{62279}{1728}\right)$$

This corresponds to the right angled triangle with the right angled side lengths equal to $\frac{4920}{1519}, \frac{1519}{492}$. In general, $n \cdot P$ for any positive integer n gives rise to a solution to the congruent number problem, and they are all distinct.

We can make two observations about the example.

(1) The formula for a' and b' are quotients of two polynomials of degree 4 (in the case of b', use $a^3 = b^2 + Na$), so given P, we expect four points Q (with complex coordinates) such that $2Q = P$.
(2) The numerators and denominators of $2P$ are much larger than those of P.

These are general facts about elliptic curves. Property (2) will form the basis of our proof by infinite descent.

5.4.2. *Example:* $y^2 = x^3 - x$. This section will prove that this equation has no rational point other than the three we have seen. Including the point at infinity $\mathcal{O}$, the theorem becomes

Theorem 5.7. *Let* $E : y^2 = x^3 - x$, *then* $E(\mathbb{Q}) = \{(0, 0), (\pm 1, 0), \mathcal{O}\}$.

Our exposition is taken from the book *Number Theory 1: Fermat's Dream* by Kato, Kurokawa, and Saito. This textbook is a wonderful introduction to many different aspects of modern number theory.

Here is the strategy: suppose P is another rational point not in our list, then we will show that $P = 2Q$ for some $Q \in E(\mathbb{Q})$. But the numerator and denominator of Q are smaller than those of P, so by the well-ordering principle for $\mathbb{N}$, we have a contradiction. This is a modern version of Fermat's proof by infinite descent. For convenience, we introduce a notion of *height*.

Definition. Let $x = \frac{a}{b}$ be a rational number in its lowest terms, so $\gcd(a, b) = 1$. Its *height* is defined to be $H(x) := \max(|a|, |b|)$.

Remark. It is more usual to define the height as $\log H(x)$. This has the property that $\log H(x)$ is approximately a quadratic form on $E(\mathbb{Q})$.

Suppose by contradiction that theorem does not hold, then there exists a solution $(x, y) \in E(\mathbb{Q})$ which minimizes $H(x)$. We may suppose $x > 1$. Indeed, if (x, y) is a solution, then so is

$$(x, y) + (0, 0) = \left(-\frac{1}{x}, \frac{y}{x^2}\right)$$

Moreover, x and $-\frac{1}{x}$ have the same height. Now write $x = \frac{r}{s}$, where r, s are positive integers and $\gcd(r, s) = 1$. They can't be both even. If they are both odd, then we can construct another solution by addition

$$(x, y) + (1, 0) = (x', y') := \left(\frac{x+1}{x-1}, -\frac{2y}{(x-1)^2}\right)$$

However,

$$x' = \frac{r+s}{r-s} = \frac{\frac{1}{2}(r+s)}{\frac{1}{2}(r-s)}$$

so $H(x') < H(x)$. This contradicts the choice of (x, y).

Now, multiplying both sides of the equation $y^2 = x^3 - x$ by s^4 gives

$$(s^2 y)^2 = rs(r+s)(r-s)$$

The right hand side is an integer, so the left hand side is also an integer. It follows by Theorem 2.20 that $s^2 y$ is an integer. After our reduction steps, the four terms on the right hand side are positive and pairwise coprime. Since their product is an integer square, each of them is individually a square. Therefore, we can write

$$x = u^2, x - 1 = v^2, x + 1 = w^2$$

where u, v, w are positive rational numbers.

For this elliptic curve, the doubling formula reads

$$2(a, b) = (a', b') := \left(\frac{(a^2+1)^2}{4b^2}, \frac{(a^2+1)(a^2-2a-1)(a^2+2a-1)}{8b^3}\right)$$

We need so solve $2(a, b) = (a', b')$. Observe that

$$a' = \left(\frac{a^2+1}{2b}\right)^2, \quad a' - 1 = \left(\frac{a^2-2a-1}{2b}\right)^2, \quad a' + 1 = \left(\frac{a^2+2a-1}{2b}\right)^2$$

It is therefore reasonable to set the terms in parentheses as u, v, w respectively and try to solve for a and b. After some experimentation, we get

$$a = (u+v)(u+w), \quad b = (u+v)(u+w)(v+w)$$

It is slightly messy but not difficult to check that $2(a, b) = (x, y)$.

In summary, starting with a non-trivial solution (x, y) with $y \neq 0$, we have found a new solution (a, b) with $b \neq 0$ such that

$$x = \frac{(a^2+1)^2}{4b^2} = \frac{(a^2+1)^2}{4(a^3-a)}$$

It remains to compute the height $H(a)$. If $a = \frac{p}{q}$ with $\gcd(p, q) = 1$, then

$$x = \frac{(p^2+q^2)^2}{4(p^3q - pq^3)}$$

Observe the following identity

$$-(12p^2 - 16q^2)(p^2+q^2)^2 + (3p^3 + 5pq^2) \cdot 4(p^3 - pq^2) = 16q^6$$

It follows that the GCD of the numerator and denominator of x is at most 16, so

$$H(x) \geq \frac{1}{16}\max((p^2+q^2)^2, 4(p^3q-pq^3)) > \frac{1}{16}H(a)^4$$

If $H(a) \leq 2$, then $a \in \{0, \pm 1, \pm 2, \pm\frac{1}{2}\}$. They either give $b = 0$ or does not give a rational value for b. All of these cases are excluded, so $H(a) \geq 3$. The above inequality implies $H(x) > H(a)$. This contradicts our initial choice of x, which proves the theorem.

Remark. Let $P = (x, y)$ be a point on the elliptic curve. Define $h(P) = \log H(x(P))$, then we have shown that

$$|h(2P) - 4h(P)| \leq C$$

for an explicit constant C. The Néron–Tate height $\hat{h}(P)$ is a modification such that $\hat{h}(P) = 4\hat{h}(P)$ holds exactly, and moreover $\left|h(P) - \hat{h}(P)\right|$ is bounded by an explicit constant independent of P.

5.4.3. *Comments on the general method.* For any elliptic curve E and positive integer n, a sophisticated version of this idea proves that there are finitely many $P_1, \cdots, P_k$ such that all rational points P can be written as a linear combination of them and n times another rational Q. This amounts to a description of the full set of rational points since in general $H(nQ) \gg H(Q)^{n^2}$, so we can effectively enumerate all points of $E(\mathbb{Q})$ up to a given height. Unfortunately, the proof only gives finiteness, but it does not identify the set $P_1, \cdots, P_k$.

To understand the problem, we look at the next simplest example, which is the congruent number curve with $N = p$ a prime

$$E : y^2 = x^3 - p^2x = x(x-p)(x+p)$$

Following through with our proof earlier, we get stuck in the part where $rs(r-p)(r+p)$ is a square implies each term is a square. The issue is that even assuming r, s are both odd, we may still have $\gcd(r-p, r+p) = p$. Another way to say it is that in general, x is a square times something in $\{\pm 1\} \times \{1, 2\} \times \{1, p\}$. By adding known points to x, we can remove the first two ambiguities, but then x is still either a square or a square times a prime. In the first case, the rest of the argument implies $P = 2Q$. In the second case, if such a point exists, then we can add it to the list of representatives. This is exactly what happens for $p = 5$: our point (45,300) need to be added to the list of representative. On the other hand, if $p = 3$, then there is an local obstruction at 3 preventing such a point to exist.

Deciding if such a point exist is equivalent to finding a single rational point on a related curve. In practice, if the new curve has a rational point, its height is usually much smaller than the corresponding one on E, so we can find it more easily by computer search. However, there are still cases where there are no local obstructions, but computer search does not find a rational point. This makes the algorithm inconclusive. This happens when $p = 17$. The corresponding curve was studied in Homework 11.3. In this case, quadratic reciprocity saves the day, but it might not always happen.

In other words, we can show that $E(\mathbb{Q})/nE(\mathbb{Q})$ is contained in a finite, theoretically computable set. This is the method of n-descent, which is usually only practical when $n \leq 5$. However, we don't know how to determine the image. The difference is controlled by a group which we call the *Tate–Shafarevich group* and denote by $\text{Ш}(E)$. The following conjecture is wide open and usually considered to be the hardest part of the Birch and Swinnerton-Dyer conjecture.

Conjecture. *The group* $\text{Ш}(E)$ *is finite.*

We can do n^k descent for increasing values of k and use it to either find more points or cut down the bounding set. Assuming the finiteness of Ш, the upper and lower bounds eventually agree. While we strongly believe in this conjecture, this still does not give an efficient algorithm since descent algorithms gets impractical for even moderate values of n.

5.4.4. *The Birch and Swinnerton-Dyer conjecture.* A consequence of our above method is that $E(\mathbb{Q})$ is *finitely generated*, so we can talk about its "dimension", which is actually called *rank*. For the curve $y^2 = x^3 - x$, this number is 0 since any point P satisfies $2P = \mathcal{O}$ and does not "essentially contribute". For the curve $y^2 = x^3 - 25x$, the point $P = (45, 300)$ satisfies all nP are distinct for $n \in \mathbb{Z}$, so this gives rise to a 1-dimensional space. It is possible to prove that there is no additional point, so this curve has rank 1. The curve $y^2 = x^3 - 34^2 x$ actually has rank 2, generated by the points $(-2, 48)$ and $(-16, 120)$. Using this terminology, an integer N is a congruent number if and only if the rank of $y^2 = x^3 - N^2 x$ is at least 1.

Heuristically, a curve with large rank has a lot of rational points, so if we reduce modulo p, it should have many points too. Let p be a prime, let N_p be the number of solutions to E modulo p. Recall that E has equation $y^2 = x^3 + ax + b$. Therefore,

$$N_p = 1 + p + \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + ax + b}{p} \right)$$

The first 1 comes from the point at infinity. If we expect the Legendre symbol to be "random", then the sum should be like a random walk and therefore have size $O(\sqrt{p})$. This is actually very precisely true.

Theorem 5.8 (Hasse)**.** *Let* $a_p = (p+1) - \#E(\mathbb{F}_p)$*, then* $|a_p| \leq 2\sqrt{p}$*.*

So to measure the size of N_p, it is natural to normalize $\frac{a_p}{2\sqrt{p}} \in [-1, 1]$. Surprisingly, the distribution of these numbers as p varies very minimally depend on E: unless E is very special, then the density of these numbers look like a semicircle of radius 1. This is the famous semicircle law or Sato–Tate conjecture, proven in 2007 by Clozel–Harris–Shepherd-Barron–Taylor using very sophisticated techniques which initially arose from Wiles' proof of Fermat's last theorem.

This does not depend on the rank. Instead, we consider a different measurement of size: how fast does

$$\prod_{p \leq X} \frac{N_p}{p}$$

grow as $X \to \infty$. By explicitly computing this for several curves, Birch and Swinnerton-Dyer made the following conjecture.

Conjecture (BSD conjecture, first version)**.** *Suppose* E *has rank* r*, then there exists a non-zero constant* C *such that*

$$\prod_{p \leq X} \frac{N_p}{p} \sim C (\log X)^r$$

It is later realized that this is a very difficult conjecture encompassing at least two different aspects. One is actually an analogue of the Riemann Hypothesis. The other one is the modern statement of the BSD conjecture, which has been vastly generalized in a completely different direction.

To state it, we need to define the L-function of an elliptic curve. It will be done as an Euler product

$$L(E,s) \doteq \prod_p \frac{1}{1-a_p p^{-s}+p^{1-2s}}$$

where the dot means for finitely many primes, the expression need to be modified. Using the theorem that $|a_p| \leq 2\sqrt{p}$, we can show that this infinite product converges if $s > \frac{3}{2}$. A consequence of Wiles' proof is that this definition can be extended to all complex $s \in \mathbb{C}$.

Heuristically, $\prod_{p \leq X} N_p$ should be related to $L(E,1)$, which is outside of the range where the series converges. Following up on this, Birch and Swinnerton-Dyer wrote down another version.

Conjecture (BSD conjecture, modern version)**.** *The power series expansion of $L(E,s)$ at $s=1$ starts with $C(s-1)^r$, where r is the rank, and C is a complex number with an explicit conjectural formula.*

The first version of the conjecture turns out to be a consequence of this conjecture and the Riemann hypothesis for $L(E,s)$, stating that the only zero of $L(E,s)$ with $\frac{1}{2} \leq \mathrm{Re}(s) \leq \frac{3}{2}$ lie on the line $\mathrm{Re}(s)=1$ (the shift is essentially a normalization issue). This modern version of the BSD conjecture has been generalized to higher dimensional equations, relating certain values of L-functions to the underlying geometry of the situation. The identity

$$L(1,\chi_5) = \sum_{n=0}^{\infty} \left(\frac{1}{5n+1} - \frac{1}{5n+2} - \frac{1}{5n+3} + \frac{1}{5n+4} \right) = \frac{2}{\sqrt{5}} \log \frac{1+\sqrt{5}}{2}$$

that we have seen before is a known instance of this conjecture.

In the case of the curve $y^2 = x^3 - N^2 x$, a remarkable fact first observed by Gauss is that a_p is related to the representation of p as a sum of two squares. In fact,

Theorem 5.9 (Gauss)**.** *If $p \equiv 3 \pmod 4$, then $a_p = 0$. If $p \equiv 1 \pmod 4$, then there is a unique representation $p = A^2 + B^2$, where A and B are positive integers and B is odd. In this case, $a_p = \pm 2B$.*

Shimura developed some formulae which in this case relates those numbers to representing an integer n as a sum of three squares. These kinds of results lead to the following theorem.

Theorem 5.10 (Tunnell)**.** *Let N be a square free integer. Let $a=1$ if N is odd and $a=2$ if N is even. For $\varepsilon \in \{0,1\}$, define*

$$S_\varepsilon = \left\{ (x,y,z) \in \mathbb{Z}^3 \;\middle|\; \frac{N}{a} = 2ax^2 + y^2 + 8z^2,\ z \equiv \varepsilon \pmod 2 \right\}$$

Assuming the BSD conjecture, the number N is a congruent number if and only if $\#S_0 = \#S_1$.

Observe that if $N \equiv 5, 6, 7 \pmod 8$, then S_0 and S_1 are both empty: there is a local obstruction modulo 8. It follows that, assuming the BSD conjecture, any square-free integer $N \equiv 5, 6, 7 \pmod 8$ is a congruent number. Heegner solved the problem if $N = p$ is a prime such that $p \equiv 7 \pmod 8$. This uses many classical identities of special functions in a very clever way. It has now been reinterpreted as constructing a special rational point on the elliptic curve, known as the *Heegner point*. This point exists for all $N \equiv 5, 6, 7 \pmod 8$. However, it is hard to show that this point is not one of the four trivial points, and such a statement is not true in general.

In the cases $N \equiv 1, 2, 3 \pmod 8$, we expect a generic N to not be a congruent number. The part of the BSD conjecture required is actually a theorem of Coates–Wiles. It was the first theoretical evidence for the BSD conjecture. More precisely,

Theorem 5.11 (Coates–Wiles)**.** *For the congruent number modular curve (more generally any CM curve), if $L(E, 1) \neq 0$, then $E(\mathbb{Q})$ is finite.*

As a consequence, if $\#S_0 \neq \#S_1$ in Tunnell's theorem, then it is provable N is not a congruent number. This is a computable criterion. However, in the exceptional case $\#S_0 = \#S_1$, we have no idea how to construct a solution in general. For each particular N, we can verify Tunnell's theorem by finding an explicit solution.

Thanks to progresses from the last 30 years, we *almost* completely understand the rank 0 and 1 situations (there are still some serious caveats). However, any case beyond that is a total mystery, probably requiring many new ideas.

5.4.5. **Modularity theorem**. We saw above that the number of solutions of $y^2 = x^3 - N^2x$ over a finite field is closely related to representing a number as a sum of two squares. There are further relations of this type.

Example. Let $E : y^2 + y = x^3 - x^2$, then $a_p(E)$ is the coefficient of q^p in the infinite product

$$f_E(q) = q \prod_{n\geq 1} (1 - q^n)^2 (1 - q^{11n})^2$$

This function is a *modular form*. They are a family of special functions whose study dates back at least to Gauss. The L-function of a modular form is roughly defined as

$$f = \sum_{n\geq 0} a_n q^n \quad \rightsquigarrow \quad L(f, s) \doteq \sum_{n\geq 1} \frac{a_n}{n^s}$$

Inspired by the relation in this and many other special cases, Taniyama–Shimura and Weil formulated the following question.

Conjecture (Modularity conjecture)**.** *The L-function of every elliptic curve comes from a modular form.*

There are refinements to this conjecture which predicts further properties of the modular form. This gives a finite list of modular forms to check. Frey observed that if $p > 3$ is a prime and $a^p + b^p + c^p = 0$ is a counterexample to Fermat's last theorem, then the elliptic curve

$$E_{a,b,c} : y^2 = x(x + a^p)(x - b^p)$$

should not come from a modular form. This was made precise by Ribet. Finally, Wiles proved enough of the modularity conjecture to deduce Fermat's last theorem.

Theorem 5.12 (Wiles, Taylor–Wiles, Breuil–Conrad–Diamond–Taylor)**.** *The modularity conjecture holds.*