NOTES ON ELEMENTARY NUMBER THEORY

TYLER GENAO

Contents

Date: Last updated May 2, 2026.

Preface

This is a copy of my notes for an elementary number theory course that I taught in Spring 2026 at the Ohio State University (MATH 4573), which I shared with my students during the semester. The prerequisite for this class is an introduction to proofs course (MATH 3345).

I primarily followed the classic textbook of Niven, Zuckerman and Montgomery [NZM91] content that I covered. However, I also included new sections on `SageMath` calculations (§1.5), an introduction to Diophantine geometry (§5.0), a re-imagining of plane curves and elliptic curves (§5.6, 5.7) and an additional section on extended topics in elliptic curves (§5.8, which is completely different from [NZM91, §5.8]). I have also included several new exercises and theorems, additional remarks on concepts and results from [NZM91], and extra details for several proofs from [NZM91]. Several new pictures of curves and surfaces have also been made for these notes; unless otherwise stated, such pictures were produced using Desmos.

In these notes, *Exercises* were assigned as homework for the class, and *Bonus Exercises* were optional homework problems to explore, often more open-ended than the usual exercises. I have also included other recommended exercises from [NZM91].

Errors in these notes are my own, and if you spot any, then please let me know. If you find these notes helpful, or have comments or suggestions, feel free to reach out to me at my current email address listed on tylergenao.com – I would love to hear it!

Acknowledgments

I would like to thank Sam Allen, Yaya Chen, Shivam Gupta, Sean Lipton, Bhavesh Mittal and Henry Tucek for comments that helped improve the accuracy and quality of these notes.

1. Divisibility

1.1. **Introduction.** In a broad sense, number theory is the subject which studies the *arithmetic* (additive and multiplicative properties) of the integers $0, \pm 1, \pm 2$, and so on. Conceptually, it is not so hard to add or multiply integers, but there are a large number of striking observations to make regarding the additive and multiplicative structures of these numbers. Number theory has deep connections to other areas of mathematics as well, such as algebra, geometry, analysis and topology. Several difficult problems in number theory also serve as centerpieces of classical and modern cryptography.

In this class, we learn *elementary number theory,* which colloquially means foundational number theory. We will learn about integer divisibility, modular arithmetic, basic abstract algebra, Quadratic Reciprocity and Diophantine equations. This will culminate in learning about elliptic curves, which have a rich historical context in studying Diophantine equations, as well as applications in modern cryptography. Along the way, we will also see how *computer algebra systems,* specifically `SageMath`, can be used to guide our questions in number theory.

To start these notes, let us go over a few classic results in number theory. Throughout these notes, we let $\mathbb{Z}$ denote the set of *integers* $(0, \pm 1, \pm 2, \ldots)$, $\mathbb{Z}^+$ the set of positive integers (i.e., natural numbers[1]) and $\mathbb{Q}$ the set of *rational numbers* (fractions of integers).

1. An integer n is divisible by 3 if and only if the sum of its digits is divisible by 3.
2. There are infinitely many *prime numbers*, which are natural numbers divisible only by 1 and themselves (Euclid).
3. The equation $x^2 + y^2 = z^2$ has infinitely many integer solutions $x, y, z \in \mathbb{Z}$ with no common factors (these are called *primitive Pythagorean triples*).
4. For integers $n \geq 3$, the equation $x^n + y^n = z^n$ has no positive solutions $x, y, z \in \mathbb{Z}$. This is *Fermat's Last Theorem,* stated in 1637 by Pierre de Fermat and proven in 1995 by Andrew Wiles.

One important aspect of number theory is *experimental number theory,* which is number theory informed or verified by computations. Here are some results of this flavor.

a) Every integer $n \in \mathbb{Z}^+$ is a sum of four squares, i.e. $n = a^2 + b^2 + c^2 + d^2$ for some $a, b, c, d \in \mathbb{Z}$. This is TRUE, and was proven with elementary techniques by Lagrange in 1770.

b) No n'th power is a sum of fewer than n n'th powers: i.e., if $x_1^n + x_2^n + \ldots + x_k^n = y^n$ for some $x_1, x_2, \ldots, x_k, y \in \mathbb{Z}$, then $k \geq n$. Euler conjectured this in the 18'th century, having seen some proofs for variants of Fermat's Last Theorem. However, this conjecture is FALSE: in 1987, Noam Elkies used the theory of elliptic curves to show with computers that

$$20615673^4 = 2682440^4 + 15365639^4 + 18796760^4.$$

This is also the smallest such counterexample, of which there exist infinitely many.

c) The *Goldbach Conjecture,* created in 1742 by Goldbach, states that every even integer $n \geq 4$ is the sum of two primes. For example $4 = 2 + 2$, $6 = 3 + 3$,

[1] If you are a computer scientist, then I am sorry.

$20 = 7+13$, $50 = 3+47$ and $100 = 29+71$. This conjecture is currently OPEN, but has been computationally verified for all positive integers $n < 4 \times 10^{18}$ (so up to *four quintillion integers*). However, no proof for this conjecture is known! It might be the case that a number larger than $4 \cdot 10^{18}$ disproves this conjecture, and we have not seen it yet.

Number theory is a *very* large subject. However, much research in modern number theory falls into at least one of the following subfields, which are broadly summarized.

1. *Algebraic number theory* studies generalizations of $\mathbb{Z}$ and $\mathbb{Q}$, whose elements, called *algebraic integers* or *algebraic numbers,* which mimic the arithmetic behavior of integers and rational numbers.
2. *Analytic number theory* studies estimates on proportions and distributions of positive integers, using analytic techniques.
3. *Arithmetic geometry* studies rational and integral solutions to algebraic curves (or more generally, *algebraic varieties*).

There are even more subfields of number theory, including (but not limited to) arithmetic combinatorics, arithmetic statistics, ergodic theory and probabilistic number theory.

To wrap this section up, let us recall two important principles learned in a foundations of mathematics class. These show up in many proofs in elementary number theory, and are in fact equivalent statements.

1. **The Principle of Well-Ordering** says that any nonempty subset $X \subseteq \mathbb{Z}^+$ has a *least positive element,* i.e., a smallest number in it.
2. **The Principle of Induction** says that if a collection of statements $\{P(n)\}$ about all integers $n \in \mathbb{Z}^+$ is true when $n = 1$, and $P(n)$ is true whenever $P(n-1)$ is true, then $P(n)$ is true for all $n \in \mathbb{Z}^+$.

1.2. **Divisibility.** Here are our main goals for this section.

- Review some basic divisibility properties of integers.
- State and prove the *Division Algorithm.*
- Define the *greatest common divisor* (GCD) of two integers.
- State and prove the applicable *Euclidean Algorithm.*
- Review the alternatively useful *Blankinship's Algorithm.*

Divisibility is an extremely important concept in elementary number theory: it provides a way to understand an integer via smaller, simpler components called *divisors,* or *factors.* This process is called *division* or *factorization* of an integer.

Definition 1.2.1. An integer a is **divisible** by nonzero $b \in \mathbb{Z}$ if there exists $x \in \mathbb{Z}$ with $bx = a$. We say that b divides a, call b a **factor,** or **divisor,** of a, and write $b \mid a$. If also $|b| < |a|$, we say that b is a **proper divisor** of a. When b does not divide a, we instead write $b \nmid a$.

Example 1.2.1. We check that:

- $1 \mid 12$.
- $-3 \mid 15$.
- $20 \mid 100$.
- $7 \nmid 5$.

Try to prove each case yourself! The first three examples require you to find the "complementary factor" for each divisor, and the fourth requires a one or two-line proof.

The following theorem is a collection of basic results on divisibility. This is from the textbook [NZM91], which my notes predominately follow. In general, if I am pulling a lemma/proposition/theorem from another source, then I will cite the source with its numbering for said result.

Theorem 1.2.1. [NZM91, Theorem 1.1] *The following statements are true for all integers a, b and d.*

(1) If $b \mid a$, then $b \mid ac$ for all $c \in \mathbb{Z}$.
(2) If $d \mid b$ and $b \mid a$, then $d \mid a$.
(3) If $d \mid a$ and $d \mid b$, then for all $x, y \in \mathbb{Z}$ one has $d \mid (ax + by)$.
(4) If $b \mid a$ and $a \mid b$, then $b = \pm a$.
(5) If $b \mid a$ and $a, b > 0$, then $b \leq a$.
(6) If $b \mid a$ and $c \in \mathbb{Z}$ is nonzero, then $bc \mid ac$.

Proof. We will prove item (3) together, which shows that "common divisors divide all $\mathbb{Z}$-linear combinations." Assume that $d \mid a$ and $d \mid b$; then we can write $dm = a$ and $dn = b$ for some $m, n \in \mathbb{Z}$. Thus, for any $x, y \in \mathbb{Z}$, we get

$$ax + by = (dm)x + (dn)y = d(mx + ny),$$

and so $d \mid (ax + by)$, which proves (3). I highly encourage you to try and prove the other parts! $\square$

The next theorem formalizes the conclusion of dividing one integer into another until a remainder is left. Classically, it is referred to as an algorithm, but it really is more of a theorem. Nonetheless, its proof is important, and it can be turned into a proper procedural algorithm.

Theorem 1.2.2 (The Division Algorithm)**.** [NZM91, Theorem 1.2] *Given integers a and b with $b > 0$, there exist unique $q, r \in \mathbb{Z}$ with $a = qb + r$ and $0 \leq r < b$.*

Proof. Consider the set

$$\begin{aligned} A :&= \{a + kb : k \in \mathbb{Z}\} \\ &= \{\ldots a - 3b, a - 2b, a - b, a, a + b, a + 2b, a + 3b, \ldots\}. \end{aligned}$$

As a consequence of the *Well-Ordering Principle,* we know that A has a smallest non-negative element, which we denote by r. By definition of A, we can express r as

$$r = a - qb$$

for some $q \in \mathbb{Z}$. This is the same as

$$a = qb + r.$$

We claim that $r < b$. Suppose this were not true; then $b \leq r$, i.e. $b \leq a - qb$, so that $0 \leq a - (q + 1)b$. However, from our initial assumption $0 < b$, we also have

$$qb < (q + 1)b,$$

i.e.,

$$-(q + 1)b < -qb,$$

i.e.,

$$a - (q + 1)b < r.$$

Since $a - (q + 1)b \geq 0$ and $a - (q + 1)b \in A$, this contradicts minimality of r. We conclude that $r < b$.

Finally, we need to show that q and r are unique. Suppose there exists another element $r_1 \in A$ with $r_1 < b$. We will show that $r_1 = r$, and then conclude that r is unique, from which it will quickly follow that q is unique, too. Let us write $r_1 = a - q_1 b$ for some $q_1 \in \mathbb{Z}$. Then we have

$$r_1 - r = b(q - q_1),$$

so that $b \mid (r_1 - r)$.

Assume for contradiction that $r_1 \neq r$. By minimality of r in A, we know that $r_1 - r \geq 0$, which implies that $r_1 - r > 0$. From $b > 0$, we then deduce by Theorem 1.2.1.(5) that $b \leq r_1 - r$. By assumption, we have $0 \leq r_1 < b$, so combining these two inequalities gives $r_1 < r_1 - r$, and thus $r < 0$, an impossibility. We conclude that $r_1 = r$, whence we also conclude from $a - qb = r = r_1 = a - q_1 b$ that $q_1 = q$. □

Remark 1.2.1. In the Division Algorithm, the assumption $b > 0$ is not necessary: one can prove a more general result where $0 \leq r < |b|$.

One utility of the Division Algorithm is in computing shared divisors between two integers. In fact, for any two integers (not both zero), there is a *greatest common divisor* that is a multiple of all other common divisors. This leads us to the following definition.

Definition 1.2.2. Given two integers a and b, the **greatest common divisor (GCD)** of a and b, written as $\gcd(a, b)$, is the largest $d \in \mathbb{Z}^+$ for which $d \mid a$ and $d \mid b$. Similarly $\gcd(a_1, a_2, \ldots, a_k)$ denotes the simultaneous greatest common divisor of $a_1, a_2, \ldots, a_k$.

Example 1.2.2. One can check that $\gcd(4, 8) = 4$, $\gcd(6, 15) = 3$ and $\gcd(7, 20) = 1$. Try this by listing all positive divisors of each integer here!

Remark 1.2.2. In the next section, we will talk about prime factorization of integers, which will give us a conceptually simple way to describe the GCD of two integers.

A useful alternative formulation of the GCD of two integers a and b is that it can be written as a $\mathbb{Z}$-linear combination of a and b.

Theorem 1.2.3. [NZM91, Theorem 1.3] *For any integers a and b, there exist $x, y \in \mathbb{Z}$ with $\gcd(a, b) = ax + by$.*

Proof. As we will see, the main idea of this proof is that "common divisors divide all $\mathbb{Z}$-linear combinations" that we saw in Theorem 1.2.1.(3). Consider the set of all $\mathbb{Z}$-linear combinations of a and b:

$$B := \{ax + by : x, y \in \mathbb{Z}\}.$$

By the Well-Ordering Principle, there exists a least positive integer $g \in B$, which we can write as $g = ax + by$ for some $x, y \in \mathbb{Z}$. We claim that g is the GCD of a and b.

First, we will show that g is a common divisor of a and b, i.e. $g \mid a$ and $g \mid b$; this will imply that $g \leq \gcd(a, b)$, by maximality of $\gcd(a, b)$ among common divisors of a and b. For contradiction, suppose that $g \nmid a$. Then by the Division Algorithm, we can write $a = qg + r$ for some $0 < r < g$. But we check that

$$\begin{aligned} r &= a - qg \\ &= a - q(ax + by) \\ &= a(1 - qx) + b(-qy). \end{aligned}$$

Thus, we have both $r \in B$ and $r < g$, contradicting the minimality of g. We deduce that $g \mid a$; a similar argument shows that $g \mid b$.

Since $g \mid a$ and $g \mid b$, maximality of the GCD implies that $g \leq \gcd(a, b)$. But since $\gcd(a, b)$ divides both a and b, by Theorem 1.2.1.(3) we have

$$\gcd(a, b) \mid (ax + by) = g$$

("common divisors divide all $\mathbb{Z}$-linear combinations"). We conclude that $\gcd(a, b) = g = ax + by$. □

We have just proven that the GCD has two useful interpretations:

- The greatest common divisor of a and b.
- The *smallest positive* $\mathbb{Z}$-linear combination of a and b.

(The "smallest positive" part follows from the fact that $\gcd(a, b)$ divides all other $\mathbb{Z}$-linear combinations of a and b. It also follows from our proof, where the GCD was the least positive element in our set B we applied the Well-Ordering Principle to.)

Using Theorem 1.2.3, one can also prove that the GCD of a and b is a multiple of any common divisor of a and b.

Proposition 1.2.4. *For integers a and b, if $d \mid a$ and $d \mid b$ then $d \mid \gcd(a, b)$.*

Proof. By Theorem 1.2.3, we know that $\gcd(a, b)$ is a $\mathbb{Z}$-linear combination of a and b. Thus $d \mid \gcd(a, b)$ by Theorem 1.2.1.(3). □

The GCD also satisfies additional properties.

Theorem 1.2.5. [NZM91, Theorems 1.6 - 1.9] *For any integers a, b, the following hold:*

(1) For $m \in \mathbb{Z}$, one has

$$\gcd(ma, mb) = m \cdot \gcd(a, b).$$

(2) If $d \mid a$, $d \mid b$ and $d > 0$, then

$$\gcd\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{1}{d} \cdot \gcd(a, b).$$

In particular, one has

$$\gcd\left(\frac{a}{\gcd(a, b)}, \frac{b}{\gcd(a, b)}\right) = 1.$$

(3) For $m \in \mathbb{Z}$, if $\gcd(a, m) = \gcd(b, m) = 1$ then

$$\gcd(ab, m) = 1.$$

(4) For $m \in \mathbb{Z}$, one has

$$\gcd(a, b) = \gcd(b, a) = \gcd(a, -b) = \gcd(a, b + am).$$

Proof. Here's a proof of item (4), specifically that $\gcd(a, b + am) = \gcd(a, b)$. Write $g := \gcd(a, b)$ and $h := \gcd(a, b + am)$. We want to show that $g = h$.

By Theorem 1.2.3 ("GCD is $\mathbb{Z}$-linear combo."), we can write $g = aw + bx$ and $h = ay + (b + am)z$ for some $w, x, y, z \in \mathbb{Z}$. On the one hand, rearranging terms gives

$$h = a(y + mz) + bz,$$

and so $g \mid h$, since $\gcd(a, b)$ divides all $\mathbb{Z}$-linear combinations of a and b (Theorem 1.2.1.(3)). On the other hand, rearranging also gives

$$\begin{aligned} g &= aw + bx \\ &= aw + bx + amx - amx && \text{(by the "adding zero" trick)} \\ &= a(w - mx) + (b + am)x. \end{aligned}$$

Thus $g := \gcd(a, b)$ is a $\mathbb{Z}$-linear combination of a and $b + am$, which implies that $h \mid g$ by Theorem 1.2.1.(3). We deduce that $g = \pm h$, and since $g, h > 0$, we conclude that $g = h$.

I highly recommend you prove the other parts of this theorem on your own! □

We will save our examples of computing GCD's once we have a proper algorithm. For now, we continue with some terminology.

Definition 1.2.3. Say two integers a and b are **coprime** if $\gcd(a, b) = 1$. Say that $a_1, a_2, \ldots, a_n \in \mathbb{Z}$ are **pairwise coprime** if $\gcd(a_i, a_j) = 1$ for any $1 \leq i \neq j \leq n$.

Theorem 1.2.6. [NZM91, Theorem 1.11] *If $d \mid ab$ and $\gcd(b, d) = 1$, then $d \mid a$.*

Proof. Since $\gcd(b,d)=1$, we know by Theorem 1.2.5.(1) that

$$\gcd(ab,ad) = a\gcd(b,d) = a.$$

Since $d \mid ab$ (assumption) and $d \mid ad$ (obvious), we conclude that

$$d \mid \gcd(ab,ad) = a. \qquad \square$$

Determining GCD's of integers can be a fun exercise – and computationally intensive, depending on which integers you look at! However, the Division Algorithm does not immediately give us a "step-by-step" process for doing this. Here is one such process.

Theorem 1.2.7 (The Euclidean Algorithm)**.** [NZM91, Theorem 1.11] *Given integers a and b with $b>0$, one can repeatedly apply the Division Algorithm to compute* $\gcd(a,b)$*:*

$$\begin{aligned}
a &= q_1 b + r_1, \ 0 < r_1 < b;\\
b &= q_2 r_1 + r_2, \ 0 < r_2 < r_1;\\
r_1 &= q_3 r_2 + r_3, \ 0 < r_3 < r_2;\\
&\dots\\
r_{j-2} &= q_j r_{j-1} + r_j, \ 0 < r_j < r_{j-1};\\
r_{j-1} &= q_{j+1} r_j.
\end{aligned}$$

Then $\gcd(a,b) = r_j$*. This algorithm shows a way to write* $\gcd(a,b)$ *(and each subremainder r_i) as a linear combination of a and b, using that $r_i = r_{i-2} - q_i r_{i-1}$.*

Let us explain the Euclidean Algorithm a bit more. To compute $\gcd(a,b)$, add b to itself until it is strictly larger than a, say $b(q+1) > a$. Then your first (sub)remainder is $r_1 := a - qb$. Repeat this process with b and r_1 instead of a and b, to get a second subremainder r_2. Repeat this process with r_1 and r_2, etc. Eventually, your $j+1$'st subremainder r_{j+1} will be zero, in which case the previous term r_j will be equal to $\gcd(a,b)$.

Proof. To see that $r_j = \gcd(a,b)$, we will make several calculations. Before this, we note that the process above does indeed terminate, i.e., we eventually have $r_{j+1}=0$, since each remainder r_j is strictly smaller than the previous r_{j-1} in the Division Algorithm.

$$\begin{aligned}
\gcd(a,b) &= \gcd(q_1 b + r_1, b) && \text{(by the first equation above)}\\
&= \gcd(q_1 b + r_1 - q_1 b, b) && \text{(by Theorem 1.2.5.(4))}\\
&= \gcd(r_1, b)\\
&= \gcd(r_1, q_2 r_1 + r_2) && \text{(by the second equation above)}\\
&= \gcd(r_1, r_2)\\
&\dots\\
&= \gcd(r_{j-1}, r_j)\\
&= \gcd(q_{j+1} r_j, r_j)\\
&= r_j \cdot \gcd(q_{j+1}, 1) && \text{(by Theorem 1.2.5.(1))}\\
&= r_j.
\end{aligned}$$

We conclude that $r_j = \gcd(a, b)$, i.e., the GCD of a and b is the "final nonzero remainder" of the repeated Division Algorithm.

Next, we will show that each subremainder r_i in the Euclidean Algorithm (including $r_j = \gcd(a, b)$) is a $\mathbb{Z}$-linear combination of a and b. We proceed by *induction* on the index i: for the base case $i = 1$, clearly $r_1 = a - q_1 b$ is a $\mathbb{Z}$-linear combination of a and b. Suppose then that for an index $i < j$, we can write r_k as a linear combination of a and b for all $k < i$. We want to show that r_i is a linear combination of a and b; we know that

$$r_i = r_{i-2} - q_i r_{i-1},$$

and since both r_{i-2} and r_{i-1} are linear combinations of a and b by our *inductive hypothesis*, so is r_i. □

We can now compute GCD's with ease.

Example 1.2.3. Let us compute the GCD of 42823 and 6409, and express it as a $\mathbb{Z}$-linear combination of the two. We take a step-by-step process via the Euclidean Algorithm:

$$\begin{aligned} 42823 &= 6 \cdot 6409 + 4369; \\ 6409 &= 1 \cdot 4369 + 2040; \\ 4369 &= 2 \cdot 2040 + 289; \\ 2040 &= 7 \cdot 289 + 17; \\ 289 &= 17 \cdot 17. \end{aligned}$$

We conclude by the Euclidean Algorithm that $\gcd(42823, 6409) = 17$.

Continuing this example, suppose we want to express the GCD – or more generally, the subremainders – as a $\mathbb{Z}$-linear combination of $a := 42823$ and $b := 6409$. The Euclidean Algorithm shows us how to do this: each step has r_i ($i \geq 2$) as a linear combination of both r_{i-1} and r_{i-2} from previous steps, so if we write the first step in terms of a and b, then we can write the remaining steps in terms of a and b inductively.

In our calculations, we have $r_1 = 4369, r_2 = 2040, r_3 = 289$ and $r_4 = 17 = \gcd(a, b)$. Let us rewrite our Euclidean Algorithm steps with these variables:

$$\begin{aligned} a &= 6b + r_1; \\ b &= r_1 + r_2; \\ r_1 &= 2r_2 + r_3; \\ r_2 &= 7r_3 + r_4; \\ r_3 &= 17r_4. \end{aligned}$$

Going inductively from top to bottom, we see how to express each subremainder as a $\mathbb{Z}$-linear combination of a and b:

$$\begin{aligned}
r_1 &= a - 6b; \\
r_2 &= b - r_1 = b - (a - 6b) = -a + 7b; \\
r_3 &= r_1 - 2r_2 = (a - 6b) - 2(-a + 7b) = 3a - 20b; \\
r_4 &= r_2 - 7r_3 = (-a + 7b) - 7(3a - 20b) = -22a + 147b \ (= \gcd(a, b)).
\end{aligned}$$

We thus conclude the following:

$$\begin{aligned}
4369 &= 42823 - 6 \cdot 6409; \\
2040 &= -42823 + 7 \cdot 6409; \\
289 &= 3 \cdot 42823 - 20 \cdot 6409; \\
17 &= -22 \cdot 42823 + 147 \cdot 6409 = \gcd(42823, 6409).
\end{aligned}$$

You can double-check these with a calculator, such as Desmos.

Next, we will give another way to use the Euclidean Algorithm to compute the GCD of a and b (and the subremainders) as a $\mathbb{Z}$-linear combination of a and b. This method is called **Blankinship's Algorithm,** and it computes $\gcd(a, b)$ via computing each r_i one step at a time, using elementary row operations. In comparison to the Euclidean Algorithm, the upshot of Blankinship's Algorithm is that it will calculate the subremainders *and* their expression as a linear combination of a and b *at the same time,* without having to substitute out into previous terms.

Here is Blankinship's Algorithm:

1. Given integers a and b, start with a 2×3 matrix
$$\left[\begin{array}{c|cc} a & 1 & 0 \\ b & 0 & 1 \end{array}\right].$$
The first row is Equation A, and the second is Equation B: these are $a = 1 \cdot a + 0 \cdot b$ and $b = 0 \cdot a + 1 \cdot b$, respectively.
2. Since $r_1 = a - q_1 b$, subtract $q_1 \cdot B$ from A to get
$$\left[\begin{array}{c|cc} a & 1 & 0 \\ b & 0 & 1 \end{array}\right] \xrightarrow{A \mapsto A - q_1 B} \left[\begin{array}{c|cc} a - bq_1 & 1 & -q_1 \\ b & 0 & 1 \end{array}\right] = \left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ b & 0 & 1 \end{array}\right].$$
3. Since $r_2 = b - q_2 r_1$, subtract $q_2 \cdot A$ from B to get
$$\left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ b & 0 & 1 \end{array}\right] \xrightarrow{B \mapsto B - q_2 \cdot A} \left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ b - q_2 r_1 & -q_2 & q_1 q_2 \end{array}\right] = \left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ r_2 & -q_2 & q_1 q_2 \end{array}\right].$$
4. Repeat this until you obtain 0 on the left column somewhere, i.e., until you get to $r_{j+1} = 0$. then the remaining nonzero number in the first column is $r_j = \gcd(a, b)$, and its row expresses it as a $\mathbb{Z}$-linear combination of a and b. The previous matrices also show how to express each r_i as a linear combination of a and b.

Example 1.2.4. We will write the GCD of 267 and 112 as a $\mathbb{Z}$-linear combination of the two, using Blankinship's Algorithm:

$$\begin{aligned}
\left[\begin{array}{c|cc} 267 & 1 & 0 \\ 112 & 0 & 1 \end{array}\right] &\xrightarrow{A\mapsto A-2\cdot B} \left[\begin{array}{c|cc} 43 & 1 & -2 \\ 112 & 0 & 1 \end{array}\right] \\
&\xrightarrow{B\mapsto B-2A} \left[\begin{array}{c|cc} 43 & 1 & -2 \\ 26 & -2 & 5 \end{array}\right] \\
&\xrightarrow{A\mapsto A-B} \left[\begin{array}{c|cc} 17 & 3 & -7 \\ 26 & -2 & 5 \end{array}\right] \\
&\xrightarrow{B\mapsto B-A} \left[\begin{array}{c|cc} 17 & 3 & -7 \\ 9 & -5 & 12 \end{array}\right] \\
&\xrightarrow{A\mapsto A-B} \left[\begin{array}{c|cc} 8 & 8 & -19 \\ 9 & -5 & 12 \end{array}\right] \\
&\xrightarrow{B\mapsto B-A} \left[\begin{array}{c|cc} 8 & 8 & -19 \\ 1 & -13 & 31 \end{array}\right]
\end{aligned}$$

We conclude that $\gcd(267, 112) = 1 = -13 \cdot 267 + 31 \cdot 12$.

To wrap this section up, we will define the *least common multiple* of two integers, which is an analog of the GCD.

Definition 1.2.4. Given integers a and b, the **least common multiple (LCM)** of a and b, written as $\operatorname{lcm}(a, b)$, is the smallest positive *common multiple* c of a and b (so $a \mid c$ and $b \mid c$). Similarly $\operatorname{lcm}(a_1, a_2, \ldots, a_k)$ denotes the simultaneous least common multiple of $a_1, a_2, \ldots, a_k$.

The following theorem shows that the LCM is not only smaller than any common multiple, but divides them (just as the GCD is a multiple of any common divisor).

Theorem 1.2.8. [NZM91, Theorem 1.12] *For integers a and b, any common multiple c satisfies* $\operatorname{lcm}(a, b) \mid c$.

Proof. Suppose that $a \mid c$ and $b \mid c$. By the Division Algorithm, we can write

$$c = q \cdot \operatorname{lcm}(a, b) + r$$

for some $0 \leq r < \operatorname{lcm}(a, b)$. Thus $r = c - q \cdot \operatorname{lcm}(a, b)$, and since a and b divide both c and $\operatorname{lcm}(a, b)$, it follows that $a \mid r$ and $b \mid r$. Thus r is a common multiple of a and b; by $r < \operatorname{lcm}(a, b)$ and minimality of $\operatorname{lcm}(a, b)$, this forces $r = 0$, so that $\operatorname{lcm}(a, b) \mid c$. □

Exercises. From [NZM91, §1.2], pages 17–18: #6, 9, 10, 12, 13, 14, 15, 17, 20, 21, 23.

Exercise 1.2.1. We will prove the following result using **induction:**

Proposition. *For any integer $n > 0$, one has $5 \mid (9^n - 4^n)$.*

a) First, convince yourself that this result might be true: for each integer $1 \leq n \leq 4$, compute $9^n - 4^n$ and write it as a multiple of 5.

We will break this proof into steps:

b) *Check the base case:* write down your answer for $n = 1$, and confirm that it is a multiple of 5.
c) *Induction hypothesis:* assume that the proposition is true for all integers k with $1 \leq k < n$. Use this to prove the proposition for $k = n$. (*Hint:* observe that $9^n - 4^n = (5 + 4) \cdot 9^{n-1} - 4 \cdot 4^{n-1}$.)

Exercise 1.2.2. Prove the following result via **contradiction:**

Proposition. *For any integer $n \geq 0$, one has $4 \nmid (n^2 + 2)$.*

Exercise 1.2.3. We will prove the following result by proving its equivalent **contrapositive:**

Proposition. *Let $a > 1$ be an integer. Then $2^a + 1$ is not divisible by $2^a - 1$.*

a) First, state the contrapositive of the proposition: i.e. $\neg q \Rightarrow \neg p$.
b) Prove the contrapositive statement.

Since the proposition and its contrapositive are equivalent, we have thus proven the proposition.

c) Based on your work above, make a conjecture about the GCD of $2^a - 1$ and $2^a + 1$. Try and prove it if you can!

Exercise 1.2.4. This exercise will get you acquainted with GCD calculations by hand.

a) Use the Euclidean Algorithm to compute the GCD of $a = 2026$ and $b = 365$.
b) Next, compute the GCD of $a = 2026$ and $b = 365$ using Blankinship's Algorithm. Using this work, also write your GCD as a $\mathbb{Z}$-linear combination of a and b.
c) Use either the Euclidean Algorithm or Blankinship's Algorithm to compute the GCD of $a = 1995$ and $b = 163$, and to write this GCD as a $\mathbb{Z}$-linear combination of a and b.

Exercise 1.2.5. Show that if positive integers a and b satisfy $\gcd(a, b) = \operatorname{lcm}(a, b)$, then $a = b$.

Exercise 1.2.6. Show that for integers a, b and c, one has $a \mid bc$ if and only if $\frac{a}{\gcd(a,b)} \mid c$.

Bonus Exercise 1.2.7. Suppose that a and b are integers with $\gcd(a, b) = 1$. Show that if $a \mid n$ and $b \mid n$, then $ab \mid n$.

Bonus Exercise 1.2.8. If you have some experience programming, create a script which computes the GCD of any two integers and expresses it as a $\mathbb{Z}$-linear combination of the two.

1.3. **Primes.** In this section, we will learn about *prime numbers,* which are the "building blocks" of the integers. Here are our main goals for this section:

- Define prime numbers.
- Prove the *Fundamental Theorem of Arithmetic,* and discuss unique factorization of integers.
- Review Euclid's proof of the infinitude of primes.
- Describe the GCD and LCM in terms of prime factorizations.

Definition 1.3.1. An integer $p > 1$ is a **prime number** (or **prime**) if it has no proper divisors, i.e., no $1 < d < p$ with $d \mid p$. An integer which is not prime is called a **composite number** (or **composite**). Every composite number n can be written as $n = ab$, where $1 < a, b < n$.

Example 1.3.1. Here are some examples of prime numbers:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 39, 41, 43, 47, 53, \ldots.$$

If you continue coming up with prime numbers, your list could theoretically never stop – there are an *infinite* amount of primes. We will prove this later in the section. What is also interesting is that there are no (nontrivial) known functions which generate all of the prime numbers.

Composites are also abundant in the integers. For example $4 = 2 \cdot 2$ is composite, and so are $6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, \ldots$.

As mentioned at the start of this section, primes are the building blocks of the integers. Here is one way to quantify this.

Theorem 1.3.1. [NZM91, Theorem 1.14] *Every integer $n > 1$ is a product of primes.*

Proof. This is an "inducting down" argument. If n is already prime, then we are done. Otherwise n is composite, and we can write $n = ab$ for some $1 < a, b < n$. If a and b are prime, then n is a product of two primes, and we are done. If n is not prime, and for example a is composite, then a is a product of integers strictly less than a, and so on. This process terminates since we can write any positive composite number as a product of two strictly smaller integers. □

By the above theorem, each integer $n > 1$ has a factorization into primes. Grouping equal primes together gives us a *prime-power factorization:*

$$n = \prod_{i=1}^{r} p_i^{e_i},$$

where each p_i is a prime distinct from the other p_j's, and each $e_i > 0$. As we will show, this factorization is *unique* for n, up to reordering the prime-power factors.

Theorem 1.3.2 (The Fundamental Theorem of Arithmetic (FTA))**.** *The factorization of an integer $n > 1$ is unique up to reordering its prime-power factors.*

Before we prove the FTA, we need one crucial lemma – it is the defining property of a prime number (and more generally, the defining property of an *irreducible* element in a *ring*).

Lemma 1.3.3. [NZM91, Theorem 1.15] *If p is prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. More generally, if $p \mid a_1 a_2 \cdots a_k$, then $p \mid a_i$ for some $1 \leq i \leq k$.*

Proof. Assume that $p \mid ab$. If $p \nmid a$, then $\gcd(p, a) = 1$, and so we have $p \mid b$ by Theorem 1.2.6 (which said that $d \mid ab, \gcd(d, a) = 1 \Rightarrow d \mid b$) ([NZM91, Theorem 1.11]).

The second result follows from induction on k: if $p \mid a_1 \cdot a_2 a_3 \cdots a_k$ and $p \nmid a_1$, then by the previous part we have $p \mid a_2 a_3 \cdots a_k$, which is a product of less than k terms. □

Proof of the FTA. Suppose we have two factorizations of n into products of primes:

$$\prod_{i=1}^{r} p_i^{e_i} = n = \prod_{j=1}^{s} q_j^{f_j}, \tag{1}$$

where the p_i's are distinct from one another, and the q_j's are distinct from one another.

For each $1 \leq i \leq r$, we have by Equation (1) that

$$p_i \mid q_1^{f_1} q_2^{f_2} \cdots q_s^{f_s},$$

so by Lemma 1.3.3 we have $p_i \mid q_j^{f_j}$ for some $1 \leq j \leq s$. Reinterpreting this as $p_i \mid q_j^{f_j} = q_j q_j \cdots q_j$, we reapply Lemma 1.3.3 and deduce that $p_i \mid q_j$, and thus $p_i = q_j$ since q_j is prime. This applies to *every* prime p_i, whence we deduce that $r \leq s$. Applying the same argument to the q_j's shows that $s \leq r$, so we conclude that $r = s$. Furthermore, since $p_i \neq p_j$ when $i \neq j$, each p_i is equal to a *unique* q_j. We use this observation to replace the q's with p's in Equation (1), relabeling the primes so that each $p_i = q_i$:

$$\prod_{i=1}^{r} p_i^{e_i} = \prod_{i=1}^{r} p_i^{f_i}. \tag{2}$$

Next, we show that the exponents in Equation (2) match up, i.e., that each $e_i = f_i$. If $e_i \neq f_i$ for some $1 \leq i \leq r$, say $e_i < f_i$, *without loss of generality.*[2] Then canceling $p_i^{e_i}$ from both sides shows that

$$p_i \mid p_1^{e_1} p_2^{e_2} \cdots \widehat{p_i^{e_i}} \cdots p_r^{e_r},$$

where $\widehat{p_i^{e_i}}$ means we exclude $p_i^{e_i}$ from the product. By Lemma 1.3.3, this implies that $p_i \mid p_j$ for some $j \neq i$, which is impossible. We conclude that the two prime factorizations are equivalent up to reindexing. □

Example 1.3.2. Let us try to factorize a few integers into prime powers:

- $10 = 2 \cdot 5$.
- $80 = 2^4 \cdot 5$.
- $-40 = -1 \cdot 2^2 \cdot 5$. (The FTA extends to negative integers, too.)
- $31 = 31$.

[2] In general, the phrase "without loss of generality" refers to the fact that the alternative case is essentially the same work. In this example, the other case to consider is $e_i > f_i$, which is essentially the same proof, except we consider in Equation (2) product of prime powers on the right-hand side instead of the left.

Abstract Algebra Digression 1.3.3. The ring of integers $\mathbb{Z}$ is part of a special class of rings: it is a *unique factorization domain* (UFD), wherein every nonzero, non-unit element has a unique factorization into prime/irreducible elements, up to units. However, *algebraic number rings* from algebraic number theory, which are meant to generalize $\mathbb{Z}$ in $\mathbb{Q}$ to finite degree field extensions beyond $\mathbb{Q}$, are not necessarily UFD's. For example, the ring

$$\mathbb{Z}[\sqrt{-5}] := \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$$

is not a UFD, as there are *two* distinct (so non-associate) factorizations of 6 into irreducible elements:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}),$$

where $2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}$ are irreducible and do not differ from one another by a unit. Nonetheless, any algebraic number ring, including $\mathbb{Z}[\sqrt{-5}]$, has a unique factorization theorem for its *ideals;* this is the unique factorization property that is inherited from $\mathbb{Z}$. Try and determine what the prime-power ideal factorization of $(6) := 6\mathbb{Z}[\sqrt{-6}]$ is!

We have shown that every integer has a unique factorization into prime powers. Conversely, if we start with a finite set of prime numbers, we can construct new integers by taking products of various powers of them. A natural question is then: how many prime numbers are there?

Theorem 1.3.4 (Euclid)**.** *The number of primes is infinite.*

Proof. This proof will show that any finite set of prime numbers "generates" a new prime not in that set. Fix primes $p_1, p_2, \ldots, p_r$. Consider the integer

$$n := p_1 p_2 \ldots p_r + 1.$$

Observe that n is not divisible by any prime p_i for $1 \leq i \leq r$: otherwise, we would have

$$p_i \cdot \left(\frac{n}{p_i} - p_1 p_2 \cdots \widehat{p_i} \cdots p_r \right) = 1,$$

and thus $p_i \mid 1$, which is absurd. However, since $n > 1$, we know that n must have a prime factor (e.g. by the "inducting down" argument from the proof of Theorem 1.3.1). Therefore, there exists a prime $p \mid n$ that is different from $p_1, p_2, \ldots, p_r$. □

Remark 1.3.1. Euclid's proof above does not necessarily generate every prime number. In fact, it is currently unknown whether there exists a (nontrivial) function $f \colon \mathbb{Z}^+ \to \mathbb{Z}^+$ for which $f(n)$ is the n'th prime number. However, there are several *polynomials* which are known to generate prime numbers. For more on *prime-generating polynomials,* see here.

Definition 1.3.2. Given an integer $n > 1$ with prime factorization

$$n = \prod_{i=1}^{r} p_i^{e_i},$$

we often call each e_i the **exponent of p_i dividing n,** or the **p_i-adic valuation of n,** denoted $v_{p_i}(n) := e_i$.

The p-adic valuation notation above can be useful, as we will sometimes write a factorization of an integer n as an "infinite" product:

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

However, this product is in fact finite, since for any prime $q \nmid n$ one has $v_q(n) = 0$. Finally, we write $p^e \parallel n$ to say that p^e is the largest power of p which divides n. Thus, we always have $p^{v_p(n)} \parallel n$.

Unique factorization of integers makes GCD's and LCM's easy to compute in principle:

$$\gcd(a,b) = \gcd\left(\prod_p p^{v_p(a)}, \prod_p p^{v_p(b)}\right) = \prod_p p^{\min\{v_p(a), v_p(b)\}}.$$

Similarly,

$$\operatorname{lcm}(a,b) = \prod_p p^{\max\{v_p(a), v_p(b)\}}.$$

Example 1.3.4. Let $a = 108$ and $b = 225$. Then $a = 2^2 \cdot 3^3$ and $b = 3^2 \cdot 5^2$, and thus

$$\gcd(a,b) = 2^0 \cdot 3^2 \cdot 5^0 = 9$$

and

$$\operatorname{lcm}(a,b) = 2^2 \cdot 3^3 \cdot 5^2 = 2700.$$

Exercises. From [NZM91, §1.3], pages 28–32: #1, 6, 7, 9, 12, 13, 15, 22, 32.

Exercise 1.3.1. Determine whether the following statements are true or false. If a statement is true, then prove it; if it is false, then provide a counterexample.

a) For a prime p, if $p \mid a^2$ then $p \mid a$.
b) If $\gcd(a,b) = \gcd(a,c)$ then $\operatorname{lcm}(a,b) = \operatorname{lcm}(a,c)$.
c) If $n \mid a^2$ then $n \mid a$.
d) If $n \mid a^2 - 1$ then $n \mid a^4 - 1$.

Exercise 1.3.2. Say that an integer a is a *perfect square* if it is the square of an integer, i.e. $a = n^2$ for some $n \in \mathbb{Z}$. Prove that if $x, y \in \mathbb{Z}^+$ are odd, then $x^2 + y^2$ is not a perfect square.

Exercise 1.3.3. Prove that if an integer n is odd, then $n^2 - 1$ is divisible by 8. Show that if also $3 \nmid n$, then we have the stronger divisibility $24 \mid n^2 - 1$.

Exercise 1.3.4.

a) Show that if an integer $n > 0$ is a composite number, then it must have a prime divisor $p \in \mathbb{Z}^+$ which satisfies $p \leq \sqrt{n}$.
b) Use part a) to check by hand whether 283 is a prime number. (You may use a calculator to approximate $\sqrt{283}$.)

Exercise 1.3.5. This problem explores a special case of *Dirichlet's Theorem on Primes in Arithmetic Progressions.*

Theorem (Dirichlet's Theorem on Primes in Arithmetic Progressions)**.** *Given positive coprime integers a and b, there exist infinitely many primes of the form $a+bk$.*

This exercise focuses on a proof for primes of the form $3+4k$ ($a=3, b=4$).

a) Show that a positive integer n of the form $3+4k$ has at least one prime factor of the same form.

b) Mimicking Euclid's proof of the infinitude of primes (see Theorem 1.3.4 ([NZM91, Theorem 1.17])), use part a) to prove the following:

Theorem. *There are infinitely many primes of the form $3+4k$.*

(*Hint:* in mimicking this proof, construct an integer n of the form $3+4k$.)

Bonus Exercise 1.3.6. Prove the following variant of Exercise 1.3.5:

Theorem. *There are infinitely many primes of the form $1+4k$.*

(*Hint:* Use the following special case of [NZM91, Theorem 2.12]:

Corollary. *For any integer n and any odd prime p, if $p \mid (n^2+1)$ then p is of the form $1+4k$.*

Then mimic Euclid's proof, constructing an integer n of the form $1+(2k)^2$.)

Bonus Exercise 1.3.7. This exercise will give a topological proof that there are infinitely many primes, due to H. Furstenberg.

Let us define a topology on $\mathbb{Z}$ as follows. Say that a subset $U \subseteq \mathbb{Z}$ is open if and only if it is a union of *arithmetic progressions,* i.e., sets of the form

$$S(a,b) := \{a+bn : n \in \mathbb{Z}\}.$$

a) Show that such a definition for open sets satisfies the axioms for a topology on $\mathbb{Z}$.

b) Show that for $a, b \in \mathbb{Z}$ one has

$$S(a,b) = \mathbb{Z} \smallsetminus \bigcup_{r=1}^{b-1} S(a+r,b).$$

Deduce that $S(a,b)$ is closed.

c) Show that

$$\mathbb{Z} \smallsetminus \{1,-1\} = \bigcup_{\text{prime } p} S(0,p).$$

Argue that $\mathbb{Z} \smallsetminus \{1,-1\}$ cannot be closed. Then using part b), conclude that there are infinitely many primes.

Bonus Exercise 1.3.8. The Fundamental Theorem of Arithmetic is a statement about the elements of $\mathbb{Z}$ different from $0, \pm 1$. In comparison, not all commutative rings admit an analogous unique factorization theorem for their elements. This exercise explores a particular example of an *algebraic number ring* which fails to have unique factorization.

Consider the ring

$$R := \mathbb{Z}[\sqrt{-6}] := \{a+b\sqrt{-6} : a, b \in \mathbb{Z}\}.$$

Define a *norm map* $N\colon \mathbb{Z}[\sqrt{-6}] \to \mathbb{Z}_{\geq 0}$ via

$$N(a+b\sqrt{-6}) := (a+\sqrt{-6})(a-\sqrt{-6}) = a^2+6b^2.$$

a) Show that for $\alpha, \beta \in R$ we have $N(\alpha\beta) = N(\alpha)N(\beta)$.

Recall that an element $u \in R$ is a *unit* if there exists $v \in R$ with $uv = 1$.

b) Show that an element $\alpha \in R$ is a unit if and only if $N(\alpha) = 1$.

For elements $\alpha, \beta \in R$, we say that β *divides* α, written $\beta \mid \alpha$, if $\alpha = \beta\gamma$ for some $\gamma \in R$; call β a *proper divisor* of α if $1 < N(\beta) < N(\alpha)$. We say that a non-unit element $\alpha \in R$ is *irreducible* if whenever $\alpha = \beta\gamma$, one has that either β or γ is a unit.

c) Show that an element $\alpha \in R$ is irreducible iff α has no proper divisors. Thus, an irreducible element of R is analogous to a prime in $\mathbb{Z}$.

d) Using the previous parts, show that every non-unit element in R has a factorization into irreducible elements.

Part d) shows that, just like in $\mathbb{Z}$, all non-unit elements of $\mathbb{Z}[\sqrt{-6}]$ factorize into products of irreducibles. However, unlike $\mathbb{Z}$, not all elements of R have a *unique* factorization.

e) Observe that

$$10 = 2 \cdot 5 = (2+\sqrt{-6})(2-\sqrt{-6}).$$

Using the norm map, show that $2, 5, 2+\sqrt{-6}$ and $2-\sqrt{-6}$ are irreducible. Thus 10 has two distinct factorizations into irreducible elements in $\mathbb{Z}[\sqrt{-6}]$.

We conclude from part e) that $\mathbb{Z}[\sqrt{-6}]$ does not have a "fundamental theorem of arithmetic." However $\mathbb{Z}[\sqrt{-6}]$, and any algebraic number ring in general, will have a unique factorization theorem for its *ideals.* (This is true of any ring that is a *Dedekind domain.*)

1.4. **The Binomial Theorem.** The main goal of this section is to state the Binomial Theorem, which in its most common form describes the coefficients in the expansion of a binomial $(x+y)^n$ when n is an integer. For us, the Binomial Theorem will specifically have later applications in solving algebraic problems in general rings (§2.11), as well as in understanding primitive roots (§2.8). However, we will not go into too much detail for this section.

Definition 1.4.1. Let k and n be integers with $n \geq k \geq 0$. Then the **binomial coefficient** $\binom{n}{k}$, read "n choose k," is defined as

$$\binom{n}{k} := \frac{n!}{(n-k)!k!},$$

where $n! := n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$ is "n factorial." By convention $0! := 1$.

Remark 1.4.1. Intuitively $n!$ describes the number of ways to arrange n objects, and $\binom{n}{k}$ is the number of ways to choose (but not arrange) k objects out of n objects.

We will take for granted the following interpretation of $\binom{n}{k}$.

Lemma 1.4.1. [NZM91, Theorem 1.20] *For any set S of n elements, the number of subsets with $k \geq 0$ elements is $\binom{n}{k}$.*

Theorem 1.4.2 (The Binomial Theorem)**.** [NZM91, Theorem 1.22] *For any integer $n \geq 1$ and real numbers x and y, one has*

$$\begin{aligned}(x+y)^n &= \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \\ &= y^n + nxy^{n-1} + \ldots + nx^{n-1}y + x^n.\end{aligned}$$

Proof. First, for a list of real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$, we consider the product

$$\prod_{i=1}^{n}(x_i + y_i) = (x_1 + y_1)(x_2 + y_2) \cdots (x_n + y_n).$$

Multiplying this out, we get 2^n monomials[3] of the form

$$\prod_{i \in I} x_i \prod_{i \notin I} y_i$$

where I ranges over each possible subset of $\{1, 2, \ldots, n\}$. Intuitively, each monomial from the expansion of $(x+y)^n$ gets a unique index set I determined by which x terms it includes in its product: for example $I = \{1, 2, n\}$ corresponds to the monomial $x_1 \cdot x_2 \cdot y_3 \cdot y_4 \cdots y_{n-1} \cdot x_n$, which comes from the product of the terms in purple:

$$(x_1 + y_1)(x_2 + y_2)(x_3 + y_3)(x_4 + y_4) \cdots (x_{n-1} + y_{n-1})(x_n + y_n).$$

Now, for each integer $0 \leq k \leq n$, we consider the monomials above corresponding to some index subset I with k elements; each such monomial is the product of k x-terms and $n-k$ y-terms. By Lemma 1.4.1, there are $\binom{n}{k}$ such indexing sets I, i.e. $\binom{n}{k}$

[3]Recall that a *monomial* is a singular term in a sum, such as $2xy$ in $(x+y)^2 = x^2 + 2xy + y^2$.

such monomials. However, if we assume that each $x_i = x$ and $y_i = y$, then each such monomial is equal to $x^k y^{n-k}$. Therefore, all indexing sets I of size k sum together to contribute the term $\binom{n}{k} x^k y^{n-k}$ to the expansion of $(x+y)^n$. □

Remark 1.4.2. It is worth noting that our proof of the Binomial Theorem is purely algebraic, and can apply to binomial expansions in ring theory. There are also analytic proofs which apply to binomial expansions of complex numbers with a real power α instead of n (although we would need to adjust the binomial coefficient definition).

Remark 1.4.3. There is a triangle called *Pascal's Triangle,* whose n'th row has as its k'th entry the binomial coefficient $\binom{n}{k}$ for each $0 \leq k \leq n$. Pascal's Triangle satisfies some interesting arithmetic with its rows. For example, the sum of the entries of any row is twice the sum of the entries of the preceding row. Additionally, each entry is the sum of the two entries directly above it. Both of these examples give additional structure to the binomial coefficients $\binom{n}{k}$.

Exercises. From [NZM91, §1.4], pages 40–41: #2, 6.

Exercise 1.4.1. Use the Binomial Theorem to show that for each integer $n \geq 0$, one has

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

1.5. **Some SageMath Computations.** In this section (not from the book), we will review some fundamental basics of SageMath. The documentation on the main site for Sage here is excellent – the purpose of this section is to get a quick rundown of how to code with Sage, so that you know how to use it for quick calculations in number theory. We will follow many of the first steps of the linked tutorial. You can use the online calculator SageCell to do computations with Sage.

Sage mostly follows the Python language. It uses = for assignment, and == for comparison. Here is an example of this, written as example code along with its output in Sage. Comments following # do not affect the code – they are just my notes to you.

(It is worth noting that SageCell will only output the most recent calculation once your code is evaluated. If you would like to output several statements at once, you will need to use print() to print the previous statements. Most of my example code here excludes the print option, since we will be doing it one line at a time.)

```
sage: a = 7
sage: a
7
sage: a == 7
true
sage: a == 1
false
sage: a >= 1 #Greater than or equal to
true
```

In Sage, different types of objects can have different object *types,* and depending on its type, one can apply various functions to said object. One example of an object type is Integers. You can read what type of object something is by applying type() to it.

```
sage: 3^2
9
sage: 3**2
9
sage: 7 % 4 #Means "7 modulo 4," which returns the remainder of 7
when divided by 4
3
sage: 7//4 #Returns the "floor" of the fraction a/b, which is
the closest integer to a/b below a/b
1
sage: type(5)
<class 'sage.rings.integer.Integer'>
```

If you do arithmetic with an object, it can change its type. Here is an example.

```
sage: a = 49
sage: type(a)
<class 'sage.rings.integer.Integer'>
a = a/8 #You replace a with its previous definition divided by 8
sage: type(a)
<class 'sage.rings.rational.Rational'>
sage: a
49/8
```

Several functions in Sage also compute approximations of real numbers to a certain amount of decimal places (which you can sometimes specify). However, several functions "prefer" to return exact values. If you want real number approximations, you often must specify this, either by taking your number and appending `.numerical_approx()` or plugging your number into `n()` or `N()`.

```
sage: sqrt(2)
sqrt(2)
sage: sqrt(RealNumber(2))
1.41421356237310
sage: n(sqrt(2))
1.41421356237310
sage: sqrt(4)
2
sage: n(sqrt(4))
2.00000000000000
sage: exp(1) #exp(x) is e^x
e
sage: exp(1).numerical_approx()
2.71828182845905
sage: n(exp(1))
2.71828182845905
sage: N(sin(1), digits=10) #Can specify how many digits you want
0.8414709848
```

Sometimes, you want to return *strings* in your code. These are simply strings of characters, and can include messages you would like to print while running code. You can put `''` or `""` around your string to have Sage recognize it as a string object.

```
sage: a = 'Hello!'
sage: a
'Hello!'
sage: type(a)
<class 'str'>
sage: b = 6
sage: type(b)
<class 'sage.rings.integer.Integer'>
sage: c = "6"
sage: type(c)
<class 'str'>
sage: c
'6'
```

It might be the case that you would like to create a *function* in Sage, which you can call on repeatedly in your code. A function needs to be initialized by the *definition* command `def`. Here is an example; indentation indicates what is included in the function's definition, and is preceded by the ellipses.

```
sage: def test_function(n):
....:   return n^2-1 #'return' is the output of the function
sage: test_function(5)
24
```

(Sometimes (but not on SageCell) you need to press "Enter" twice to exit the function's definition.)

The utility of working with computer algebra systems such as `Sage` is that there are an abundance of useful mathematical functions already implemented in it. Here are just two examples.

```
sage: is_prime(2)
True
sage: is_prime(10)
False
sage: is_prime(101)
True
sage: is_prime(1001)
False
sage: factor(1001)
7 * 11 * 13
```

It is worth noting that you can use `?` to help describe an object or function.

```
sage: sin?
Signature:      sin(self, coerce=True, hold=False,
dont_call_method_on_arg=False, *args)
Type:           Function_sin
String form:    sin
File:           /home/sc_serv/sage/src/sage/functions/trig.py
Docstring:
   The sine function.

   EXAMPLES:

      sage: sin(0)
      0
      sage: sin(x).subs(x==0)
      0
      sage: sin(2).n(100)
      0.90929742682568169539601986591
      sage: sin(x)._sympy_()
      sin(x)
#The documentation continues!
#Press 'q' to quit the documentation.
```

Next, we will briefly review loops. These can be useful when trying to iterate over a set of objects. The first is a *for loop.*

```
sage: for i in range(5): #iterates over all integers in [0,5)
....:   print(i % 2) #Returns remainder when dividing by 2
....:
0
1
0
1
0
```

In general, `range(b)` ranges over integers in the interval $[0, b)$, while `range(a,b)` ranges over $[a, b)$. **An important note:** If you are using SageCell, then you have to use `print()` on calculations in loops in order to see them as output, as we noted earlier. If you are using something other than SageCell, you might not have to specify `print()` each time.

A *while loop* can be equally useful.

```
sage: n = 2
sage: while n <= 1000:
....:   n = n * 2
....:
sage: n
1024 #Think about what this output means!
```

The next topic is on *lists*. Lists are a useful way to keep track of objects in your code. They can even track objects with different types! Lists are ordered, **and start at index 0.** You can add objects to the end of a list with the `.append()` command, and `len(L)` gives the length of a list `L`. You can also range over elements of a list in for loops.

```
sage: List = [1,2,'a']
sage: List[0]
1
sage: List[2]
'a'
sage: len(List)
3
sage: for item in List: #Can also do: "for i in range(len(List))"
....:   item
....:
1
2
'a'
sage: List.append(pi)
sage: List
[1, 2, 'a', pi]
```

Let's practice interpreting math code. Here is an example of some code with output:

```
sage: List = [1,1]
sage: c = List[0] + List[1]
sage: while c <= 1000:
....:   print(c)
....:   List.append(c)
....:   c = List[len(List)-1] + List[len(List)-2]
....:
....:
2
3
5
8
13
```

```
21
34
55
89
144
233
377
610
987
```

Mathematically, this code looks like it checks whether the sum of the previous two terms in the list is less than 1000, and if it is, it prints this sum and adds it to the end of the list. As it turns out, this list is the beginning of the *Fibonacci sequence,* whose *Fibonacci numbers* are defined recursively via $a_1 := 1$, $a_2 := 1$ and $a_k := a_{k-1} + a_{k-2}$ when $k \geq 2$. The Fibonacci numbers satisfy a lot of interesting properties, some of which can be read here: A000045.

Our next example will be slightly more complicated: we will explore primality in the list of integers of the form $101, 1001, 10001, \ldots$.

```
sage: n = '11' #written as a string to make it easy to 'insert' zeroes
sage: for i in range(100):
....:   n = n.replace('1','10',1) #replace 100...1 with 1000....1
....:   if is_prime(Integer(n)) == True:
....:           print(n+" is prime!")
....:   else:
....:           print(n+" is composite!")
....:
101 is prime!
1001 is composite!
10001 is composite!
100001 is composite!
1000001 is composite!
10000001 is composite!
100000001 is composite!
1000000001 is composite!
10000000001 is composite!
100000000001 is composite!
1000000000001 is composite!
10000000000001 is composite!
#etc.
```

What do you observe? If you extend the number of zeroes inserted, it seems to remain composite! Note that the terms of this sequence can be recursively defined via $a_k := 10^k + 1$ where $k \geq 1$; this provides a more mathematical definition.

We can study this sequence further by factorizing each such number:

```
sage: for k in range(1,30):
....:   print(str(10^k+1)+" = 10^"+str(k)+" = "+str(factor(10^k+1)))
....:
11 = 10^1 = 11
101 = 10^2 = 101
```

```
1001 = 10^3 = 7 * 11 * 13
10001 = 10^4 = 73 * 137
100001 = 10^5 = 11 * 9091
1000001 = 10^6 = 101 * 9901
10000001 = 10^7 = 11 * 909091
100000001 = 10^8 = 17 * 5882353
1000000001 = 10^9 = 7 * 11 * 13 * 19 * 52579
10000000001 = 10^10 = 101 * 3541 * 27961
100000000001 = 10^11 = 11^2 * 23 * 4093 * 8779
1000000000001 = 10^12 = 73 * 137 * 99990001
10000000000001 = 10^13 = 11 * 859 * 1058313049
100000000000001 = 10^14 = 29 * 101 * 281 * 121499449
1000000000000001 = 10^15 = 7 * 11 * 13 * 211 * 241 * 2161 * 9091
10000000000000001 = 10^16 = 353 * 449 * 641 * 1409 * 69857
100000000000000001 = 10^17 = 11 * 103 * 4013 * 21993833369
1000000000000000001 = 10^18 = 101 * 9901 * 999999000001
10000000000000000001 = 10^19 = 11 * 909090909090909091
100000000000000000001 = 10^20 = 73 * 137 * 1676321 * 5964848081
1000000000000000000001 = 10^21 = 7^2 * 11 * 13 * 127 * 2689 * 459691 *
909091
10000000000000000000001 = 10^22 = 89 * 101 * 1052788969 * 1056689261
100000000000000000000001 = 10^23 = 11 * 47 * 139 * 2531 *
549797184491917
1000000000000000000000001 = 10^24 = 17 * 5882353 * 9999999900000001
#etc.
```

Notice any patterns? Try this for larger exponents k, and come up with a conjecture! This sequence has been studied previously, see here: A000533.

For our final example, let us try to understand the asymptotic behavior of primes. First, we design a function which counts the number of primes up to any positive integer n. This is called the *prime-counting function,* and is often denoted by $\pi\colon \mathbb{Z}^+ \to \mathbb{Z}^+$. For example $\pi(10) = 4$.

```
sage: def prime_count(n):
....:     counter = 0
....:     for k in range(n):
....:             if is_prime(k) == True:
....:                     counter = counter + 1
....:     return counter
....:
sage: prime_count(10)
4
sage: prime_count(100)
25
sage: prime_count(1000000)
78498
```

Next, let us try to calculate the *proportion* of primes up to n, as $n \to \infty$. We will increment on powers of 10. Do not forget to include our previous code for the prime counting function – or, you can make a new prime-proportion function using the implemented `prime_pi` function in `Sage`.

```
sage: for k in range(1,8):
....:   print("The proportion of primes up to 10^"+str(k)+" is "+
str(n(prime_count(10^k)/10^k)))
....:
The proportion of primes up to 10^1 is 0.400000000000000
The proportion of primes up to 10^2 is 0.250000000000000
The proportion of primes up to 10^3 is 0.168000000000000
The proportion of primes up to 10^4 is 0.122900000000000
The proportion of primes up to 10^5 is 0.0959200000000000
The proportion of primes up to 10^6 is 0.0784980000000000
The proportion of primes up to 10^7 is 0.0664579000000000
```

Maybe these values are approaching something specific, but it is hard to tell with limited data. One issue is that computing this proportion can be quite slow. However, we will leave here the asymptotic values of a related function.

```
sage: for k in range(1,8):
....:   print(n(1/log(10^k)))
....:
0.434294481903252
0.217147240951626
0.144764827301084
0.108573620475813
0.0868588963806504
0.0723824136505420
0.0620420688433217
```

You might see a faint pattern, or need more convincing, but either way it turns out that one has an *asymptotic equivalence* of these two functions:

$$\frac{\pi(n)}{n} \sim \frac{1}{\log(n)},$$

i.e.,

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

In particular, the growth of the proportion of primes up to a real number x is approximately given by $\frac{1}{\log(x)}$. This is the *Prime Number Theorem,* one of the most important results in analytic number theory.

Exercises.

Exercise 1.5.1. This exercise studies the **Mersenne primes.** These are the primes of the form $a^k - 1$ where $a, k \geq 2$ are integers.

a) Create `Sage` code for a function called `MersennePrimes`, which takes as input positive integers B and k, and outputs all Mersenne primes up to B of the form $a^k - 1$. Also include calculations with your function for $k = 2, 3, \ldots, 30$ when $B = 10000000$ (ten million).

b) What do you observe with your output from part a)? Make a conjecture based on it; you can also range over larger B and k. (One point)

c) Prove that if p is a Mersenne prime, then $p = 2^k - 1$ for some $k \geq 2$. (*Hint:* use the algebraic identity $1 + x + x^2 + \ldots + x^{n-1} = \frac{x^n-1}{x-1}$.)
d) Continuing part d), show that k must also be prime.
e) Do you think there are finitely or infinitely many Mersenne primes? (One point)

The Mersenne primes can be read about here: A000043.

Bonus Exercise 1.5.2. Given a pair of positive integers (a, b), let $\pi_{a,b} \colon \mathbb{R} \to \mathbb{Z}^+$ be the function such that $\pi_{a,b}(x)$ counts the number of primes $1 \leq p \leq x$ of the form $a + bk$. For example $\pi_{1,3}(20) = 3$.

a) Create code which calculates $\pi_{3,4}(x)$ for $x = 10, 10^2, \ldots, 10^{10}$. (To reiterate, we are counting the number of primes $p \leq x$ of the form $3 + 4k$.)
b) For each x above, compute the ratio $\pi_{3,4}(x)/\pi(x)$. (This is the proportion of primes in $[1, x]$ of the form $3 + 4k$.)
c) What does the limit
$$\lim_{x \to \infty} \frac{\pi_{3,4}(x)}{\pi(x)}$$
seem to equal? Make a conjecture for primes of the form $3 + 4k$ based on this.
d) Do the same analysis for primes of the form $1 + 4k$. Explore this for other a, b as well. Can you come up with a general conjecture for the proportion of primes of the form $a + bk$, where $a, b \in \mathbb{Z}^+$ are coprime?

Bonus Exercise 1.5.3. The following theorem is a conjecture based on Dirichlet's Theorem above on Primes in Arithmetic Progressions, vastly generalizing it.

Conjecture (The Bunyakovsky Conjecture)**.** *Let $f(x)$ be a polynomial with integer coefficients, satisfying the following three properties:*

i) The leading coefficient of $f(x)$ is positive;
ii) $f(x)$ is irreducible over $\mathbb{Z}$;
iii) $\gcd(f(1), f(2), f(3), \ldots) = 1$.[4]

Then there are infinitely many primes of the form $f(n)$ where n ranges over positive integers.

a) Show that Dirichlet's Theorem on Primes in Arithmetic Progressions is a special case of the Bunyakovsky Conjecture.
b) Show that the following well-known conjecture is a special case of the Bunyakovksy Conjecture:

Conjecture. *There are infinitely many primes of the form $n^2 + 1$.*

c) The Bunyakovsky Conjecture is currently open for all polynomials of degree greater than 1 satisfying $i) - iii)$ above. Pick your favorite polynomial in $\mathbb{Z}[x]$ and use computations to try to understand whether $f(n)$ is prime for various values of n. If your polynomial doesn't satisfy all of $i) - iii)$, what do you observe goes wrong?

[4]Recall that $\gcd(a_1, a_2, \ldots, a_k)$ is the greatest *simultaneous* divisor of the integers $a_1, a_2, \ldots, a_k$, i.e., the largest integer $d \in \mathbb{Z}$ for which $d \mid a_1, d \mid a_2, \ldots, d \mid a_k$. Then $\gcd(a_1, a_2, a_3, \ldots)$ is the greatest common divisor between all of the a_i terms.

2. Congruences

In this chapter, we will work with *congruences.* Congruences allow one to describe when two integers have the same remainder once divided by a "modulus." However, congruences capture so much more information on the arithmetic of integers, and can help prove new results. We will learn about modular arithmetic, and study solutions to congruences: this will include learning about Fermat's Little Theorem and Euler's Theorem, Wilson's Theorem, the Chinese Remainder Theorem and Hensel's Lemma. We will also learn some basic group theory, and see how it recontextualizes what we know about modular arithmetic, culminating in a study of primitive roots and power residues.

2.1. Congruences.

The main goals of this section are the following:

- Prove basic results on congruences.
- Learn about residue systems.
- Review Euler's phi function $\varphi(n)$.
- Prove Fermat's Little Theorem, Euler's Theorem and Wilson's Theorem.
- Study roots of $x^2 + 1$ modulo primes.

Definition 2.1.1. For integers a, b and m, we say that a **is congruent to** b **modulo** m**,** and write

$$a \equiv b \pmod{m},$$

if there exists $k \in \mathbb{Z}$ with

$$a = b + mk.$$

Note that $a \equiv b \pmod{m}$ is equivalent to

$$m \mid (a - b).$$

In this context, we call m the **modulus.** Without qualification, we will always assume that $m > 0$.

If $m \nmid (a - b)$, we write $a \not\equiv b \pmod{m}$, and say that a and b are not congruent (or *incongruent*) modulo m.

Example 2.1.1. Convince yourself of the following:

- $13 \equiv 1 \pmod{12}$.
- $4 \equiv -3 \pmod{7}$.
- For any odd number n, one has $n^2 \equiv 1 \pmod{4}$. To see this, write $n = 2k + 1$ for some $k \in \mathbb{Z}$. Then

$$\begin{aligned} n^2 &= 4k^2 + 4k + 1 \\ &\equiv 1 \pmod{4}. \end{aligned}$$

Here are some basic properties of congruences.

Theorem 2.1.1. [NZM91, Theorem 2.1] *Let a, b, c and d be integers.*

(1) The following are equivalent:

- $a \equiv b \pmod{m}$.
- $b \equiv a \pmod{m}$.

- $a - b \equiv 0 \pmod{m}$.

(2) *If* $a \equiv b \pmod{m}$ *and* $b \equiv c \pmod{m}$, *then* $a \equiv c \pmod{m}$.
(3) *If* $a \equiv b \pmod{m}$ *and* $c \equiv d \pmod{m}$*:*
  (a) *then* $a + c \equiv b + d \pmod{m}$,
  (b) *and* $ac \equiv bd \pmod{m}$.
(4) *If* $a \equiv b \pmod{m}$ *and* $d \mid m$ *with* $d > 0$, *then* $a \equiv b \pmod{d}$.
(5) *If* $a \equiv b \pmod{m}$, *then for all* $c \in \mathbb{Z}^+$ *one has* $ac \equiv bc \pmod{mc}$.

Proof. We will just prove (3). Since $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, we can write $a = b + mk$ and $c = d + m\ell$ for some $k, \ell \in \mathbb{Z}$. Then we check that

$$\begin{aligned} a + c &= b + mk + d + m\ell \\ &= b + d + m(k + \ell), \end{aligned}$$

whence we deduce that $a + c \equiv b + d \pmod{m}$. We also check that

$$\begin{aligned} ac &= (b + mk)(d + m\ell) \\ &= bd + m(kd + b\ell + mk\ell), \end{aligned}$$

so that $ac \equiv bd \pmod{m}$. □

What is also interesting is that polynomials preserve congruences.

Theorem 2.1.2. [NZM91, Theorem 2.2] *Let* $f(x) \in \mathbb{Z}[x]$, *i.e., let* $f(x)$ *be a polynomial with integer coefficients. If* $a \equiv b \pmod{m}$ *then* $f(a) \equiv f(b) \pmod{m}$.

Proof. Let us write

$$f(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n$$

where each $c_i \in \mathbb{Z}$. By Theorem 2.1.1.(3).(b) ([NZM91, Theorem 2.1.(4)]), since

$$a \equiv b \pmod{m},$$

we find that for each $i \geq 0$ we also have

$$a^i \equiv b^i \pmod{m},$$

and thus

$$c_i a^i \equiv c_i b^i \pmod{m}.$$

Then by Theorem 2.1.1.(3).(a) ([NZM91, Theorem 2.1.(3)]), adding each of these terms together gives

$$c_0 + c_1 a + c_2 a^2 + \ldots + c_n a^n \equiv c_0 + c_1 b + c_2 b^2 + \ldots + c_n b^n \pmod{m},$$

i.e.,

$$f(a) \equiv f(b) \pmod{m}.$$ □

Remark 2.1.1. As we will see later in this chapter (and further in the course), finding solutions to polynomials modulo m is an intrinsically interesting problem, and can be connected to finding integer and rational solutions to polynomials, which is a classical Diophantine problem.

This next theorem further describes basic properties of modular arithmetic.

Theorem 2.1.3. [NZM91, Theorem 2.3]

(1) $ax \equiv ay \pmod{m}$ *if and only if* $x \equiv y \pmod{\frac{m}{\gcd(a,m)}}$.
(2) *If* $\gcd(a,m) = 1$, *then* $ax \equiv ay \pmod{m}$ *iff* $x \equiv y \pmod{m}$.
(3) *Given integers* $m_1, m_2 \ldots, m_r$, *one has for all* $1 \leq i \leq r$ *that*

$$x \equiv y \pmod{m_i}$$

iff

$$x \equiv y \pmod{\operatorname{lcm}(m_1, m_2, \ldots, m_r)}.$$

Proof.

(1) Suppose that $ax \equiv ay \pmod{m}$. Then $ax = ay + mk$ for some $k \in \mathbb{Z}$. Thus $a \cdot (x - y) = m \cdot k$, and so dividing both sides by $\gcd(a,m)$ gives

$$\frac{a}{\gcd(a,m)} \cdot (x - y) = \frac{m}{\gcd(a,m)} \cdot k.$$

We deduce that

$$\frac{m}{\gcd(a,m)} \mid \frac{a}{\gcd(a,m)} \cdot (x - y).$$

However, we know that $\frac{m}{\gcd(a,m)}$ and $\frac{a}{\gcd(a,m)}$ are coprime since they share no common divisors (recall Theorem 1.2.5.(2) ([NZM91, Theorem 1.7])). We thus have by Theorem 1.2.6 ([NZM91, Theorem 1.11]) that

$$\frac{m}{\gcd(a,m)} \mid (x - y),$$

and so

$$x \equiv y \pmod{\frac{m}{\gcd(a,m)}}.$$

Try and prove the converse direction on your own!

(2) This is a consequence of part (1).

(3) We will prove the forward direction. Assume that for each $1 \leq i \leq r$, we have

$$x \equiv y \pmod{m_i},$$

i.e.,

$$m_i \mid (x - y).$$

Then $x - y$ is a common multiple of each $m_1, m_2, \ldots, m_r$, and so

$$\operatorname{lcm}(m_1, m_2, \ldots, m_r) \mid (x - y)$$

by an inductive version of Theorem 1.2.8 ([NZM91, Theorem 1.12]). This is equivalent to

$$x \equiv y \pmod{\operatorname{lcm}(m_1, m_2, \ldots, m_r)}.$$ □

In general, given an integer $m > 0$, for any $a \in \mathbb{Z}$ we get by the Division Algorithm that

$$a = mq + r$$

for some $q, r \in \mathbb{Z}$ with $0 \leq r < m$. This implies that

$$a \equiv r \pmod{m}.$$

Therefore, every integer is congruent modulo m to *exactly one* of $0, 1, 2, \ldots, m-1$. This motivates the following definitions.

Definition 2.1.2. Given an integer a, if $a \equiv r \pmod{m}$, then say r is a **residue of** a **modulo** m**.** A set of integers $\{r_1, r_2, \ldots, r_m\}$ is called a **complete residue system modulo** m (or CRS mod m) if for any integer a, there exists a unique r_i with $a \equiv r_i \pmod{m}$.

Definition 2.1.3. Given integers a and $m > 0$, the set of integers

$$\begin{aligned} S(a,m) &:= \{a + mk \ : \ k \in \mathbb{Z}\} \\ &= \{\ldots, a-3m, a-2m, a-m, a, a+m, a+2m, a+3m, \ldots\} \end{aligned}$$

is an arithmetic progression, called the **congruence class,** or **residue class, of** a **modulo** m**.** It is often written as $\overline{a} \mod m$ or $[a] \mod m$, or simply $\overline{a}$ or $[a]$ when the modulus is clear.

Remark 2.1.2. When talking about congruence classes modulo m, we often refer to them by their representatives. For example, the congruence class of odd numbers, denoted $S(1,2) = [1] \mod 2 = \overline{1} \mod 2$, is often just called the class of 1 modulo 2.

Remark 2.1.3. Each number in $S(a,m)$ is congruent to a modulo m. There are m distinct residue classes mod m, given by $S(0,m), S(1,m), S(2,m) \ldots, S(m-1,m)$. We thus have

$$\bigcup_{k=0}^{m-1} S(k,m) = \mathbb{Z}.$$

There is another type of residue system which is important when doing *multiplicative* calculations modulo m.

Definition 2.1.4. A **reduced residue system modulo** m (or RRS mod m) is a set of integers $\{r_1, r_2, \ldots, r_n\}$ with each r_i coprime to m, such that any integer a *coprime to* m*,* there exists a unique r_i with $a \equiv r_i \pmod{m}$.

Remark 2.1.4. One can show that any set of m integers is a complete residue system modulo m if and only if no two elements in the set are congruent modulo m: this is because there are only m possible remainders in $[0, m)$. In particular, any two complete residue systems must share the same cardinality m. Similarly, any two reduced residue systems modulo m will share the same size.

The following theorem says that congruent numbers modulo m share the same GCD with m. This will help us construct reduced residue systems from complete residue systems, by simply removing the classes which are not coprime to m.

Theorem 2.1.4. [NZM91, Theorem 2.4] *If* $a \equiv b \pmod{m}$*, then* $\gcd(a,m) = \gcd(b,m)$.

Proof. Write $a = b + mk$ for some $k \in \mathbb{Z}$. Then

$$\begin{aligned} \gcd(a,m) &= \gcd(b+mk, m) \\ &= \gcd(b+mk-(mk), m) \quad \text{(by Theorem 1.2.5.(4) [NZM91, Theorem 1.9]))} \\ &= \gcd(b,m). \end{aligned}$$

□

As noted in Remark 2.1.4, any two reduced residue systems modulo m will have the same cardinality. Thus, this size is an invariant of m; it is worth giving it a name.

Definition 2.1.5. Given an integer $n > 0$, we let $\varphi(n)$ denote the number of elements in any reduced residue system modulo n. This number is called the **Euler phi function,** or **Euler's totient function.**

Here is a useful formulation for $\varphi(n)$.

Theorem 2.1.5. [NZM91, Theorem 2.5] *For each integer $n > 0$, the value $\varphi(n)$ is the number of integers in $[1, n]$ that are coprime to n.*

Proof. This follows from Theorem 2.1.4 ([NZM91, Theorem 2.4]): from the complete residue system $S := \{0, 1, 2, \ldots, n\}$, one constructs the reduced residue system by removing the integers in S which are not coprime to n. □

Euler's phi function appears in many contexts, and is interesting to study in its own right. We will see more of it soon. You can read about its sequence of values here: A000010.

Example 2.1.2. Let us do some basic calculations with $\varphi(n)$:

- $\varphi(7) = 6$, since the integers in $[1, 7]$ that are coprime to 7 are $a = 1, 2, 3, 4, 5, 6$.
- $\varphi(8) = 4$. For example, a RRS mod 8 is $\{1, 3, 5, 7\}$.
- $\varphi(6) = 2$: the only two integers in $[1, 6]$ coprime to 6 are 1 and 5.
- $\varphi(12) = 4$, since a RRS mod 12 is $\{1, 5, 7, 11\}$.

There are natural choices for representatives of congruences classes modulo m, and thus for residue systems: namely, the integers in $[0, m)$. However, you can modify a residue system by multiplying each representative by a fixed number coprime to m.

Theorem 2.1.6. [NZM91, Theorem 2.6] *Let* $\gcd(a, m) = 1$. *If $\{r_1, r_2, \ldots, r_n\}$ is a complete or reduced residue system modulo m, then so is $\{ar_1, ar_2, \ldots, ar_n\}$.*

Example 2.1.3. We saw that a RRS mod 8 is $\{1, 3, 5, 7\}$. By this theorem, so is $\{3, 9, 15, 21\}$.

Proof. Let $\{r_1, r_2, \ldots, r_n\}$ be a complete/reduced residue system modulo m. Assume that $\gcd(a, m) = 1$. To show that $\{ar_1, ar_2, \ldots, ar_n\}$ is also a CRS/RRS, it suffices to show that these two sets have the same size, since that implies $\{ar_1, ar_2, \ldots, ar_n\}$ contains precisely all of the remainders modulo m when $\{r_1, r_2, \ldots, r_n\}$ is a CRS, and precisely all of the remainders of integers coprime to m when $\{r_1, r_2, \ldots, r_n\}$ is a RRS.

To show these two sets have the same cardinality, it suffices to prove that no two ar_i are congruent modulo m. If

$$ar_i \equiv ar_j \pmod{m},$$

then since $\gcd(a, m) = 1$, we have by Theorem 2.1.3.(2) ([NZM91, Theorem 2.3]) that

$$r_i \equiv r_j \pmod{m},$$

which forces $i = j$. This concludes our proof. □

We next turn towards proving several results about the multiplicative structure of integers under a modulus. If you have taken an abstract algebra course before, you might recognize a few of these results as applications of group theory – we will revisit this interpretation later in this chapter.

Our first result is attributed to Euler.

Theorem 2.1.7 (Euler's Theorem)**.** *If* $\gcd(a, m) = 1$*, then*

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

Proof. Let $\{r_1, r_2, \ldots, r_n\}$ be a reduced residue system modulo m. Then by Theorem 2.1.6 ([NZM91, Theorem 2.6]), so is $\{ar_1, ar_2, \ldots, ar_n\}$; these two residue systems are equal up to reducing the representatives modulo m. We thus have

$$\begin{aligned}\prod_{i=1}^{n} r_i &\equiv \prod_{i=1}^{n} ar_i \pmod{m} \\ &= a^{\varphi(m)} \prod_{i=1}^{n} r_i \pmod{m} \qquad \text{(since the size of a RRS mod } m \text{ is } \varphi(m)).\end{aligned}$$

We can cancel out the product $\prod_{i=1}^{n} r_i$ from both sides since $\gcd\left(\prod_{i=1}^{n} r_i, m\right) = 1$, as per Theorem 2.1.3.(2) ([NZM91, Theorem 2.3]). We conclude that

$$1 \equiv a^{\varphi(m)} \pmod{m}. \qquad \square$$

The following is a special case of Euler's Theorem, and was historically proven first.

Theorem 2.1.8 (Fermat's Little Theorem)**.** *For a prime* p*, if* $p \nmid a$ *then one has*

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

Proof. By Euler's Theorem, it suffices to show that $\varphi(p) = p - 1$. This is true by Theorem 2.1.5 ([NZM91, Theorem 2.5]), since all integers $1 \leq k < p$ are coprime to p. $\square$

Example 2.1.4. Let us double-check that Euler's Theorem holds for a few cases.

- Taking $m = 6$ and $a = 5$, we have $\varphi(6) = 2$, so by Euler's Theorem
$$5^2 \equiv 1 \pmod{6}.$$
Indeed $6 \mid (25 - 1) = 24$.
- Taking $m = 12$ and $a = 7$, we have $\varphi(12) = 4$, so by Euler's Theorem
$$7^4 \equiv 1 \pmod{12}.$$
Indeed $12 \mid (7^4 - 1) = 2400 = 12 \cdot 200$.
- Taking $m = 11$ and $a = 2$, we have $\varphi(11) = 10$, so by Euler's Theorem (or Fermat's Little Theorem)
$$2^{10} \equiv 1 \pmod{11}.$$
Indeed $11 \mid (2^{10} - 1) = 1023 = 11 \cdot 93$.

The concept of a *multiplicative inverse* of an integer is important. While the reciprocal of an integer a is usually not an integer, there is a notion of an inverse to a modulo m, which multiplied with $\overline{a}$ equals $\overline{1}$.

Theorem 2.1.9. [NZM91, Theorem 2.9] *If* $\gcd(a, m) = 1$*, then there exists* $x \in \mathbb{Z}$ *with* $ax \equiv 1 \pmod{m}$*. Such an* x *is unique modulo* m*.*

Proof. We harken back to §1.2: by Theorem 1.2.3 ([NZM91, Theorem 1.3]), we know that $\gcd(a, m) = 1$ is a $\mathbb{Z}$-linear combination of a and m; let us write

$$ax + my = 1$$

for some $x, y \in \mathbb{Z}$. Then reducing modulo m shows that

$$ax \equiv 1 \pmod{m}.$$

This shows that the multiplicative inverse of a modulo m exists.

Next, we show uniqueness of the inverse. Continuing the above, suppose $y \in \mathbb{Z}$ is such that

$$ay \equiv 1 \pmod{m}.$$

Let us multiply both sides by x:

$$\begin{aligned} x &\equiv x \cdot ay && \pmod{m} \\ &= (xa)y && \pmod{m} \\ &\equiv y && \pmod{m} \qquad \text{(since } x \text{ is an inverse of } a \text{ modulo } m\text{).} \end{aligned}$$

Thus $x \equiv y \pmod{m}$. □

Definition 2.1.6. Given integer a with $\gcd(a, m) = 1$, writing $ax \equiv 1 \pmod{m}$ we call (the residue class of) x the **multiplicative inverse of** a **modulo** m**.** It is denoted $a^{-1} := x$, where we allow a to represent both itself and its residue class mod m.

Remark 2.1.5. When $\gcd(a, m) = 1$, we have at least three ways to get the multiplicative inverse of a modulo m:

1. Using the Euclidean/Blankinship's Algorithm to write $1 = ax + my$ for some $x, y \in \mathbb{Z}$. It then follows that $a^{-1} \equiv x \pmod{m}$.
2. Deducing that $a^{-1} \equiv a^{\varphi(m)-1} \pmod{m}$ from Euler's Theorem.
3. If m is small, then we can reasonably multiply a by each integer $1 \leq k < m$ individually, until we find $ak \equiv 1 \pmod{m}$; then $a^{-1} \equiv k \pmod{m}$.

Example 2.1.5. We will find multiplicative inverses for the following congruence classes.

- $2 \mod 3$: since 3 is small, we can multiply 2 by all integers $1 \leq k < 3$ until we find the inverse. We see that $2 \cdot 2 \equiv 1 \pmod{3}$, so that
$$2^{-1} \equiv 2 \pmod{3}.$$
- $14 \mod 5$: $14 \equiv -1 \pmod{5}$, and since $(-1)^2 = 1$, it follows that
$$14^{-1} \equiv -1 \equiv 14 \pmod{5}.$$

- 2 mod 37: since 37 is odd, by Fermat's Little Theorem we have $2^{\varphi(37)} \equiv 1 \pmod{37}$. Since $\varphi(37) = 36$, we deduce that

$$2^{-1} \equiv 2^{35} \pmod{37}.$$

Suppose we want to find a smaller representative for 2^{35} modulo 37. One way is to compute the power 2^{35} "one step at a time," by repeatedly converting it to a power of a smaller power and then reducing that smaller power mod 37. Here is one way to do this:

$$\begin{aligned}
2^{35} &= (2^5)^7 && \pmod{37} \\
&= (32)^7 && \pmod{37} \\
&\equiv (-5)^7 && \pmod{37} && (\text{since } 32 \equiv -5 \pmod{37}) \\
&= -5^7 && \pmod{37} \\
&= -5 \cdot (5^2)^3 && \pmod{37} \\
&\equiv -5 \cdot (-12)^3 && \pmod{37} && (\text{since } 5^2 \equiv 25 \equiv -12 \pmod{37}) \\
&= 5 \cdot 12^3 && \pmod{37} \\
&= 60 \cdot 144 && \pmod{37} \\
&\equiv -14 \cdot -4 && \pmod{37} && (\text{since } 60 \equiv -14 \pmod{37} \text{ and } 144 = 148 - 4 \equiv -4 \pmod{37}) \\
&= 56 && \pmod{37} \\
&\equiv 19 && \pmod{37}.
\end{aligned}$$

We conclude that

$$2^{-1} \equiv 19 \pmod{37}.$$

Remark 2.1.6. The algebra we did above is good practice for dealing with finding inverses over a larger modulus, as we will see later in this chapter (and in Chapter 3).

So far, we have two characterizations of prime numbers in $\mathbb{Z}$. An integer $p > 1$ is prime if one of the following holds.

1. It has no nontrivial proper divisors.
2. For any $a, b \in \mathbb{Z}$, if $p \mid ab$ then $p \mid a$ or $p \mid b$.

We saw that 1. implies 2. in Lemma 1.3.3 ([NZM91, Theorem 1.15]), but one can also show that 2. implies 1.; I encourage you to try and prove this.

We will provide a third characterization of prime numbers; this is also a classic result in elementary number theory.

Theorem 2.1.10 (Wilson's Theorem)**.** *An integer $p > 1$ is prime if and only if*

$$(p-1)! \equiv -1 \pmod{p},$$

i.e. $p \mid ((p-1)! + 1)$.

Wilson's Theorem has applications in determining the existence of roots of $x^2 + 1$ modulo primes p; see the upcoming Theorem 2.1.12 ([NZM91, Theorem 2.12]), which also has later application towards studying quadratic residues.

To prove Wilson's Theorem, we will need one small lemma.

Lemma 2.1.11. [NZM91, Lemma 2.10] *Let p be prime. Then for all $x \in \mathbb{Z}$, one has $x^2 \equiv 1 \pmod{p}$ if and only if $x \equiv \pm 1 \pmod{p}$.*

This lemma says that $x^2 - 1$ *only* has the obvious roots ± 1 modulo p. We will study solutions to more complicated polynomials mod p throughout this chapter.

Proof of lemma. The backward direction is clear. For the forward direction, if $x^2 \equiv 1 \pmod{p}$, then $p \mid (x^2 - 1) = (x+1)(x-1)$. Since p is prime, this implies $p \mid (x+1)$ or $p \mid (x-1)$, whence we have $x \equiv 1 \pmod{p}$ or $x \equiv -1 \pmod{p}$. □

Proof of Wilson's Theorem. First, we prove the backward direction. Suppose that

$$(p-1)! \equiv -1 \pmod{p}.$$

Then we can write

$$pk = (p-1)! + 1$$

for some $k \in \mathbb{Z}$. Towards a contradiction: if p is composite, then $p = ab$ for some $1 < a, b < p$. Since $a \leq p - 1$, we know that $a \mid (p-1)!$. Since $a \mid p$, we get

$$a \mid (pk - (p-1)!) = 1,$$

which is impossible since $a > 1$. We conclude that p is prime.

Next, we prove the forward direction; suppose that p is prime. We want to show that

$$(p-1)! \equiv -1 \pmod{p}.$$

We will do this by proving the fact that every term k in the factorial

$$(p-1)! := (p-1) \cdot (p-2) \cdots k \cdots 2 \cdot 1$$

has a multiplicative inverse modulo p that also shows up in the factorial.

For each $1 \leq k \leq p-1$, since $\gcd(k, p) = 1$ there exists a unique $1 \leq x_k \leq p-1$ with $kx_k \equiv 1 \pmod{p}$; here x_k is the multiplicative inverse of k mod p. This is Theorem 2.1.9 ([NZM91, Theorem 2.9]). If $k \equiv x_k \pmod{p}$, then we have $k^2 \equiv 1 \pmod{p}$, so by Lemma 2.1.11 ([NZM91, Lemma 2.10]) this forces $k \equiv \pm 1 \pmod{p}$. We deduce that when $1 < k < p-1$, we have $k \neq x_k$.

Next, we claim that each integer $1 < k < p-1$ *corresponds* to a unique inverse in $1 < x_k < p-1$, i.e., if $1 < \ell < p-1$ has $x_\ell = x_k$, then $\ell = k$. To see this, multiply the equation $x_\ell = x_k$ by $k\ell$, and reduce modulo p: then

$$k \equiv \ell \pmod{p},$$

whence $p \mid (k - \ell)$. However, from $0 < k, \ell < p-1$ we have $0 \leq |k - \ell| < p$, which from $p \mid (k - \ell)$ forces $k = \ell$.

We conclude from the above considerations that each term k in

$$(p-2)! := (p-2) \cdot (p-3) \cdots k \cdots 2$$

corresponds to a unique term x_k in $(p-2)!$ with $x_k \neq k$. In particular, we can pair off terms in $(p-1)!$, and turn them into products of one:

$$\begin{aligned}
(p-1)! :&= \prod_{k=1}^{p-1} k \\
&\equiv 1 \cdot (-1) \cdot \prod_{k=2}^{p-2} k && \pmod{p} \\
&= -1 \cdot \prod kx_k && \pmod{p} && \text{(by our observations above)} \\
&\equiv -1 \cdot \prod 1 && \pmod{p} \\
&= -1 && \pmod{p}.
\end{aligned}$$

We conclude that

$$(p-1)! \equiv -1 \pmod{p}. \qquad \square$$

One application of Wilson's Theorem is to study solutions to x^2+1 modulo p. Note that over $\mathbb{C}$, the polynomial x^2+1 always has roots $\pm i$, but algebraic solutions to x^2+1 are *not* guaranteed modulo arbitrary p; for example, you cannot necessarily reduce i "modulo p."

Theorem 2.1.12. [NZM91, Theorem 2.12] *For a prime p, the congruence*

$$x^2 \equiv -1 \pmod{p}$$

has solutions if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

A proof of this theorem will also be given via applications of primitive roots in §2.8. This theorem is especially important when we study quadratic residues in Chapter 3.

Proof. When $p = 2$, we have $-1 \equiv 1 \pmod 2$, so take $x = 1$. Assume then that $p > 2$. By Wilson's Theorem, we have

$$(p-1)! \equiv -1 \pmod{p},$$

i.e.,

(3)
$$\left(1 \cdot 2 \cdots k \cdots \frac{p-1}{2}\right) \cdot \left(\left(\frac{p-1}{2}+1\right) \cdots (p-k) \cdots (p-2) \cdot (p-1)\right) \equiv -1 \pmod{p}.$$

For each $1 \leq k \leq \frac{p-1}{2}$, observe that both $\frac{p-1}{2} < p-k \leq p-1$ and

$$k \cdot (p-k) \equiv -k^2 \pmod{p}.$$

Thus, pairing off each term k with $p-k$ in Equation (3) implies that

$$(p-1)! \equiv \prod_{k=1}^{\frac{p-1}{2}} (-k^2) \equiv -1 \pmod{p},$$

and thus

$$(-1)^{\frac{p-1}{2}} \cdot \left(\prod_{k=1}^{\frac{p-1}{2}} k\right)^2 \equiv -1 \pmod{p},$$

so that

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

where $a := \prod_{k=1}^{\frac{p-1}{2}} k = \left(\frac{p-1}{2} - 1\right)!$. In particular, if $p \equiv 1 \pmod 4$, then writing $p = 1 + 4k$ we have

$$(-1)^{\frac{p+1}{2}} = (-1)^{\frac{2+4k}{2}} = (-1)^{1+2k} = -1,$$

so that a is a solution to $x^2 \equiv -1 \pmod p$.

Conversely, suppose that

$$a^2 \equiv -1 \pmod{p}$$

for some $a \in \mathbb{Z}$. Then taking both sides to the $\frac{p-1}{2}$'th power gives

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

However, by Fermat's Little Theorem we know $a^{p-1} \equiv 1 \pmod p$, and so

$$(-1)^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Thus $p \mid ((-1)^{\frac{p-1}{2}} - 1)$, and since $p > 2$ this forces $(-1)^{\frac{p-1}{2}} = 1$. In particular $\frac{p-1}{2}$ must be even, so that $\frac{p-1}{2} = 2k$ for some $k \in \mathbb{Z}$, and thus $p = 1 + 4k$, i.e. $p \equiv 1 \pmod 4$. □

Example 2.1.6. To illustrate the validity of the theorem above, here are some examples of determining solutions to $x^2 + 1$ modulo p.

- $5 \equiv 1 \pmod 4$, and we can check that $x^2 + 1$ modulo 5 has solutions

  $$x \equiv \pm 2 \equiv 2, 3 \pmod 5,$$

  i.e.,

  $$2^2, 3^2 \equiv -1 \pmod 5.$$

- $7 \equiv 3 \pmod 4$, and so by the theorem $x^2 + 1$ should have no roots modulo 7. We can verify this by computing all squares mod 7. These are:
  - $0^2 = 0$.
  - $1^2 = 1$.
  - $2^2 = 4$.
  - $3^2 \equiv 2 \pmod 7$.
  - $4^2 \equiv 2 \pmod 7$.
  - $5^2 \equiv 4 \pmod 7$.
  - $6^2 \equiv 1 \pmod 7$.

  None of these squares are -1 modulo 7. (As an aside, do you notice a pattern in the calculations?)
- $13 \equiv 1 \pmod 4$; by the theorem, we know that $x^2 + 1$ has roots modulo 13. What are they?

The next theorem connects solutions of x^2+1 modulo p, to representing p as a sum of two squares. This also has connections to algebraic number theory – see Bonus Exercise 2.1.13.

Theorem 2.1.13. [NZM91, Lemma 2.13] *For an odd prime p, one has*

$$p \equiv 1 \pmod 4$$

if and only if

$$p = a^2 + b^2$$

for some $a, b \in \mathbb{Z}$.

For this theorem, we will use the *Pigeonhole Principle:* if m items are put into n containers and $m > n$, then at least two items are in the same container.

Proof. First, the backwards direction: suppose we can write $p = a^2 + b^2$. Then since any integer $x \in \mathbb{Z}$ satisfies $x^2 \equiv 0, 1 \pmod 4$, we find that $p \equiv 0, 1, 2 \pmod 4$. Since p is odd, we have $p \not\equiv 0, 2 \pmod 4$, which forces $p \equiv 1 \pmod 4$.

For the forward direction, assume that $p \equiv 1 \pmod 4$. Then by Theorem 2.1.12 ([NZM91, Theorem 2.12]), there exists $c \in \mathbb{Z}$ with

$$c^2 \equiv -1 \pmod p. \tag{4}$$

Since p is prime, we know that $\sqrt{p}$ is not an integer. Let us set the *floor* $K := \lfloor \sqrt{p} \rfloor$, the largest integer less than $\sqrt{p}$. Then we have $K < \sqrt{p} < K + 1$. Consider pairs $(u, v) \in \mathbb{Z}^2$ where

$$0 \leq u, v \leq K.$$

Then u and v each take on $K + 1$ possible values, and so there are $(K + 1)^2$ possible pairs. Since $K + 1 > \sqrt{p}$, there are more than p such pairs.

Define a function $f \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by

$$f(u, v) := u + cv.$$

We know that $f(u, v)$ can only take at most p distinct values modulo p. However, there are more than p pairs (u, v), so by the Pigeonhole Principle there are two distinct pairs (u_1, v_1) and (u_2, v_2) with the same image mod p under f:

$$f(u_1, v_1) \equiv f(u_2, v_2) \pmod p,$$

i.e.,

$$u_1 + cv_1 \equiv u_2 + cv_2 \pmod p.$$

Thus

$$u_1 - u_2 \equiv -c(v_1 - v_2) \pmod p.$$

Then squaring both sides gives

$$\begin{aligned} (u_1 - u_2)^2 &\equiv c^2(v_1 - v_2)^2 && \pmod p \\ &\equiv -(v_1 - v_2)^2 && \pmod p \qquad \text{(by Equation (4)).} \end{aligned}$$

Setting $a := u_1 - u_2$ and $b := v_1 - v_2$, we can write this as

$$a^2 \equiv -b^2 \pmod p,$$

so that

$$p \mid (a^2 + b^2). \tag{5}$$

We claim the divisibility (5) is an equality. First, we note that

$$a^2 + b^2 > 0 \tag{6}$$

since $a \neq 0$ or $b \neq 0$ (as the original pairs (u_1, v_1) and (u_2, v_2) were distinct). Since $0 \leq u_i, v_i \leq K$, we also know that

$$0 \leq |a|, |b| \leq K.$$

We can square both sides of the above and get from $K < \sqrt{p}$ that

$$0 \leq a^2, b^2 < p.$$

Combining this with (6) gives

$$0 < a^2 + b^2 < 2p.$$

However, we know by (5) that $p \mid (a^2 + b^2)$, and the only multiple of p strictly between 0 and $2p$ is p. We conclude that $p = a^2 + b^2$. □

Remark 2.1.7. In this section, we have proven the following equivalencies for any odd prime p.

$$\begin{aligned} & p \equiv 1 \pmod 4 \\ \Leftrightarrow\ & p = a^2 + b^2 \text{ for some } a, b \in \mathbb{Z} \\ \Leftrightarrow\ & x^2 \equiv -1 \pmod p \text{ has a solution.} \end{aligned}$$

As noted previously, there is a fourth equivalency from algebraic number theory, via the *splitting behavior* of p in $\mathbb{Z}[i]$, see Bonus Exercise 2.1.13.

Example 2.1.7. Since 5, 13, 17 and 29 are 1 modulo 4, we can write them as sums of two squares: $5 = 1 + 4$, $13 = 4 + 9$, $17 = 1 + 16$, and $29 = 4 + 25$. However, the primes 7, 11, 19 and 23 are 3 modulo 4, and thus are not sums of two squares – this can be verified by hand.

Exercises. From [NZM91, §2.1], pages 56–57: $\#1 - 6, 10 - 15, 17, 32$.

Exercise 2.1.1.

a) List all integers $1 \leq n \leq 100$ which are congruent to 3 modulo 18.
b) Give a complete residue system modulo 9 comprised of multiples of 4.
c) Give a reduced residue system modulo 14. From this, compute $\varphi(14)$.
d) Give a reduced residue system modulo 11. For each representative r of this residue system, give a multiplicative inverse $r^{-1} \in \mathbb{Z}$ with $0 \leq r^{-1} < 11$.

Exercise 2.1.2. Recall the *Freshman's Dream,* which is the erroneous conclusion that for all real numbers x and y and for any integer $n \geq 2$, one has

$$(x + y)^n = x^n + y^n.$$

In spite of the above, show that for any integers a and b and prime p, one has

$$(a + b)^p \equiv a^p + b^p \pmod p.$$

Exercise 2.1.3. Show that for prime powers $p^e > 1$, one has $\varphi(p^e) = p^e - p^{e-1}$.

Exercise 2.1.4. Show that for any integer $m > 0$, one has that $a \in \mathbb{Z}$ is a root of $x^{\varphi(m)} - 1$ modulo m if and only if $\gcd(a, m) = 1$. Thus, any reduced residue system mod m is equivalent to the set of all roots of $x^{\varphi(m)} - 1 \pmod m$, up to congruence mod m.

Exercise 2.1.5.

a) Show that for integers e, f, m where $m > 0$, if
$$e \equiv f \pmod{\varphi(m)},$$
then for all integers a coprime to m one has
$$a^e \equiv a^f \pmod m.$$
(*Hint:* Euler's Theorem.)

b) Let p be a prime. Show that for any polynomial $f(x) \in \mathbb{Z}[x]$, there exists $g(x) \in \mathbb{Z}[x]$ of degree less than $p - 1$ such that for all $a \in \mathbb{Z}$ coprime to p one has
$$f(a) \equiv g(a) \pmod p.$$

c) Show that for all integers a with $13 \nmid a$, one has
$$a^{16} + 42a^{12} + 11a^4 + 1 \equiv 4 - a^4 \pmod{13}.$$

d) Give an interesting example of a polynomial $f(x) \in \mathbb{Z}[x]$ that "reduces" to a polynomial $g(x) \in \mathbb{Z}[x]$ of strictly smaller degree modulo a prime p, in the sense of part b). (One point)

Bonus Exercise 2.1.6. Show that for an integer $m \geq 3$, the set $\{0^2, 1^2, \ldots, (m-1)^2\}$ is not a complete residue system modulo m.

Bonus Exercise 2.1.7. Show that if $\{x_1, x_2, \ldots, x_r\}$ is a reduced residue system modulo m, then so is $\{x_1^{-1}, x_2^{-1}, \ldots, x_r^{-1}\}$, where each x_i^{-1} is any integer that is a multiplicative inverse to x_i mod m.

Bonus Exercise 2.1.8.

a) Show that for all integers n and k, if $7 \nmid n$ then $7 \mid (n^{6k} - 1)$.

b) Show that for any integer n, one has $42 \mid (n^7 - n)$.

The following two exercises deal with *primitive roots* and *discrete logarithms.* We will talk about the former in §2.8, and the latter is important to cryptography. Both topics are closely connected.

Bonus Exercise 2.1.9. Given an integer $m \in \mathbb{Z}^+$, we say that a positive integer g is a *primitive root modulo* m if the powers $g^0 = 1, g, g^2, \ldots, g^{\varphi(m)-1}$ form a reduced residue system modulo m.

a) Show that if g is a primitive root modulo m, then for all integers a coprime to m, there exists a unique integer $0 \leq e < \varphi(m)$ such that $g^e \equiv a \pmod m$. In particular, a primitive root modulo m, if it exists, will generate all reduced residue classes mod m.

b) Determine whether a primitive root exists modulo the following numbers.
   i) Modulo 6.
   ii) Modulo 8.
   iii) Modulo 9.
c) Use Bonus Exercise 2.1.7 to show that if g is a primitive root modulo m, then so is $g^{-1} \mod m$.
d) Use part c) to show that for any prime $p > 3$, the product of primitive roots modulo p is congruent to 1 modulo p.

We will explore primitive roots more closely in §2.8.

Bonus Exercise 2.1.10. Given a primitive root g modulo m, we can define a *discrete logarithm modulo m with base g* as follows. As noted in Bonus Exercise 2.1.9, for each integer a there exists a unique integer $0 \leq e < m$ with $g^e \equiv a \pmod{m}$. This e is called the *discrete logarithm of a modulo m,* written as $\log_g(a) := e$. The discrete logarithm depends on the choice of g.

a) Compute the following powers modulo 13, reducing them to representatives between 0 and 12.
   i) 2^3.
   ii) 2^9.
   iii) 2^{11}.
b) Compute the following discrete logarithms modulo 13, with base 2.
   i) $\log_2(6)$.
   ii) $\log_2(5)$.
   iii) $\log_2(7)$.

Computing discrete logarithms mod m for large m can take an extremely long time, even with a computer (though there are ways to get around this if m is a "vulnerable" or unsafe modulus). The computational intractability of the discrete logarithm makes it an important component of many algorithms in public-key cryptography.

Bonus Exercise 2.1.11. This problem explores primes and their connection to numbers of the form $n! + 1$ for integers $n > 0$. Wilson's Theorem gives one such connection.

a) Show that if p is prime, then $(p-1)! + 1$ is a power of p if and only if $p \leq 5$.
b) Using part a) and Wilson's Theorem, show that there are infinitely many integers $n > 0$ such that $n! + 1$ is divisible by at least two distinct primes.

In contrast to part b), it is an open problem to determine whether $n! + 1$ is prime for infinitely many $n \in \mathbb{Z}^+$. Such primes are called *factorial primes.* Some of the known factorial primes are listed on the OEIS: A002981.

Bonus Exercise 2.1.12. Prove that no polynomial $f(x) \in \mathbb{Z}[x]$ of degree greater than one has the property that $f(n)$ is prime for all $n \in \mathbb{Z}^+$. See also the Bunyakovsky Conjecture (Bonus Exercise 1.5.3).

Bonus Exercise 2.1.13. This extends the characterization of primes $p \equiv 1 \pmod 4$ given in Theorem 2.1.13, see Remark 2.1.7. This is similar in flavor to Bonus Exercise 1.3.8. We have shown for odd primes p that

$$p \equiv 1 \pmod 4 \Leftrightarrow \exists\, a, b \in \mathbb{Z} : p = a^2 + b^2 \Leftrightarrow x^2 + 1 \text{ has a root modulo } p.$$

In this exercise, we will give another equivalency, studying how primes p behave in the *Gaussian integer ring*

$$\mathbb{Z}[i] := \{a + bi : a, b \in \mathbb{Z}\}.$$

This ring is generated over $\mathbb{Z}$ by i, which is a root of $x^2 + 1$ over $\mathbb{C}$.

a) Prove the following result.

Theorem. *A prime p satisfies $p \equiv 1 \pmod 4$ if and only if p* splits *in $\mathbb{Z}[i]$, i.e. $p = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$ with $\alpha \neq \beta$.*

b) Show that if a prime p splits in $\mathbb{Z}[i]$, then $x^2 + 1$ splits into distinct linear polynomials modulo p. Show that the converse also holds.

c) Using parts *a)* and *b)*, show that an odd prime p is an *irreducible* element in $\mathbb{Z}[i]$ iff $p \equiv 3 \pmod 4$, iff $x^2 + 1$ is irreducible modulo p.

d) How does $p = 2$ factor in $\mathbb{Z}[i]$? How does $x^2 + 1$ factor modulo 2?

This exercise shows that the factorizing behavior of a prime $p \in \mathbb{Z}$ in $\mathbb{Z}[i]$ is determined by the factorization of $x^2 + 1$ modulo p. This is not a coincidence – in a more general setting, this is a consequence of a theorem of Dedekind and Kummer.

2.2. **Solutions of Congruences.** Our main goals for this (relatively short) section are the following.

- Make concrete the notion of polynomials and solutions modulo m.
- Characterize solutions to linear polynomials modulo m.

As we saw in the previous section, determining solutions to congruences can lead to interesting discoveries about the integers; this was summarized in Remark 2.1.7, where we concluded that a prime p is a sum of two squares if and only if $x^2 + 1$ has roots modulo p. Determining solutions to congruences is a major theme of this chapter, and will have connections to the remaining two chapters on quadratic residues and Diophantine equations.

Let us set forth official definitions for some terms we have used so far.

Definition 2.2.1. For a polynomial $f(x) \in \mathbb{Z}[x]$ and integer $m > 0$, we say that a congruence class $\overline{a}$ modulo m is a **solution, zero** or **root** of $f(x)$ modulo m, if

$$f(a) \equiv 0 \pmod{m}.$$

Remark 2.2.1. By Theorem 2.1.2 ([NZM91, Theorem 2.2]), we know that if $a \equiv b \pmod{m}$ then $f(a) \equiv f(b) \pmod{m}$. Therefore, solutions modulo m are *independent of representatives for congruence classes.* Thus, when referring to solutions modulo m, we will often just refer to the integer representatives of such congruence classes.

Example 2.2.1. The polynomial $f(x) := x^2 + 5$ has two solutions modulo 7: they are $x \equiv 3, 4 \mod 7$, since $3^2 + 5 = 2 \cdot 7$ and $4^2 + 5 = 3 \cdot 7$. Observe that 10 is also a solution mod 7 since $10^2 + 5 = 15 \cdot 7$, but as $10 \equiv 3 \pmod{7}$, both 10 and 3 represent the same root of $f(x) \mod 7$.

Some polynomials can "degenerate" when working modulo m. For example, the polynomial $f(x) := mx$ is essentially the same as $g(x) := 0$ modulo m, since for all $a \in \mathbb{Z}$ we have $f(a) \equiv 0 \pmod{m}$. This leads us to the following definition.

Definition 2.2.2. Let a polynomial $f(x) \in \mathbb{Z}[x]$ be written as

$$f(x) = c_0 + c_1 x + \ldots + c_n x^n$$

with $c_n \neq 0$. Recall that the *degree* of f is $\deg(f) := n$. Given an integer $m > 0$, the **degree of f modulo m,** written as $\deg_m(f)$, is the greatest nonnegative integer $i \leq n$ such that $c_i \not\equiv 0 \pmod{m}$. Similar to the usual convention, if each $c_i \equiv 0 \pmod{m}$ we say that $\deg_m(f) := -\infty$.

This section focuses on characterizing the simplest case for analyzing solutions modulo m: linear polynomials of the form $f(x) := ax - b$, where $a, b \in \mathbb{Z}$. What this will highlight is that under a modulus m, *a polynomial $f(x)$ can have more than $\deg_m(f)$ solutions.*

Theorem 2.2.1 (Linear Congruence Theorem)**.** [NZM91, Theorem 2.17] *Fix integers a and b, and $m > 0$. Then $ax - b$ has a root modulo m – i.e., the congruence*

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

has a solution modulo m – if and only if

$$\gcd(a, m) \mid b.$$

In this case, there are $\gcd(a, m)$ *distinct solutions modulo* m*, given by*

$$c = (a')^{-1} \cdot \frac{b}{\gcd(a, m)} + m' \cdot k,$$

where $0 \leq k < \gcd(a, m)$*,* $a' := \frac{a}{\gcd(a,m)}$*,* $m' := \frac{m}{\gcd(a,m)}$ *and* $(a')^{-1}$ *is an integer representative of the multiplicative inverse of* a' *mod* m'*.*

Remark 2.2.2. Something to note is that when $\gcd(a, m) = 1$, this theorem says there is *exactly* one solution to $ax \equiv b \pmod{m}$, given by

$$c = a^{-1}b \pmod{m}.$$

This is simply the multiplicative inverse of a modulo m; thus, the theorem generalizes this construction.

Proof. Let $g := \gcd(a, m)$, and write $a = a'g$ and $m = m'g$. Our congruence

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

is the same as

$$a'gx \equiv b \pmod{m'g}. \tag{7}$$

If (7) has a solution, then there exists $c \in \mathbb{Z}$ with

$$a'gc = b + m'g,$$

which forces $g \mid b$. In the other direction, if we assume $g \mid b$, then cancellation in (7) gives

$$a'x \equiv \frac{b}{g} \pmod{m'}.$$

Since $\gcd(a', m') = 1$, it follows that a' has a multiplicative inverse modulo m', which we denote by $(a')^{-1} \in \mathbb{Z}$. Thus,

$$c := (a')^{-1} \cdot \frac{b}{g} \pmod{m'}$$

is a solution to (7). This proves the first part.

As shown above, when a solution to $ax \equiv b \pmod{m}$ exists, then any integer c with

$$c \equiv (a')^{-1} \cdot \frac{b}{g} \pmod{m'}$$

gives a solution after multiplying the congruence by $\gcd(a, m)$. However, this congruence for c is the same as

$$c = (a')^{-1} \cdot \frac{b}{g} + m'k \tag{8}$$

for some $k \in \mathbb{Z}$. It is clear that any $k \in \mathbb{Z}$ will produce such a solution c. We claim that there are *only* $\gcd(a, m)$ distinct solutions modulo m, and these can be represented by $0 \leq k < \gcd(a, m)$. To see this, observe that for $k \geq \gcd(a, m)$ we can write

$$k = q \cdot \gcd(a, m) + r$$

for some $q, r \in \mathbb{Z}$ with $0 \leq r < \gcd(a, m)$. Then one has in (8) that

$$\begin{aligned} c &= (a')^{-1} \cdot \frac{b}{g} + m'(gq + r) \\ &= (a')^{-1} \cdot \frac{b}{g} + m'r + mq \\ &\equiv (a')^{-1} \cdot \frac{b}{g} + m'r \pmod{m}. \end{aligned}$$

In particular, our solution c given by

$$c = (a')^{-1} \cdot \frac{b}{g} + m'k$$

with $k \geq \gcd(a, m)$, satisfies

$$c \equiv (a')^{-1} \cdot \frac{b}{g} + m'r \pmod{m}$$

for some $0 \leq r < \gcd(a, m)$. This concludes our proof. □

As we will see later on, studying solutions to non-linear polynomials will get more complicated – but also more interesting!

Remark 2.2.3. Studying solutions to polynomials $f(x) \in \mathbb{Z}[x]$ over various moduli m can give information about integer solutions to $f(x)$ in $\mathbb{Z}$ (or even $\mathbb{Q}$). For example, if $f(x)$ has a solution $a \in \mathbb{Z}$, then from $f(a) = 0$ we have for all $m > 0$ that

$$f(a) \equiv 0 \pmod{m}.$$

Therefore, if there exists $m \in \mathbb{Z}^+$ such that $f(x)$ has *no* roots modulo m, then $f(x)$ has no solutions in $\mathbb{Z}$. This observation can be extremely useful when studying solutions to Diophantine equations, which we will do in Chapter 5 (this includes elliptic curves).

The converse question to the above is also interesting: *if a polynomial $f(x) \in \mathbb{Z}[x]$ has solutions modulo* every *$m > 0$, must it have a solution in $\mathbb{Z}$?* The answer is *not necessarily.* However, there are certain types of polynomials $f(x)$ for which this is always true, and in such a case one says that $f(x)$ satisfies the *local-global principle.*

Exercises. From [NZM91, §2.2], pages 62–63: $\#1 - 6, 8 - 9$.

Exercise 2.2.1. Determine all integer solutions to the following congruences. If no solution exists, explain why.

a) $15x \equiv 4 \pmod{35}$.
b) $137x \equiv 64 \pmod{255}$.
c) $63x \equiv 21 \pmod{69}$.

2.3. **The Chinese Remainder Theorem.** Here are our main goals for this section:

- Prove the Chinese Remainder Theorem (CRT).
- Prove analogs of CRT for residue systems and solutions to polynomials.

In the previous section, we precisely determined when a linear congruence

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

has a solution: when $\gcd(a, m) \mid b$, it has exactly $\gcd(a, m)$ solutions, all of which are given by a formula and uniquely determined modulo m. In this section, we are interested in understanding when *several* linear congruences share a simultaneous solution. An important case of this is determining solutions for a system of *monic* congruences:

$$\begin{aligned} x &\equiv a_1 \pmod{m_1}, \\ x &\equiv a_2 \pmod{m_2}, \\ &\dots \\ x &\equiv a_r \pmod{m_r}. \end{aligned}$$

We know by Theorem 2.2.1 ([NZM91, Theorem 2.17]) that each congruence $x \equiv a_i \pmod{m_i}$ individually has a solution, but does there exist a solution that works for all congruences at the same time? The answer is *yes,* assuming pairwise coprimality conditions on the moduli: this result is called the **Chinese Remainder Theorem,** and is the focus of this section (along with some variants).

Theorem 2.3.1 (Chinese Remainder Theorem (CRT))**.** *Let $m_1, m_2, \dots, m_r > 0$ be pairwise coprime integers. Then for any integers $a_1, a_2, \dots, a_r$, there exists a solution to the following system of congruences:*

$$\begin{aligned} x &\equiv a_1 \pmod{m_1}, \\ x &\equiv a_2 \pmod{m_2}, \\ &\dots \\ x &\equiv a_r \pmod{m_r}. \end{aligned}$$

Furthermore, all solutions are congruent modulo $m := m_1 \cdot m_2 \cdots m_r$: fixing a solution $x_0 \in \mathbb{Z}$, an integer y is also a solution if and only if for each $1 \leq i \leq r$ we have

$$y \equiv x_0 \pmod{m_i},$$

i.e.,

$$y \equiv x_0 \pmod{m}.$$

As we will see, our proof of CRT is constructive, which means we can explicitly create solutions to these systems when they exist.

Proof. Let us define

$$m := m_1 \cdot m_2 \cdots m_r.$$

Observe that for each $1 \leq i \leq r$ one has $\gcd\left(\frac{m}{m_i}, m_i\right) = 1$. Thus $\frac{m}{m_i}$ is invertible modulo m_i, so there exists $b_i \in \mathbb{Z}$ with

$$\frac{m}{m_i} \cdot b_i \equiv 1 \pmod{m_i}. \tag{9}$$

Let us also note that for each $1 \leq j \leq r$ with $j \neq i$, we have $m_j \mid \frac{m}{m_i}$, and so

$$\frac{m}{m_i} \equiv 0 \pmod{m_j}.$$

Let us define the integer

$$x_0 := \sum_{i=1}^{r} \frac{m}{m_i} \cdot b_i \cdot a_i. \tag{10}$$

We check that for each $1 \leq j \leq r$, we have

$$\begin{aligned} x_0 &\equiv \frac{m}{m_i} \cdot b_i \cdot a_i && \pmod{m_j} && \left(\text{since each } \frac{m}{m_i} \equiv 0 \pmod{m_j} \text{ when } i \neq j\right) \\ &\equiv a_j && \pmod{m_j} && \left(\text{since } \frac{m}{m_i} \cdot b_i \equiv 1 \pmod{m_j}\right). \end{aligned}$$

We conclude that x_0 is a solution to the system of linear congruences.

For uniqueness, observe that if y is also a solution, then for each $1 \leq i \leq r$ we have

$$y \equiv x_0 \pmod{m_i}.$$

This implies that

$$y \equiv x_0 \pmod{\operatorname{lcm}(m_1, m_2, \ldots, m_r)}$$

by Theorem 2.1.3.(3) ([NZM91, Theorem 2.3.(3)]). However, since the m_i's are pairwise coprime, we have $\operatorname{lcm}(m_1, m_2, \ldots, m_r) = m_1 \cdot m_2 \cdots m_r$, whence we conclude that y is congruent to x_0 modulo $m_1 \cdot m_2 \cdots m_r$. □

When the hypotheses of CRT are satisfied, we can use Equations (9) and (10) from the proof of CRT to construct explicit solutions.

Example 2.3.1. We wish to find a solution to the system

$$\begin{aligned} x &\equiv 5 \pmod{7}, \\ x &\equiv 7 \pmod{11}, \\ x &\equiv 3 \pmod{13}, \end{aligned}$$

if one exists. Since $7, 11$ and 13 are pairwise coprime, indeed a solution exists by CRT. Here is how we can construct one such solution. Let us set

$$\begin{aligned} a_1 &:= 5, \\ a_2 &:= 7, \\ a_3 &:= 3. \end{aligned}$$

Take

$$\begin{aligned} m_1 &:= 7, \\ m_2 &:= 11, \\ m_3 &:= 13, \end{aligned}$$

and set

$$m := 7 \cdot 11 \cdot 13 = 1001.$$

We break this up into steps.

1. Following Equation (9), we need to solve for b_i in each of the following congruences:
$$\begin{aligned}\frac{m}{m_1}\cdot b_1 &\equiv 1 \pmod 7,\\ \frac{m}{m_2}\cdot b_2 &\equiv 1 \pmod{11},\\ \frac{m}{m_3}\cdot b_3 &\equiv 1 \pmod{13}.\end{aligned}$$
These are
$$\begin{aligned}143\cdot b_1 &\equiv 1 \pmod 7,\\ 91\cdot b_2 &\equiv 1 \pmod{11},\\ 77\cdot b_3 &\equiv 1 \pmod{13},\end{aligned}$$
and reduce to
$$\begin{aligned}3\cdot b_1 &\equiv 1 \pmod 7,\\ 3\cdot b_2 &\equiv 1 \pmod{11},\\ -1\cdot b_3 &\equiv 1 \pmod{13}.\end{aligned}$$
We can then take
$$\begin{aligned}b_1 &:= 5,\\ b_2 &:= 4,\\ b_3 &:= -1.\end{aligned}$$
We note that other congruent choices for each b_i will also work, and will not change the congruence class of our solutions modulo $m = 1001$.
2. Following Equation (10), we can construct an explicit integer solution
$$\begin{aligned}x_0 &:= \sum_{i=1}^{r}\frac{m}{m_i}\cdot b_i\cdot a_i\\ &= 143\cdot 5\cdot 5 + 91\cdot 4\cdot 7 + 77\cdot(-1)\cdot 3\\ &= 5892.\end{aligned}$$
You can double-check that this is a solution to our original system of equations. (One way to remember this formula is that the product $\frac{m}{m_i}\cdot b_i$ is congruent to 1 mod m_i, so that $\frac{m}{m_i}\cdot b_i\cdot a_i$ is the original desired congruence $a_i \mod m_i$.)

From our solution $x_0 = 5892$ modulo $m = 1001$, we can quickly obtain a *least positive solution:* simply divide m into x_0 using the Division Algorithm:
$$x_0 = 5\cdot 1001 + 887.$$
Then
$$x_0 \equiv 887 \pmod m,$$
and therefore also satisfies our original system of congruences (check it!).

The condition that the moduli m_i in CRT are pairwise coprime is necessary to guarantee that a solution exists, as our next example illustrates.

Example 2.3.2. We will show that the following system of congruences does *not* have a solution:

$$\begin{aligned} x &\equiv 29 \pmod{52}, \\ x &\equiv 19 \pmod{72}. \end{aligned}$$

For contradiction, suppose a solution $a \in \mathbb{Z}$ exists. Since 4 divides both 52 and 72, we can then reduce both equations modulo 4 and get

$$\begin{aligned} a &\equiv 1 \pmod{4}, \\ a &\equiv 3 \pmod{4}. \end{aligned}$$

In particular, we have $1 \equiv 3 \pmod{4}$, so that $4 \mid (3-1) = 2$, which is impossible.

Despite the above example, it is still possible for a system of congruences to have a solution even when the moduli are *not* pairwise coprime.

Example 2.3.3. Consider the system

$$\begin{aligned} x &\equiv 3 \pmod{10}, \\ x &\equiv 5 \pmod{84}. \end{aligned}$$

Then CRT does not immediately apply since $\gcd(10, 84) = 2 > 1$. However, we can break these up into more congruences based on the prime factorizations of the moduli. Observe that

$$x \equiv 3 \pmod{10} \Rightarrow \begin{cases} x \equiv 1 \pmod{2}, \\ x \equiv 3 \pmod{5}. \end{cases}$$

Since $84 = 3 \cdot 4 \cdot 7$, we also get

$$x \equiv 5 \pmod{84} \Rightarrow \begin{cases} x \equiv 2 \pmod{3}, \\ x \equiv 1 \pmod{4}, \\ x \equiv 5 \pmod{7}. \end{cases}$$

Since $\gcd(10, 84) = 2$, we should check that the two new congruences modulo powers of 2 do not contradict one another (like in the last example). Since $x \equiv 1 \pmod{4}$ implies $x \equiv 1 \pmod{2}$, these are indeed compatible. Therefore, the new congruences that we will consider are:

$$\begin{aligned} x &\equiv 1 \pmod{4}, \\ x &\equiv 2 \pmod{3}, \\ x &\equiv 3 \pmod{5}, \\ x &\equiv 5 \pmod{7}, \end{aligned}$$

where we chose $x \equiv 1 \pmod{4}$ instead of $x \equiv 1 \pmod{2}$ since it is a stronger congruence requirement.

Since $4, 3, 5$ and 7 are pairwise coprime, a solution can be constructed from CRT using $m = 4 \cdot 3 \cdot 5 \cdot 7 = 420$ and solving

$$\begin{aligned} 105 \cdot b_1 &\equiv 1 \pmod 4, \\ 140 \cdot b_2 &\equiv 1 \pmod 3, \\ 84 \cdot b_3 &\equiv 1 \pmod 5, \\ 60 \cdot b_4 &\equiv 1 \pmod 7. \end{aligned}$$

We can take

$$\begin{aligned} b_1 &:= 1, \\ b_2 &:= 2, \\ b_3 &:= -1, \\ b_4 &:= 2. \end{aligned}$$

This gives the solution

$$\begin{aligned} x_0 &= 105 \cdot 1 \cdot 1 + 140 \cdot 2 \cdot 2 + 84 \cdot (-1) \cdot 3 + 60 \cdot 2 \cdot 5 \\ &= 1013. \end{aligned}$$

A least positive solution is then $x = 173$. Check directly that this solution satisfies both our original system of congruences and the modified one!

CRT finds solutions to systems of (monic) linear congruences, provided that the moduli are pairwise coprime. As illustrated in the previous example, one can turn a single congruence

$$x \equiv a \pmod m$$

into a system of congruences

$$\begin{aligned} x &\equiv a \pmod{p_1^{e_1}}, \\ x &\equiv a \pmod{p_2^{e_2}}, \\ &\dots \\ x &\equiv a \pmod{p_r^{e_r}}, \end{aligned}$$

based on the prime factorization

$$m = \prod_{i=1}^{r} p_i^{e_i}.$$

This idea of factorizing a modulus is useful when trying to understand how to "factorize" other multiplicative structures related to the integers. The first example of this is that reduced residue systems satisfy their own "Chinese Remainder Theorem," providing us a multiplicative formula for Euler's phi function $\varphi(m)$. In the following, for an integer $m > 0$ we let

$$R(m) \subseteq \{0, 1, 2, \dots, m-1\}$$

denote the canonical reduced residue system modulo m (i.e., whose representatives are in $[0, m)$).

Theorem 2.3.2. [NZM91, Theorem 2.19] *Given pairwise coprime integers $m_1, m_2, \ldots, m_r > 0$, define*

$$m := m_1 \cdot m_2 \cdots m_r.$$

Then the natural reduction map

$$R(m) \to R(m_1) \times R(m_2) \times \ldots \times R(m_r),$$
$$a \mapsto (a_1, a_2, \ldots, a_r),$$

where each $1 \leq a_i < m_i$ is congruent to a modulo m_i, is a bijection. *Thus*

$$\varphi(m) = \varphi(m_1) \cdot \varphi(m_2) \cdots \varphi(m_r).$$

In particular, given an integer $n > 1$ with prime factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, one has

$$\varphi(n) = \prod_{i=1}^{r} \varphi(p_i^{e_i}).$$

Proof. Let us call the map above Φ. We must show that Φ is a bijection.

1. Φ is injective. If integers a and b in $R(m)$ satisfy
$$\Phi(a) = \Phi(b),$$
then
$$(a_1, a_2, \ldots, a_r) = (b_1, b_2, \ldots, b_r).$$
This is equivalent to
$$a \equiv b \pmod{m_i}$$
for each $1 \leq i \leq r$. Thus, by Theorem 2.1.3.(3) ([NZM91, Theorem 2.3.(3)]) we deduce that
$$a \equiv b \pmod{\operatorname{lcm}(m_1, m_2, \ldots, m_r)},$$
and since the m_i are pairwise coprime, this is equivalent to
$$a \equiv b \pmod{m}.$$
This forces $a = b$ since $0 \leq a, b < m$. We deduce that Φ is injective.
2. Φ is surjective. This follows from CRT: given a tuple
$$(a_1, a_2, \ldots, a_r) \in R(m_1) \times R(m_2) \times \cdots \times R(m_r),$$
since the moduli $m_1, m_2, \ldots, m_r$ are pairwise coprime, by CRT there exists $a \in \mathbb{Z}$ with
$$\begin{aligned} a &\equiv a_1 \pmod{m_1}, \\ a &\equiv a_2 \pmod{m_2}, \\ &\ldots \\ a &\equiv a_r \pmod{m_r}. \end{aligned}$$
Since each a_i is coprime to m_i and $a \equiv a_i \pmod{m_i}$, it follows that a is coprime to each m_i by Theorem 2.1.4 ([NZM91, Theorem 2.4]). In particular $a \in R(m)$.
We can assume that $0 \leq a < m$. Then we have
$$\Phi(a) = (a_1, a_2, \ldots, a_r).$$

We deduce that Φ is a surjection, and thus a bijection.

Since there exists a bijection between the finite sets $R(m)$ and $R(m_1) \times R(m_2) \times \cdots \times R(m_r)$, their cardinalities are equal:

$$\#R(m) = \prod_{i=1}^{r} \#R(m_i),$$

so that

$$\varphi(m) = \prod_{i=1}^{r} \varphi(m_i)$$

by definition of $\varphi(n)$.

The final part of the theorem follows immediately, since in a prime factorization

$$n = \prod_{i=1}^{r} p_i^{e_i},$$

the $p_i^{e_i}$ are pairwise coprime. □

Remark 2.3.1. One conclusion of this theorem is that for coprime integers $m, n > 0$, one has

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

This says that $\varphi(n)$ is a *multiplicative function.* In particular, the values of $\varphi(n)$ are completely determined by its values on prime powers. In fact, one has a formula for $\varphi(n)$ on prime powers:

$$\varphi(p^e) = p^e - p^{e-1} = p^{e-1}(p-1).$$

This is Exercise 2.1.3. This gives us a complete formula for $\varphi(n)$: for an integer $n > 1$ with prime factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, we have

$$(*) \qquad \varphi(n) = \prod_{i=1}^{r} p_i^{e_i - 1}(p_i - 1).$$

Example 2.3.4. Using the formula for $\varphi(n)$ above, we make the following calculations.

- $\varphi(18) = \varphi(2 \cdot 3^2) = \varphi(2) \cdot \varphi(3^2) = 1 \cdot 3(3-1) = 6$. Thus, there are six integers in $[1, 18]$ that are coprime to 18.
- $\varphi(100) = \varphi(2^2 \cdot 5^2) = \varphi(2^2) \cdot \varphi(5^2) = 2(2-1) \cdot 5(5-1) = 40$. Thus, there are forty integers in $[1, 100]$ that are coprime to 100.
- $\varphi(210) = \varphi(2 \cdot 3 \cdot 5 \cdot 7) = \varphi(2) \cdot \varphi(3) \cdot \varphi(5) \cdot \varphi(7) = 1 \cdot 2 \cdot 4 \cdot 6 = 48$. Thus, there are forty-eight integers in $[1, 210]$ that are coprime to 210.

Remark 2.3.2. The bijection in Theorem 2.3.2 ([NZM91, Theorem 2.19]) comes from an identical bijection

$$\Phi \colon C(m) \to C(m_1) \times C(m_2) \times \ldots \times C(m_r),$$
$$a \mapsto (a_1, a_2, \ldots, a_r),$$

where $C(n) := \{0, 1, 2, \ldots, n-1\}$ denotes the canonical complete residue system modulo n, the m_i are positive and pairwise coprime, and $m := m_1 \cdot m_2 \cdots m_r$. However, less

new information is gleamed from this version: for example, a size comparison shows that

$$\#C(m) = \prod_{i=1}^{r} \#C(m_i),$$

which is just

$$m = \prod_{i=1}^{r} m_i.$$

In this last part of the section, we will see how CRT helps us study solutions/roots/zeroes of polynomials modulo m.

Definition 2.3.1. Given a polynomial $f(x) \in \mathbb{Z}[x]$, for an integer $m > 0$ we define the **zero set of f modulo m** as

$$V(f, m) := \{0 \leq a < m : f(a) \equiv 0 \pmod{m}\}.$$

We let $\varphi_f(m)$ denote the size of $V(f, m)$.

There is a strong connection between our solution-counting function $\varphi_f(n)$ and Euler's phi function $\varphi(n)$.

Theorem 2.3.3 (CRT for solutions)**.** [NZM91, Theorem 2.20] *Given a polynomial $f(x) \in \mathbb{Z}[x]$ and pairwise coprime integers $m_1, m_2, \ldots, m_r > 0$, writing $m := m_1 \cdot m_2 \cdots m_r$, the natural reduction map*

$$\begin{aligned} V(f, m) &\to V(f, m_1) \times V(f, m_2) \times \ldots \times V(f, m_r), \\ a &\mapsto (a_1, a_2, \ldots, a_r) \end{aligned}$$

is a bijection. *Thus*

$$\varphi_f(m) = \varphi_f(m_1) \cdot \varphi_f(m_2) \cdots \varphi_f(m_r).$$

In particular, given an integer $n > 1$ with prime factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, we have

$$\varphi_f(n) = \prod_{i=1}^{r} \varphi_f(p_i^{e_i}).$$

The key idea for proving that this map is a bijection is that if a_i is a root of $f(x)$ modulo m_i, then any solution $a \in \mathbb{Z}$ to the simultaneous congruences

$$\begin{aligned} x &\equiv a_1 \pmod{m_1}, \\ x &\equiv a_2 \pmod{m_2}, \\ &\cdots \\ x &\equiv a_r \pmod{m_r} \end{aligned}$$

via CRT is also a root of $f(x)$ modulo $m := m_1 \cdot m_2 \cdots m_r$.

Proof. This proof will be similar to the one for factorizing residue systems. In fact, noting that $V(f, m) \subseteq C(m)$, this map is simply a restriction of the reduction map

$$\begin{aligned} \Phi \colon C(m) &\to C(m_1) \times C(m_2) \times \ldots \times C(m_r), \\ a &\mapsto (a_1, a_2, \ldots, a_r). \end{aligned}$$

To see that the codomain is correct, note that if

$$f(a) \equiv 0 \pmod{m},$$

then for each $1 \leq i \leq r$ we have

$$f(a) \equiv 0 \pmod{m_i}.$$

Thus Φ does take $V(f,m)$ into $V(f,m_1) \times V(f,m_2) \times \ldots V(f,m_r)$.

We must show that Φ is a bijection.

1. Φ is an injection on $V_f(m)$. This is true since Φ is injective on the larger set $C(m)$.
2. Φ surjects onto $V(f,m_1) \times V(f,m_2) \times \ldots V(f,m_r)$. Start with a tuple

$$(a_1, a_2, \ldots, a_r) \in V(f,m_1) \times V(f,m_2) \times \ldots V(f,m_r).$$

For each $1 \leq i \leq r$, we have

$$a_i \in V(f,m_i),$$

so that

$$f(a_i) \equiv 0 \pmod{m_i}. \tag{11}$$

By CRT there exists $a \in C(m)$ with

$$\begin{aligned} a &\equiv a_1 \pmod{m_1}, \\ a &\equiv a_2 \pmod{m_2}, \\ &\ldots \\ a &\equiv a_r \pmod{m_r}. \end{aligned}$$

For each $1 \leq i \leq r$, since

$$a \equiv a_i \pmod{m_i},$$

we have by Theorem 2.1.2 ([NZM91, Theorem 2.2]) that

$$f(a) \equiv f(a_i) \pmod{m_i},$$

which by (11) implies that

$$f(a) \equiv 0 \pmod{m_i}.$$

We deduce that

$$f(a) \equiv 0 \pmod{\operatorname{lcm}(m_1, m_2 \ldots, m_r)},$$

and thus

$$f(a) \equiv 0 \pmod{m}.$$

It follows that $a \in V(f,m)$, whence we deduce that Φ is surjective.

We conclude that Φ is a bijection, and so

$$\#V(f,m) = \prod_{i=1}^{r} \#V(f,m_i),$$

i.e.,

$$\varphi_f(m) = \prod_{i=1}^{r} \varphi_f(m_i).$$

Therefore, given an integer $n > 1$ with prime factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, one has

$$\varphi_f(n) = \prod_{i=1}^{r} \varphi_f(p_i^{e_i}). \qquad \square$$

Remark 2.3.3. The above shows that $\varphi_f(n)$ is a multiplicative function, much like $\varphi(n)$. However, unlike Euler's totient function, it is possible for $\varphi_f(n)$ to be zero. For example, consider $f(x) := x^2 + 1$ and $n := 7$ from Example 2.1.6.

Theorem 2.3.3 ([NZM91, Theorem 2.20]) shows that finding roots of a polynomial $f(x) \in \mathbb{Z}[x]$ modulo m amounts to finding roots modulo the prime-power divisors of m, and then applying CRT to all possible pairings of the roots from each prime power. Thus, the fundamental problem of computing solutions modulo m reduces to the prime power case. We will study solutions modulo p^e in the next section §2.6.

Example 2.3.5. Consider $f(x) := x^2 + x + 1$. We would like to determine its roots modulo 21. To do this, we factorize $21 = 3 \cdot 7$ and looks for roots of $f(x)$ modulo 3 and modulo 7. Once we have these (if they exist), we then apply CRT to each pair of solutions whose entries are mod 3 and mod 7, respectively.

We check by hand that

$$f(x) \equiv 0 \pmod 3 \Leftrightarrow x \equiv 1 \pmod 3,$$

and

$$f(x) \equiv 0 \pmod 7 \Leftrightarrow x \equiv 2, 4 \pmod 7.$$

Pairing up these solutions, we have two cases to apply CRT towards.

1. Consider the system

$$\begin{aligned} x &\equiv 1 \pmod 3, \\ x &\equiv 2 \pmod 7. \end{aligned}$$

Then solving

$$\begin{aligned} 7b_1 &\equiv 1 \pmod 3, \\ 3b_2 &\equiv 1 \pmod 7, \end{aligned}$$

we can take $b_1 := 1$ and $b_2 := 5$, and construct the solution

$$\begin{aligned} x_0 :&= 7 \cdot 1 \cdot 1 + 3 \cdot 5 \cdot 2 \\ &= 37 \\ &\equiv 16 \pmod{21}. \end{aligned}$$

We can check directly that $x_0 := 16$ is a solution to $f(x) \mod 21$:

$$f(16) = 16^2 + 16 + 1 = 273 = 13 \cdot 21,$$

and so $f(16) \equiv 0 \pmod{21}$.

2. Consider the system

$$x \equiv 1 \pmod 3,$$
$$x \equiv 4 \pmod 7.$$

Then the same work from the first part applies, since the b_i's remain unchanged (they do not depend on the a_i). We can take $b_1 := 1$ and $b_2 := 5$, and thus have a solution

$$\begin{aligned} x_0 &:= 7 \cdot 1 \cdot 1 + 3 \cdot 5 \cdot 4 \qquad \text{(in purple is the only change, which is } a_2\text{)} \\ &= 67 \\ &\equiv 4 \pmod{21}. \end{aligned}$$

We double-check that $x_0 := 4$ is a solution to $f(x) \mod 21$:

$$f(4) = 4^2 + 4 + 1 = 21,$$

and so $f(4) \equiv 0 \pmod{21}$.

We conclude that the two solutions to $f(x)$ modulo 21 are 4 and 16 (mod 21). As it turns out, this gives an equivalence of polynomials modulo $m = 21$:

$$x^2 + x + 1 \equiv (x - 4)(x - 4^2) \pmod{21}.$$

Exercises. From [NZM91, §2.3], pages 71–72: $\#1 - 4, 7 - 8, 10 - 15, 17, 32$.

Exercise 2.3.1. For each part, use the Chinese Remainder Theorem to determine all integers x which satisfy the simultaneous congruences. If no solution exists, then prove it.

a) $x \equiv 1 \pmod 3$ and $x \equiv 4 \pmod 7$.
b) $x \equiv 0 \pmod 4$, $x \equiv 2 \pmod 5$ and $x \equiv 3 \pmod 9$.
c) $8x \equiv 1 \pmod 6$ and $7x \equiv 10 \pmod{15}$.

Exercise 2.3.2. Prove the following multilinear version of Dirichlet's Theorem on Primes in Arithmetic Progressions.

Theorem. *Given positive integers $m_1, m_2, \ldots, m_r$ which are pairwise coprime, for any integers $a_1, a_2, \ldots, a_r$ where each a_i is coprime to m_i, there exist infinitely many primes p such that for all $1 \leq i \leq r$ one has*

$$p \equiv a_i \pmod{m_i}.$$

(*Hint:* Use the Chinese Remainder Theorem and Dirichlet's Theorem on Primes in Arithmetic Progessions (Exercise 1.3.5) to prove this.)

Exercise 2.3.3.

a) Determine all integers $n \in \mathbb{Z}^+$ for which $\varphi(n)$ is odd.
b) Show that if every prime p which divides m also divides n, then $\varphi(mn) = m\varphi(n)$.

Exercise 2.3.4.

a) Prove that the polynomial $f(x) := x^2 + 5x + 24$ has no integer solutions.

b) Using the Chinese Remainder Theorem, prove that $f(x)$ from part a) has solutions modulo 36; **write these solutions as least positive integers mod 36.** Then check that your solutions work by writing their values under $f(x)$ as a multiple of 36.

Bonus Exercise 2.3.5. Prove that for integers $m, n \in \mathbb{Z}^+$ with $m > 1$, if

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

then $m = 2$ and n is odd. Characterize the set of positive integers n satisfying

$$\varphi(2n) = \varphi(n).$$

Bonus Exercise 2.3.6. Show that for a fixed integer $n \in \mathbb{Z}^+$, the equation

$$\varphi(x) = n$$

has a finite number of solutions.

Bonus Exercise 2.3.7. This exercise explores which positive integers are not in the image of Euler's totient function $\varphi(x)$.

a) Show that for an odd integer $n > 0$, the equation

$$\varphi(x) = n$$

has a solution if and only if $n = 1$. Thus, the image $\varphi(\mathbb{Z}^+)$ contains no odd $n \geq 3$.

b) Show that there does not exist a solution to $\varphi(x) = 14$.

c) Show that 14 is the *smallest* positive even integer not in $\varphi(\mathbb{Z}^+)$. Then determine the next smallest such integer.

2.6. **Prime Power Moduli (Hensel's Lemma).** We saw in the last section that studying solutions to polynomials under a modulus reduces to studying solutions under a prime-power modulus; this was Theorem 2.3.3 ([NZM91, Theorem 2.20]). In this section, we will learn about a technique to studying solutions modulo prime powers called *Hensel's Lemma,* which is about lifting mod p^e solutions to mod p^{e+1} solutions.

Definition 2.6.1. Given integers $d, m \in \mathbb{Z}^+$ with $d \mid m$, and congruence classes $a \bmod d$ and $b \mod m$, we say that $b \pmod m$ is a **lift** of $a \pmod d$ if

$$b \equiv a \pmod d.$$

Theorem 2.6.1 (Hensel's Lemma)**.** [NZM91, Theorem 2.23] *Consider a polynomial $f(x) \in \mathbb{Z}[x]$, and let p^e be a prime power. Assume that the following hold:*

1. *a is a* root *of $f(x)$ modulo p^e, i.e.,*
$$f(a) \equiv 0 \pmod{p^e}.$$
2. *One has*
$$f'(a) \not\equiv 0 \pmod p$$
(i.e. a is not a repeated root *of $f(x)$ modulo p).*

Then there exists a lift b of a such that b is a root of $f(x)$ mod p^{e+1}, i.e.,

$$f(b) \equiv 0 \pmod{p^{e+1}}.$$

We can take

$$b := a + tp^e$$

where

$$t := -f'(a)^{-1} \cdot \frac{f(a)}{p^e} \mod p,$$

with $f'(a)^{-1}$ denoting the multiplicative inverse of $f'(a)$ mod p. (We have that t is unique mod p.)

Remark 2.6.1. It is worth double-checking that Hensel's Lemma is necessary, i.e., that a solution to a polynomial modulo p^e is not necessarily a solution modulo p^{e+1}. One example is to take $f(x) := x^2 + 1$, and observe that $a = 1$ is a root of $f(x)$ mod 2, but not of $f(x)$ mod 4. (You can also check that $f'(x)$ is identically zero mod 2.)

Remark 2.6.2. In practice, when searching for roots of $f(x)$ modulo m, if for a prime $p \mid m$ we find a root a for $f(x)$ modulo p and have $f'(a) \not\equiv 0 \pmod p$, then we can apply Hensel's Lemma repeatedly to lift a until we have a solution modulo p^e, where $p^e \parallel m$. Recall that once we have solutions modulo every prime power divisor of m, we can then apply CRT and calculate all solutions mod m.

Before proving Hensel's Lemma, we will need a proper lemma for it; this is from [NZM91, §1.4].

Lemma 2.6.2. [NZM91, Theorem 1.21] *The product of k consecutive integers is a multiple of $k!$.*

Proof of Hensel's Lemma. Given a solution

$$f(a) \equiv 0 \pmod{p^e}$$

with

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

our primary goal is to show that for

$$b := a + tp^e$$

where

$$t = -f'(a)^{-1} \cdot \frac{f(a)}{p^e} \mod p, \tag{12}$$

we have that b is a solution to $f(x) \mod p^{e+1}$. Note that the multiplicative inverse of $f'(a)$ exists mod p since $p \nmid f'(a)$. We will first prove that for any integer of the form $a + tp^e$ with $t \in \mathbb{Z}$, there is a simple expression for $f(a + tp^e)$ modulo p^{e+1}:

$$f(a + tp^e) \equiv f(a) + f'(a) \cdot tp^e \pmod{p^{e+1}}. \tag{13}$$

We will then show that

$$f(a) + f'(a) \cdot tp^e \equiv 0 \pmod{p^{e+1}}$$

for our specific t in Equation (12).

To prove (13), we will use some Calculus II. Recall that for any function $f(x)$, for each integer $n \geq 0$ the *n'th Taylor polynomial of* $f(x)$ *at* $x = a$ is

$$\begin{aligned} f(x) &= \sum_{k=0}^{n} \frac{f^{(k)}(x)(x-a)^k}{k!} \\ &= f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2} + \ldots + \frac{f^{(n)}(a)(x-a)^n}{n!}. \end{aligned}$$

This Taylor polynomial gives us a way to approximate $f(x)$ near $x = a$: the larger n is, the better an approximation the Taylor polynomial is of $f(x)$ near $x = a$; and taking $n \to \infty$ gives the *Taylor series of* $f(x)$ *at* $x = a$ which gives exact values for $f(x)$ near $x = a$. In our application $f(x)$ is a polynomial, and so the Taylor series is simply the n'th Taylor polynomial where $n := \deg(f)$, which equals $f(x)$ itself – try to prove this yourself!

With the above in mind, let us evaluate $f(x)$ at $x := a + tp^e$ via its n'th Taylor polynomial expansion at $x = a$, noting that $(a + tp^e) - a = tp^e$:

$$f(a + tp^e) = f(a) + f'(a) \cdot tp^e + \frac{f''(a) \cdot t^2 p^{2e}}{2} + \ldots + \frac{f^{(n)}(a) \cdot t^n p^{ne}}{n!}. \tag{14}$$

To get (13), we claim that all but the first two terms of $f(a + tp^e)$ are necessarily multiples of p^{e+1}. Such a term is of the form

$$\frac{f^{(i)}(a) \cdot t^i p^{ie}}{i!} = t^i p^{ie} \cdot \frac{f^{(i)}(a)}{i!}$$

for $i \geq 2$, so it suffices to show that

$$\frac{f^{(i)}(a)}{i!}$$

is an integer. To see this, note that if $c_k x^k$ is a monomial term in the sum $f(x)$ with $k \geq i$, then the corresponding term in $f^{(i)}(a)$ is

$$c_k \cdot k(k-1)(k-2)\cdots(k-(i-1)) \cdot a^{k-i}.$$

Since $k, k-1, k-2, \ldots, k-(i-1)$ are consecutive integers, Lemma 2.6.2 ([NZM91, Theorem 1.21]) implies that $i!$ divides this corresponding term. Therefore $i!$ divides *all* terms in $f^{(i)}(a)$. This proves our claim that $\frac{f^{(i)}(a)}{i!}$ is an integer, and so we deduce that for $i \geq 2$

$$p^{e+1} \mid \frac{f^{(i)}(a) \cdot t^i p^{ie}}{i!}.$$

We thus conclude from Equation (14) that (13) holds for any $t \in \mathbb{Z}$.

Next, we will show that the right-hand side term in (13) is congruent to 0 modulo p^{e+1} for the choice of t in (12). Observe that this choice of t satisfies

$$f'(a) \cdot t \equiv -\frac{f(a)}{p^e} \pmod{p}.$$

We can multiply this expression by p^e and get

$$f'(a) \cdot tp^e \equiv -f(a) \pmod{p^{e+1}},$$

i.e.,

$$f(a) + f'(a) \cdot tp^e \equiv 0 \pmod{p^{e+1}}.$$

Then Equation (13) for $f(a + tp^e)$ becomes

$$f(a + tp^e) \equiv 0 \pmod{p^{e+1}},$$

as desired. We conclude that $a + tp^e$ for t in (12) is a solution to $f(x) \bmod p^{e+1}$, and that $t \in \mathbb{Z}$ is unique mod p. Since $a + tp^e \equiv a \pmod{p^e}$, we also conclude that $a + tp^e \bmod p^{e+1}$ is a lift of $a \mod p^e$. □

Hensel's Lemma shows that if $f(x)$ has a root $a \bmod p^e$, then with the extra assumption that $f'(a) \not\equiv 0 \pmod{p}$, one can lift it to a solution mod p^{e+1}. This latter condition is important enough to give a name to.

Definition 2.6.2. For a polynomial $f(x) \in \mathbb{Z}[x]$ and a prime power p^e, say that a $a \in \mathbb{Z}$ is a **nonsingular** root of $f(x) \bmod p^e$ if $f(a) \equiv 0 \pmod{p^e}$ and $f'(a) \not\equiv 0 \pmod{p}$. If a is a root of $f(x) \mod p^e$ but $f'(a) \equiv 0 \pmod{p}$, then call a a **singular** root of $f(x) \bmod p^e$.

Remark 2.6.3. Recast differently, Hensel's Lemma shows that a *nonsingular* root of $f(x) \bmod p^e$ constructively lifts to a root of $f(x) \bmod p^{e+1}$. Nonetheless, for singular roots there will exist either 0 or p lifts to a solution mod p^{e+1}. The exercises will cover this a bit.

Hensel's Lemma can be applied repeatedly to continually lift roots modulo powers of p. For example, if a_1 is a nonsingular root of $f(x)$ mod p, then a_1 lifts to a root a_2 of $f(x)$ mod p^2. Since

$$f'(a_2) \equiv f'(a) \not\equiv 0 \pmod{p},$$

we know that a_2 is a nonsingular for $f(x)$ mod p^2, and so Hensel's Lemma implies a_2 lifts to a root a_3 of $f(x)$ mod p^3, and so on. We thus have a sequence

$$a_1, a_2, a_3, \ldots$$

of roots for $f(x)$ modulo the powers $p, p^2, p^3, \ldots$, respectively. Furthermore, these roots can be defined recursively: in the homework, you will show that for each $k \geq 1$ one has the formula

$$(*) \qquad a_{k+1} \equiv a_k - f'(a_1)^{-1} \cdot f(a_k) \pmod{p^{k+1}},$$

where $f'(a_1)^{-1}$ represents the inverse of $f'(a_1)$ mod p (which just needs to be computed once). This gives an easy formula for constructing lifts of solutions using Hensel's Lemma. It is worth noting that this construction is analogous to Newton's Method for finding roots of a real differentiable function.

Remark 2.6.4. We saw above that a nonsingular solution a_1 to $f(x)$ mod p lifts to any desired modulus p^e, producing a sequence of integers $(a_1, a_2, a_3, \ldots)$ where for each $k \leq \ell$ one has

$$a_\ell \equiv a_k \pmod{p^k}.$$

Such sequences are called *p-adic integers.* The set (ring) of p-adic integers is written as $\mathbb{Z}_p$. The p-adic integers can be used to analyze integer solutions to polynomials or Diophantine equations "arbitrarily locally." In particular, if $f(x)$ has a nonsingular root a_1 modulo p, then the corresponding p-adic integer $(a_1, a_2, a_3, \ldots)$ is a "root" of $f(x)$ in $\mathbb{Z}_p$.

Example 2.6.1. Let us illustrate Hensel's Lemma with an example. Consider the polynomial $f(x) := x^3 - 2$. Note that $f(x)$ has no integer solutions, since $\sqrt[3]{2} \notin \mathbb{Z}$. However, one checks directly that it has exactly one solution mod 5, which is $a_1 := 3$. We calculate $f'(x) = 3x^2$, and so $f'(3) = 27 \not\equiv 0 \pmod 5$, which means a_1 is a nonsingular root of $f(x) \mod 5$. Therefore, by Hensel's Lemma a_1 can be lifted to a mod-25 solution

$$a_2 := a_1 + t_1 p,$$

where

$$t_1 := -f'(3)^{-1} \cdot \frac{f(3)}{5} \mod 5.$$

We check that

$$\begin{aligned} f'(3)^{-1} &= 27^{-1} \\ &\equiv 2^{-1} \pmod 5 \\ &\equiv 3 \pmod 5. \end{aligned}$$

We also see that

$$\frac{f(3)}{5} = \frac{25}{5} = 5 \equiv 0 \pmod 5.$$

Thus, we can take

$$t_1 := -3 \cdot 0 = 0,$$

so our solution lift is

$$a_2 := a_1 = 3.$$

We can verify that $f(a_2) = f(3) \equiv 0 \pmod{25}$, since $f(3) = 25$.

Let us apply Hensel's Lemma again to lift a_2 to a solution modulo $p^3 = 125$. This is possible because a_2 is a nonsingular root of $f(x) \bmod 25$, since

$$\begin{aligned} f(a_2) &\equiv f(a_1) \pmod 5 \\ &\equiv 3 \pmod 5 \\ &\not\equiv 0 \pmod 5. \end{aligned}$$

Such a lift is

$$a_3 := a_2 + t_2 p^2,$$

where

$$t_2 := -f'(3)^{-1} \cdot \frac{f(3)}{25} \mod 5$$

(in purple is the only difference from the previous t). We still have $f'(3)^{-1} \equiv 3 \pmod 5$, but this time we calculate $\frac{f(3)}{25} \mod 5 = 1 \mod 5$. We thus take

$$t_2 := -3 \cdot 1 = -3,$$

and conclude that

$$a_3 := 3 + (-3) \cdot 25 = -72$$

is a solution lift of 3 to $f(x) \bmod 125$. We can verify this:

$$f(-72) = -373250 = 125 \cdot (-2986).$$

Let us lift it one more time to $f(x) \mod 5^4 = 625$: we just computed $\frac{f(-72)}{125} = -2986$, and so for the lift

$$a_4 := -72 + 125 t_3$$

we take

$$\begin{aligned} t_3 :&= -f'(-72)^{-1} \cdot \frac{f(-72)}{625} \\ &= -3 \cdot -2986 \\ &\equiv -3 \cdot -1 \pmod 5 \\ &= 3 \pmod 5. \end{aligned}$$

We thus conclude that

$$a_4 := -72 + 375 = 303$$

is a root of $f(x) \mod 625$, which can be verified directly:

$$f(303) = 27818125 = 625 \cdot 44509.$$

We therefore have a sequence of solutions

$$(3 \pmod 5, 3 \pmod{25}, -72 \pmod{125}, 303 \pmod{625}, \ldots)$$

(or just $(3, 3, -72, 303, \ldots)$). By Remark 2.6.4, this sequence is a p-adic integer.

Exercises. From [NZM91, §2.6], page 91: $\#1-3, 5-7$.

Exercise 2.6.1. Fix a polynomial $f(x) \in \mathbb{Z}[x]$ and a prime p. Suppose that $f(x)$ has a nonsingular root a_1 modulo p. Then by Hensel's Lemma, we can lift a_1 repeatedly to obtain solutions a_{k+1} to $f(x)$ mod p^{k+1} for each $k \geq 1$. These solutions can be described by the formula

$$(*) \qquad a_{k+1} \equiv a_k - f'(a_1)^{-1} \cdot f(a_k) \pmod{p^{k+1}},$$

where $f'(a_1)^{-1}$ represents the multiplicative inverse of $f'(a_1)$ modulo p.

a) Using the formula $(*)$, verify that a_{k+1} is a lift of a_k. Also verify that each $f'(a_k)^{-1}$ is congruent to $f'(a_1)^{-1}$ modulo p.
b) Use the formula for solution lifts from Hensel's Lemma in class to prove the formula $(*)$.
c) Suppose that a is instead a *singular root* modulo p, i.e.,

$$f(a) \equiv 0 \pmod{p}$$

but

$$f'(a) \equiv 0 \pmod{p}.$$

Prove that there are either 0 or p lifts of a modulo p^2. (*Hint:* revisit our proof of Hensel's Lemma and see what happens when $f'(a)$ is not coprime to p.)

Exercise 2.6.2. Using Hensel's Lemma by hand, solve the congruence

$$x^4 + 2 \equiv 0 \pmod{27};$$

write your solution(s) as a least positive integer mod 27**.** Then show directly that your solution(s) works by writing its evaluation under $f(x)$ as a multiple of 27.

Exercise 2.6.3.

a) Create a `Sage` function which takes as input $(f(x), p)$, where $f(x)$ is a polynomial in $\mathbb{Z}[x]$ and p is a prime, and outputs a list of solutions for $f(x)$ modulo p, via plugging each integer in $[0, p-1]$ directly into $f(x)$ and reducing mod p. If no such solutions exist, have it return a message saying as much.

   Run output for this function for $f(x) := x^2 + 1$, over all primes $p \leq 101$.
b) What patterns do you notice in part a)? (One point)
c) Using your function from part a), create another function which takes as input $(f(x), p^k)$ where $f(x) \in \mathbb{Z}[x]$ and p^k is a prime power, and applies Hensel's Lemma to $f(x)$ modulo p^k, via the formula $(*)$ given in Exercise 2.6.1. **Have your code output solutions as least positive integers mod** p^k**.**

   (*Hint:* You may find `ZZ.valuation()` useful for returning the exponent k in p^k, as well as `inverse_mod(a,m)` for computing multiplicative inverses and `derivative()` for computing derivatives. Note, however, that some object types might mismatch, in which case you may have to turn several objects into integers to have them interact; `Integer()` can be useful for this.)

   Use this code to output solutions to $f(x) := x^2 + 1$ modulo p^3 for all primes $p \leq 101$.

Bonus Exercise 2.6.4. Extend Exercise 2.6.3 via the Chinese Remainder Theorem: program a function in `Sage` where given polynomial $f(x) \in \mathbb{Z}[x]$ and integer $m > 0$, the function takes as input $(f(x), m)$ and outputs the list of all solutions to $f(x)$ modulo m.

2.7. **Prime Modulus.** By CRT and Hensel's Lemma, the problem of finding solutions to a polynomial congruence

$$f(x) \equiv 0 \pmod{m}$$

is reduced to finding solutions

$$f(x) \equiv 0 \pmod{p}$$

for all primes $p \mid m$. However, there is no systematic method for finding roots of an arbitrary polynomial $f(x)$ modulo p beyond naïvely evaluating $f(x)$ at integers $0 \leq k < p$ and reducing mod p. Nonetheless, some polynomials have interesting characterizations for their roots modulo primes. We saw this for $f(x) := x^2 + 1$ in Remark 2.1.7: for odd primes p, one has that:

$$\begin{aligned} & x^2 + 1 \equiv 0 \pmod{p} \text{ has a solution} \\ \Leftrightarrow\ & p \equiv 1 \pmod{4} \\ \Leftrightarrow\ & p = a^2 + b^2 \text{ for some } a, b \in \mathbb{Z} \\ \Leftrightarrow\ & p \textit{ splits} \text{ in } \mathbb{Z}[i] \qquad \text{(algebraic number theory result).} \end{aligned}$$

In Section 2.8 (the final section of Chapter 2 for us) will focus on understanding when an integer polynomial of the form $x^n - a$ has roots mod p. Following that, Chapter 3 will focus on specifically characterizing solutions to polynomials of the form $x^2 - a$ mod p.

Polynomial roots modulo m can have counterintuitive properties. For example, by the *Fundamental Theorem of Algebra* any degree $n \geq 0$ polynomial $f(x) \in \mathbb{Z}[x]$ has **at most** n solutions in $\mathbb{Z}$.[5] However, over a modulus m it is possible for $f(x)$ to have more roots than its degree. For example, the polynomial $f(x) := 5x^2 + 10$ has roots $x = 0, 1, 2, 3, 4$ modulo 5, since $f(x)$ is equivalent to the zero polynomial modulo 5. As another example $g(x) := x^2 + 7x + 2$ has four solutions modulo 10, which are $x = 3, 4, 8, 9$.

However, over a prime modulus p a polynomial $f(x) \in \mathbb{Z}[x]$ of degree n cannot have more than n roots mod p if its mod p degree $\deg_p(f)$ is nonnegative.[6]

Theorem 2.7.1. [NZM91, Theorem 2.26] *For a polynomial $f(x) \in \mathbb{Z}[x]$ and a prime p, if $\deg_p(f) = n \geq 0$ then $f(x)$ has at most n roots modulo p.*

Therefore, after reducing $f(x)$ modulo p, the resulting polynomial

$$\overline{f(x)} := f(x) \mod p$$

will have the "correct" maximum amount of roots modulo p. In a more technical language, this is because the set of congruence classes modulo p is a *field.* We will touch on fields later in the chapter (§2.11).

[5]More specifically, a polynomial $f(x) \in \mathbb{Z}[x]$ of degree $n \geq 0$ has **exactly** n solutions in the *complex numbers* $\mathbb{C}$.

[6]Recall that $\deg_p(f)$ is the largest index i such that the coefficient of x^i in $f(x)$ is nonzero mod p. We take $\deg_p(f) := -\infty$ if all coefficients of $f(x)$ are multiples of p.

2.10. **Number Theory From an Algebraic Viewpoint.** For the next two sections, we will develop some basic *abstract algebra* to recontextualize what we have learned about congruences. One benefit to this is that we will have alternative proofs to classic elementary number theory results, such as Euler's Theorem (and Fermat's Little Theorem). It will also clarify our discussions about primitive roots, quadratic residues and elliptic curves in this course.

Here are our main goals for this section:

- Define what a *group* is.
- Revisit congruences in terms of group theory.
- Define *homomorphisms* and *isomorphisms* between groups.

Abstract algebra is the study of algebraic structures, which are (usually) sets with "additional structure" that give a way to "combine" elements. These structures are often generalizations of the usual addition and multiplication of numbers. Algebraic structures include *groups, rings* and *fields* – each of which we will touch on in these next two sections. However, our main focus will be in understanding the basic theory of groups, culminating in a proof of an important case of *Lagrange's Theorem.*

Much of elementary number theory can be recast in terms of abstract algebra. For example, a complete residue system modulo m is an example of a group with "reduced addition," where there is a sensible way of adding two congruence representatives. We also have that any reduced residue system modulo m is a group under multiplication. Taking a step back, we have that the set of integers $\mathbb{Z}$ is a group under addition, along with $\mathbb{Q}$ and $\mathbb{R}$. However, the sets $\mathbb{Q} \setminus \{0\}$ and $\mathbb{R} \setminus \{0\}$ are also groups under multiplication. It is very easy to come up with groups that we encounter regularly in mathematics.

Let us properly define a group. Recall that for a set X, a *binary operation* on X is a map

$$X \times X \to X.$$

Definition 2.10.1. A set G is a **group** if there exists a binary operation

$$\oplus \colon G \times G \to G$$

with the following properties.

1. G has an **identity** element $e := e_G \in G$, where for all $g \in G$ one has
$$e \oplus g = g \oplus e = g.$$
2. G is closed under **inverses:** for any $g \in G$, there exists $h \in G$ with
$$g \oplus h = h \oplus g = e.$$
One often writes g^{-1} for h.
3. $\oplus$ is **associative:** for $g, h, k \in G$, one has
$$(g \oplus h) \oplus k = g \oplus (h \oplus k).$$
One often writes this as $g \oplus h \oplus k$.

Such an operation $\oplus$ is called a **group law on G.** We sometimes write $(G, \oplus)$ instead of G to emphasize the group law.

Definition 2.10.2. Let $(G, \oplus)$ be a group.

- If for all $g, h \in G$ one has
$$g \oplus h = h \oplus g,$$
then we say that G is an **abelian** or **commutative** group.
- If G is a finite set, then we call G a **finite group.** We call the size of G its **order,** and write it as $|G|$ or $\#G$.

Remark 2.10.1. For this course, the groups that we study are almost always finite abelian groups, which simplifies our study.

Example 2.10.1. Groups are abundant in mathematics:

- The (set of) integers $\mathbb{Z}$ forms a group $(\mathbb{Z}, +)$ with the usual addition $+$ of two numbers. It it not hard to convince ourselves that $+$ satisfies the identity, inverse and associativity properties a group law must satisfy.
- $\mathbb{Z}$ is *not* a group under the usual multiplication $\cdot$, since it fails to produce inverses: for any $n \in \mathbb{Z}$ with $n \neq 0, \pm 1$, the inverse of n must be $\frac{1}{n}$, by uniqueness[7] – but $\frac{1}{n} \notin \mathbb{Z}$.
- On the other hand $\mathbb{Q} \smallsetminus \{0\}$ is a group under multiplication, since it is closed under inverses: if $\frac{a}{b} \in \mathbb{Q}$ with $a \neq 0$, then $\frac{b}{a} \in \mathbb{Q}$.
- For $m > 0$, a complete residue system $C(m)$ is a group under "reduced addition:" for $a, b \in C(m)$, we define
$$a \oplus b := r$$
where $r \in C(m)$ satisfies
$$a + b \equiv r \pmod{m};$$
thus $a \oplus b$ is the congruence class representative of $a + b$ modulo m in $C(m)$. By definition of $C(m)$, such a choice of r is unique, meaning this operation is well-defined. For $a \in C(m)$, we write $-a$ for the inverse a^{-1}, since we know this element inverse in $C(m)$ is congruent to $-a$ mod m.
- Similar to the above, a reduced residue system $R(m)$ is a group under "reduced multiplication." We discuss this in more detail later.

One can come up with more exotic groups for examples, but for this course we are primarily interested in (and only have time for!) the groups more directly connected to the integers.

The group $(C(m), \oplus)$ above is important enough to give a distinguished name.

Definition 2.10.3. For an integer $m \in \mathbb{Z}^+$, let
$$\mathbb{Z}/m\mathbb{Z} := \{0, 1, 2, \ldots, m-1\}$$
denote the canonical complete residue system modulo m. We call $\mathbb{Z}/m\mathbb{Z}$ the **(set of) integers modulo** m**.** For the group $(\mathbb{Z}/m\mathbb{Z}, \oplus)$, we often write $+$ instead of $\oplus$, and $[a] \in \mathbb{Z}/m\mathbb{Z}$ instead of $a \in \mathbb{Z}/m\mathbb{Z}$.

[7]One can show that the inverse of a group element is unique – this will be proven in Exercise 2.10.1.

Remark 2.10.2. One also has a similar definition for other complete residue systems modulo m, but working with the canonical CRS mod m is meant to be more clarifying. As you will explore in Exercise 2.10.2, the choice of CRS does not affect the group structure of $(\mathbb{Z}/m\mathbb{Z}, +)$.

Example 2.10.2. The integers modulo 8 is

$$\mathbb{Z}/8\mathbb{Z} = \{0, 1, 2, 3, 4, 5, 6, 7\}.$$

We check that

$$[3] + [6] = [1],$$

since

$$3 + 6 = 9 \equiv 1 \pmod 8.$$

We also have $-[3] = [5]$, since $-3 \equiv 5 \pmod 8$.

Remark 2.10.3. If you have studied the integers modulo m in a previous context, you might think of elements of $\mathbb{Z}/m\mathbb{Z}$ as *congruence classes,* rather than fixed integers $0 \leq k < m$. When it comes to the group structure of $\mathbb{Z}/m\mathbb{Z}$, these two notions are equivalent: this will be explored in Exercise 2.10.2, see also Remark 2.10.2.

With our introduction to groups, there are many small but important facts to prove for ourselves. Several are in Exercise 2.10.1; I leave you to prove the following proposition on your own.

Proposition 2.10.1. *Let $m > 0$ be an integer. Then the following properties hold for the group $(\mathbb{Z}/m\mathbb{Z}, +)$:*

(1) $\mathbb{Z}/m\mathbb{Z}$ is an abelian group.
(2) $[0]$ is the identity.
(3) for all $[a] \in \mathbb{Z}/m\mathbb{Z}$, one has $-[a] = [m - a]$.

Definition 2.10.4. Given a group $(G, \oplus)$ and a subset $H \subseteq G$, we say that H is a **subgroup** of G (under $\oplus$) if $(H, \oplus)$ is a group with identity $e_H = e_G$. We often write $(H, \oplus) \subseteq (G, \oplus)$ or $H \leq G$ to indicate a subgroup containment.

Example 2.10.3. We have $(\mathbb{Z}, +) \subseteq (\mathbb{Q}, +) \subseteq (\mathbb{R}, +)$, as well as $(\mathbb{Q} \smallsetminus \{0\}, \cdot) \subseteq (\mathbb{R} \smallsetminus \{0\}, \cdot)$. However, we do not have $(\mathbb{Q} \smallsetminus \{0\}, \cdot) \subseteq (\mathbb{Q}, +)$, since the two operations $+$ and $\cdot$ are not "equivalent." The notion of equivalent groups is expanded on in the next definition.

In general, one can understand an algebraic object by understanding "structure-preserving maps" to and from said object. Let us make this more precise for groups.

Definition 2.10.5. Given two groups $(G, \oplus)$ and $(G', \odot)$, we say that a set map

$$\phi \colon G \to G'$$

is a **homomorphism** if it "preserves the group structure:" that is, for all $g, h \in G$ one has

$$\phi(g \oplus h) = \phi(g) \odot \phi(h).$$

If ϕ is also a bijection, then we call ϕ an **isomorphism,** and write

$$\phi \colon G \xrightarrow{\sim} G'.$$

In this case, we say that G and G' are **isomorphic,** and also write

$$G \cong G'$$

(or $G \cong_\phi G'$ to specify the isomorphism, or $(G, \oplus) \cong (G', \odot)$ to specify the group operations).

Isomorphic groups are essentially the same algebraic structures, just with a possibly different underlying set. Observe that isomorphic groups must have equal size.

The following is an example of one of the most important homomorphisms in number theory.

Example 2.10.4. Given an integer $m > 0$, we have the **mod-m reduction map**

$$\pi_m \colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$$

defined by

$$\pi_m(a) := [r]$$

where $[r] \in \mathbb{Z}/m\mathbb{Z}$ satisfies

$$a \equiv r \pmod{m}.$$

Intuitively, this map returns the remainder of an integer when divided by m. Let us carefully check that $\pi := \pi_m$ is a homomorphism. We want to show that for all $a, b \in \mathbb{Z}$

$$\pi(a + b) = \pi(a) + \pi(b).$$

If we write $[r_1] := \pi(a + b)$, $[r_2] := \pi(a)$ and $[r_3] := \pi(b)$, then this is equivalent to showing that

$$[r_1] = [r_2] + [r_3]$$

in $\mathbb{Z}/m\mathbb{Z}$. By definition of $(\mathbb{Z}/m\mathbb{Z}, +)$, this sum is true if

$$r_1 \equiv r_2 + r_3 \pmod{m},$$

which is true since $a + b \equiv r_1 \pmod{m}$, $a \equiv r_2 \pmod{m}$ and $b \equiv r_3 \pmod{m}$.

Here are more examples of groups.

Example 2.10.5. Observe that the set

$$\mathbb{Z}^\times := \{1, -1\} \subseteq \mathbb{Z}$$

is a group under the usual multiplication $\cdot$. We have an isomorphism

$$(\mathbb{Z}/2\mathbb{Z}, +) \xrightarrow{\sim} (\mathbb{Z}^\times, \cdot)$$

via $[0] \mapsto 1$ and $[1] \mapsto -1$.

Example 2.10.6. Consider a nonstandard complete residue system modulo 6, such as

$$X := \{6, 1, 8, 33, 16, 11\}$$

Similar to $(\mathbb{Z}/6\mathbb{Z}, +)$, we have that X is a group under "reduced addition" $\oplus$: for example, one has $11 \oplus 33 = 8$ since $11 + 33 \equiv 8 \pmod{6}$. In fact, this group is isomorphic to $(\mathbb{Z}/6\mathbb{Z}, +)$: we can construct a bijection

$$\phi \colon X \to \mathbb{Z}/6\mathbb{Z}$$

by identifying congruent numbers:

$$\begin{aligned} 6 &\mapsto [0]; \\ 1 &\mapsto [1]; \\ 8 &\mapsto [2]; \\ 33 &\mapsto [3]; \\ 16 &\mapsto [4]; \\ 11 &\mapsto [5]. \end{aligned}$$

Such a map ends up being an isomorphism since it is a homomorphism, as a consequence of congruence properties:

$$\phi(a \oplus b) = \phi(a) + \phi(b).$$

Thus $(X, \oplus) \cong (\mathbb{Z}/6\mathbb{Z}, +)$.

The example above generalizes to any complete residue system modulo a fixed m:

Theorem 2.10.2. [NZM91, Theorem 2.46] *For an integer $m > 0$, any complete residue system modulo m is a group under addition mod m, and is isomorphic to $(\mathbb{Z}/m\mathbb{Z}, +)$. Thus $(\mathbb{Z}/m\mathbb{Z}, +)$ is "the" additive group mod m.*

Just as $(\mathbb{Z}/m\mathbb{Z}, +)$ represents complete residue systems modulo m as additive groups, we can represent *reduced* residue systems mod m as multiplicative groups.

Definition 2.10.6. Given $m > 0$, we define the **unit group modulo** m as

$$(\mathbb{Z}/m\mathbb{Z})^\times := \{[a] \in \mathbb{Z}/m\mathbb{Z} : \gcd(a, m) = 1\}.$$

Its group operation $\odot$ is "reduced multiplication:" for $[a], [b] \in (\mathbb{Z}/m\mathbb{Z})^\times$ one has

$$[a] \odot [b] := [r]$$

where $0 \leq r < n$ is the remainder of ab divided by m; thus

$$ab \equiv r \pmod{m}.$$

We often write $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ instead of $((\mathbb{Z}/m\mathbb{Z})^\times, \odot)$. By definition of Euler's phi function, we have

$$|(\mathbb{Z}/m\mathbb{Z})^\times| = \varphi(m).$$

Similar to [NZM91, Theorem 2.46], it turns out that all reduced residue systems modulo m are groups under multiplication, and are isomorphic to one another.

Theorem 2.10.3. [NZM91, Theorem 2.47] *For an integer $m > 0$, any reduced residue system modulo m is a group under multiplication mod m, and is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^\times$. Thus $(\mathbb{Z}/m\mathbb{Z})^\times$ is "the" multiplicative group mod m.*

Remark 2.10.4. To wrap this section up, let us introduce some standard notation for groups. For an arbitrary group $(G, \oplus)$, one often writes gh or $g \cdot h$ instead of $g \oplus h$. Thus (and for example)

$$g^3 := ggg := g \cdot g \cdot g.$$

For an integer $n \geq 0$, let us write

$$g^{-n} := (g^{-1})^n.$$

One can check that the inverse of g^n is g^{-n}; this is in Exercise 2.10.1. We also set $g^0 := e$.

If G is abelian, one may write $+$ instead of $\cdot$, e.g.

$$3g := g + g + g,$$

with inverses $-g$ instead of g^{-1}, and identity $e := 0$. An exception to writing the group additively is when the group is known to have an "inherently" multiplicative group law, such as the group $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$.

Exercises. From [NZM91, §2.10], pages 119–120: $\#1 - 2, 7 - 8$.

Exercise 2.10.1. This exercise proves some routine but important facts about groups. Let G be a group.

a) Show that the identity element $e := e_G$ of G is unique.

Consider an element $g \in G$.

b) Show that if an element $h \in G$ satisfies

$$hg = e,$$

then one also has

$$gh = e,$$

and vice-versa.

c) Show that the inverse g^{-1} of g is unique.

d) Show that $(g^{-1})^{-1} = g$.

e) Show that for any $h \in G$ one has

$$(gh)^{-1} = h^{-1}g^{-1}.$$

f) For an integer $n \geq 0$, we set $g^{-n} := (g^{-1})^n$. Show that for **all** $n \in \mathbb{Z}$, one has

$$g^{-n} = (g^n)^{-1}.$$

Let G' be another group. Consider a group homomorphism

$$\phi \colon G \to G'.$$

g) Show that $\phi(e_G) = e_{G'}$.

h) Show that for any $g \in G$ one has

$$\phi(g^{-1}) = \phi(g)^{-1}.$$

i) Show that for any integer n and an element $g \in G$, one has

$$\phi(g^n) = \phi(g)^n.$$

j) Show that if ϕ is an isomorphism, then so is its inverse map $\phi^{-1} \colon G' \to G$.

Exercise 2.10.2. This exercise gives an alternate definition of $\mathbb{Z}/m\mathbb{Z}$, the integers modulo m. In particular, it gives us a new description for the additive group $(\mathbb{Z}/m\mathbb{Z}, +)$, as well as the unit group $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$.

Let us define a relation on $\mathbb{Z}$ as follows: say

$$a \sim b$$

when one has

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

a) Show that $\sim$ is an *equivalence relation:* i.e., show it is reflexive, symmetric and transitive.

b) For an integer $a \in \mathbb{Z}$, denote its equivalence class under $\sim$ as $[a]$. Describe $[a]$ as a set.

c) Let X be the *quotient set* $\mathbb{Z}/\sim$, i.e., let X be the collection of equivalence classes under $\sim$. Show that X is a group under an addition law $\oplus$ such that

$$(X, \oplus) \cong (\mathbb{Z}/m\mathbb{Z}, +).$$

d) Let $Y \subseteq X$ be the subset of equivalence classes for integers which are coprime to m:

$$Y := \{x \in X : \exists a \in \mathbb{Z} \text{ with } \gcd(a, m) = 1 \text{ and } x = [a]\}.$$

Show that Y is a group under a multiplication law $\odot$ such that

$$(Y, \odot) \cong ((\mathbb{Z}/m\mathbb{Z})^\times, \cdot).$$

Exercise 2.10.3.

a) Prove that for $m > 1$, the groups $(\mathbb{Z}/m\mathbb{Z}, +)$ and $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ are never isomorphic.

b) Try to describe the group homomorphisms $(\mathbb{Z}/m\mathbb{Z}, +) \to ((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$.

Bonus Exercise 2.10.4. Prove that the groups $(\mathbb{Z}/6\mathbb{Z}, +)$ and $((\mathbb{Z}/7\mathbb{Z})^\times, \cdot)$ are isomorphic by giving an explicit isomorphism, writing down where each element goes.

Bonus Exercise 2.10.5. This exercise will give examples of nonabelian groups. For each integer $n \in \mathbb{Z}^+$, let $\mathrm{Mat}_{n\times n}(\mathbb{R})$ denote the set of $n \times n$ matrices with real entries.

a) Check that $\mathrm{Mat}_{n\times n}(\mathbb{R})$ is an abelian group under matrix addition.

b) Explain why $\mathrm{Mat}_{n\times n}(\mathbb{R})$ is *not* a group under matrix multiplication.

c) Define $\mathrm{GL}_n(\mathbb{R})$, the *general linear group of $n \times n$ matrices,* as the subset of invertible matrices in $\mathrm{Mat}_{n\times n}(\mathbb{R})$. Convince yourself that $\mathrm{GL}_n(\mathbb{R})$ is a group under matrix multiplication.

d) Show that $\mathrm{GL}_n(\mathbb{R})$ is abelian if and only if $n = 1$. What is $\mathrm{GL}_1(\mathbb{R})$ isomorphic to, as a group?

2.11. **Groups, Rings and Fields.** This section is a direct continuation of Section 2.10. However, we will study not just groups, but also **rings** and **fields,** which have two binary operations on them that interface with each other. The prototypical example of a ring and field is $\mathbb{Z}$ and $\mathbb{Q}$, respectively.

Here are the main goals for this section:

- Define the *order* of a group element.
- Prove a special case of *Lagrange's Theorem,* on orders of elements of finite groups dividing their group's size.
- Define rings and fields.
- Define direct products of groups and rings.
- Review CRT for $\mathbb{Z}/m\mathbb{Z}$ and $(\mathbb{Z}/m\mathbb{Z})^\times$.

As we will see, the *order* of elements in groups such as $\mathbb{Z}/m\mathbb{Z}$ and $(\mathbb{Z}/m\mathbb{Z})^\times$ are related to interesting congruence properties of integers – some of which we have seen before, such as Euler's Theorem (and Fermat's Little Theorem).

Our first result for this section concerns finiteness in groups, which we will soon use to define the order of an element.

Theorem 2.11.1. [NZM91, Theorem 2.48] *Let G be a group.*

1. *There exists a* cancellation property *for G: for elements $g, h, k \in G$, if*
$$gh = gk$$
then $h = k$. Similarly $hg = kg$ implies $h = k$.
2. *If G has finite order, then for any element $g \in G$ there exists a* least positive integer $r \in \mathbb{Z}^+$ *for which*
$$g^r = e.$$

Proof.

1. Assume that
$$gh = gk.$$
Then multiply both sides on the left by g^{-1}:
$$\begin{aligned} g^{-1} \cdot gh &= g^{-1} \cdot gk \\ (g^{-1}g) \cdot h &= (g^{-1}g) \cdot k && \text{(by associativity)} \\ e \cdot h &= e \cdot k \\ h &= k && \text{(by definition of the identity).} \end{aligned}$$
The case where $hg = kg$ is similar.
2. Assume that $|G|$ is finite. For an element $g \in G$, consider the subset
$$\{e, g, g^2, g^3, \ldots\} \subseteq G.$$
Since G is finite, so is $\{e, g, g^2, g^3, \ldots\}$. By the Pigeonhole Principle, there exist integers s and t with $t > s \geq 0$ such that
$$g^t = g^s,$$
i.e.,
$$g^{t-s} \cdot g^s = g^s.$$

By the cancellation property, this implies that

$$g^{t-s} = e.$$

Since $t - s > 0$, we conclude by the Well-Ordering Principle that there exists a least positive integer k with

$$g^k = e. \qquad \square$$

This concept of an element "wrapping back to zero" via $g^k = e$ is the concept of the *order* of an element.

Definition 2.11.1. Given a group G, finite or otherwise, if an element $g \in G$ satisfies

$$g^n = e$$

for some $n \in \mathbb{Z}^+$, then g is said to have **finite order.** In this case, the least positive integer $r \in \mathbb{Z}^+$ for which

$$g^r = e$$

is called the **order** of g, and is written as

$$|g| := r.$$

If $g^n \neq e$ for all $n \in \mathbb{Z}^+$, then g is said to have **infinite order.**

Remark 2.11.1. By Theorem 2.11.1 ([NZM91, Theorem 2.48]), any element from a finite group must have finite order.

There is an alternative formulation of the order of an element which can be useful in proofs.

Proposition 2.11.2. *Let G be a group. Then the order of an element $g \in G$ is equal to the number of distinct nonnegative powers of g.*

Proof. We first assume that $|g|$ is finite. Let $n \in \mathbb{Z}^+$ denote the number of such distinct powers. By minimality of $|g|$, we observe that the $|g|$-number-of-powers

$$e = g^0, g, g^2, \ldots, g^{|g|-1} \tag{15}$$

are distinct: otherwise, for some $0 \leq s < t \leq |g| - 1$ we have

$$g^t = g^s,$$

i.e.,

$$g^{t-s} = e,$$

which contradicts minimality of $|g|$ since $0 \leq t - s < |g|$. This implies that $|g| \leq n$.

Suppose for contradiction that $|g| < n$, i.e., suppose there are more distinct nonnegative powers of g than $|g|$. Then by the powers in (15) being distinct, there exists $k \geq |g|$ such that for each $0 \leq i < |g|$ we have

$$g^k \neq g^i.$$

Applying the Division Algorithm to k and $|g|$, we find that for some $q, r \in \mathbb{Z}$

$$k = q \cdot |g| + r$$

where $0 \leq r < |g|$. We check that

$$\begin{aligned} g^k &= g^{q \cdot |g| + r} \\ &= (g^{|g|})^q \cdot g^r \\ &= e^q \cdot g^r \\ &= e \cdot g^r \\ &= g^r. \end{aligned}$$

However, since $0 \leq r < |g|$ we know that g^r is in the list $e, g, g^2, \ldots, g^{|g|-1}$, which is absurd since $g^k = g^r$. We conclude that $|g| = n$.

Next, we assume that $|g| = \infty$. Then each of the powers

$$e, g, g^2, \ldots$$

are distinct: otherwise, for some $0 \leq s < t$ we have $g^t = g^s$, and thus $g^{t-s} = e$ for $0 \leq t - s$, contradicting that the order of g is infinite. This concludes our proof. □

Following the proof above, the distinct nonnegative powers of an element in a group define a special subgroup.

Definition 2.11.2. Given a group G, for each element $g \in G$ we define the **subgroup generated by** g as

$$\langle g \rangle := \{g^k : k \in \mathbb{Z}\}.$$

This is a subgroup of G. We say that G **cyclic** if there exists an element $g \in G$ with

$$G = \langle g \rangle.$$

Such an element g is called a **generator** for G.

Remark 2.11.2. For a group G and an element $g \in G$ with finite order, as noted in the proof of Proposition 2.11.2 one has

$$\langle g \rangle = \{e, g, g^2, \ldots, g^{|g|-1}\}.$$

We thus have $|\langle g \rangle| = |g|$.

For a finite group G, there are a few ways to detect whether G is cyclic, i.e., has a generator. Here is one such way: it is an application of Proposition 2.11.2. Prove it yourself!

Proposition 2.11.3. *If G is a finite group, then an element $g \in G$ is a generator if and only if $|g| = \#G$.*

Example 2.11.1. We claim that $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic, and generated by $[1]$. This follows from the fact that for any element $[d] \in \mathbb{Z}/m\mathbb{Z}$, one has

$$[d] = d \cdot [1].$$

We deduce that

$$(\mathbb{Z}/m\mathbb{Z}, +) = \langle [1] \rangle.$$

In Exercise 2.11.1, you will explore the other generators of $(\mathbb{Z}/m\mathbb{Z}, +)$.

In contrast to the above, it turns out that $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ is cyclic precisely when $m = 1, 2, 4, p^e$ or $2p^e$ where p^e is an odd prime power. We will investigate this in the next section (§2.8).

Cyclic groups are the simplest type of group: in fact, for each integer $n \in \mathbb{Z}^+$ there is exactly one cyclic group of order n *up to isomorphism.* Prove this one for yourself:

Proposition 2.11.4. *Let $n > 0$ be an integer. Then any two cyclic groups of order n are isomorphic.*

We are now ready to prove an important case of *Lagrange's Theorem:* generally, this theorem shows that the order of any subgroup of a finite group G divides $|G|$. We will prove Lagrange's Theorem for orders of elements, i.e., for *cyclic subgroups.* However, we will still call our result Lagrange's Theorem.

Theorem 2.11.5 (Lagrange's Theorem)**.** [NZM91, Theorem 2.49] *Let G be a finite group. Then the order of any element $g \in G$ divides $|G|$, i.e.,*

$$|g| \mid |G|.$$

In particular, one has

$$g^{|G|} = e.$$

Proof. As noted in the proof of Proposition 2.11.2, the subgroup

$$\langle g \rangle = \{e, g, g^2, \ldots, g^{|g|-1}\}$$

has $|g|$ distinct elements. If $\langle g \rangle = G$, then $|g| = |G|$ and we are done. Suppose then that $\langle g \rangle \subsetneq G$. Fixing any $h_2 \in G \smallsetminus \langle g \rangle$, consider the set

$$h_2\langle g \rangle := \{h_2, h_2 g, h_2 g^2, \ldots, h_2 g^{|g|-1}\}.$$

We claim that $\langle g \rangle$ and $h_2\langle g \rangle$ are *disjoint.* If this is not true, then for some $0 \leq k < |g|$ we have $g^k \in h_2\langle g \rangle$, so that

$$g^k = h_2 g^\ell$$

for some $0 \leq \ell < |g|$. But then $h_2 = g^{k-\ell} \in \langle g \rangle$, which contradicts our assumption on h_2. We deduce that $\langle g \rangle \cap h_2\langle g \rangle = \emptyset$. One can also show that the map

$$\langle g \rangle \to h_2\langle g \rangle, \qquad g^k \mapsto h_2 g^k$$

is a bijection, and thus $|h_2\langle g \rangle| = |\langle g \rangle| = |g|$. We conclude that

$$|\langle g \rangle \cup h_2\langle g \rangle| = 2|g|.$$

If

$$G = \langle g \rangle \cup h_2\langle g \rangle,$$

then taking sizes shows that $2|g| = |G|$, and we are done. Suppose then that $G \neq \langle g \rangle \cup h_2\langle g \rangle$; fixing $h_3 \in G \smallsetminus (\langle g \rangle \cup h_2\langle g \rangle)$, one can show that $\langle g \rangle, h_2\langle g \rangle$ and $h_3\langle g \rangle$ are pairwise disjoint, as well as $|h_3\langle g \rangle| = |g|$, and thus

$$|\langle g \rangle \cup h_2\langle g \rangle \cup h_3\langle g \rangle| = 3|g|.$$

Eventually, this process of finding h_i's must terminate since $|G|$ is finite, which means that for some $k \geq 1$ we have

$$G = \langle g \rangle \cup h_2\langle g \rangle \cup \ldots \cup h_k\langle g \rangle,$$

which implies $k|g| = |G|$. This proves Lagrange's Theorem. □

What's nice about Lagrange's Theorem is that it provides an alternative proof to some of our earlier results. For example, it reproves Euler's Theorem: given $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$, we have by Lagrange's Theorem that

$$||[a]|| \text{ divides } |(\mathbb{Z}/m\mathbb{Z})^\times| = \varphi(m),$$

so that

$$[a]^{\varphi(m)} = [1],$$

i.e.,

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

Example 2.11.2. Let us do a sanity check for Lagrange's Theorem. Consider the integers modulo 6:

$$\mathbb{Z}/6\mathbb{Z} = \{[0], [1], [2], [3], [4], [5]\}.$$

The *additive* orders of these elements are

$$||[0]|| = 1, ||[1]|| = 6, ||[2]|| = 3, ||[3]|| = 2, ||[4]|| = 3, ||[5]|| = 6.$$

Each of these orders divide $|(\mathbb{Z}/6\mathbb{Z}, +)| = 6$, so Lagrange's Theorem checks out. Do you spot a pattern to these orders?

On the other hand, the multiplicative group of the integers modulo 6 is

$$(\mathbb{Z}/6\mathbb{Z})^\times = \{[1], [5]\}.$$

The *multiplicative* order of [5] is 2. Lagrange's Theorem still checks out since $|(\mathbb{Z}/6\mathbb{Z})^\times| = 2$.

Having discussed the basics of groups, we now move on to learning the basics of *rings,* and by extension *fields.* In a way, a ring is one step above a group, as it admits two distinct binary operations which interact with each other. Rings are more analogous to the full algebraic structure of $\mathbb{Z}$, which has both an addition and a multiplication law.

Definition 2.11.3. A **ring** is a set R with the following properties.

1. R has two binary operations $\oplus = +$ and $\odot = \cdot$, such that $(R, \oplus)$ **is a commutative group.**
2. $(R, \oplus)$ has an additive identity $0 := 0_R$, and an element $1 := 1_R$ which acts like a multiplicative identity, such that $0 \neq 1$. For $r \in R$, we write $-r$ for its inverse under $\oplus$, and r^{-1} for its inverse under $\odot$ *when it exists.*
3. ($\odot$ is **associative**): for all $r, s, t \in R$ one has

   $$(r \odot s) \odot t = r \odot (s \odot t),$$

   i.e.,

   $$(r \cdot s) \cdot t = r \cdot (s \cdot t).$$

   We often write this as $r \odot s \odot t, r \cdot s \cdot t$ or rst.

4. (**distributive law**): $\odot$ distributes over $\oplus$, that is, for $r, s, t \in R$ one has

$$r \odot (s \oplus t) = (r \odot s) \oplus (r \odot t),$$

i.e.,

$$r \cdot (s + t) = r \cdot s + r \cdot t,$$

and similarly

$$(r + s) \cdot t = r \cdot t + s \cdot t.$$

Given two rings $(R, \oplus_R, \odot_R)$ and $(S, \oplus_S, \odot_S)$, we say that S is a **subring** of R if $S \subseteq R, \oplus_S = \oplus_R, \odot_S = \odot_R, 0_S = 0_R$ and $1_S = 1_R$. We write $S \subseteq R$ or $S \leq R$ to denote this.

Example 2.11.3. One can check that under the usual addition and multiplication, we have rings $(\mathbb{Z}, +, \cdot) \subseteq (\mathbb{Q}, +, \cdot) \subseteq (\mathbb{R}, +, \cdot)$.

Remark 2.11.3. One can show from the distributive law that the group $(R, \oplus)$ must be abelian, whether or not we assume it – this follows from writing $(1 + 1)(a + b)$ in two different ways. Try and prove this yourself!

Here are some additional definitions for rings.

Definition 2.11.4. Let R be a ring. If for all $r, s \in R$ one has

$$r \cdot s = s \cdot r,$$

we say that R is **commutative.**

For a ring R, we define the **unit group of** R as the subset of (multiplicatively) invertible elements:

$$R^\times := \{r \in R : \exists s \in R \text{ with } rs = sr = 1\}.$$

The elements in $R^\times$ are called *units.* In Exercise 2.11.3, you will check that $R^\times$ is a group under multiplication. One always has $R^\times \subseteq R \smallsetminus \{0\}$, but if

$$R^\times = R \smallsetminus \{0\}$$

then all nonzero elements of R are invertible, and if R is also commutative, then we call R a **field.**

Example 2.11.4. A very important example of a ring beyond $\mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ is the integers modulo m. Recall that

$$\mathbb{Z}/m\mathbb{Z} := \{[0], [1], \ldots, [m-1]\}.$$

We have previously defined "reduced addition" on $\mathbb{Z}/m\mathbb{Z}$, which makes it a group. We have also defined "reduced multiplication" on $(\mathbb{Z}/m\mathbb{Z})^\times$, but in fact, this multiplication extends to $\mathbb{Z}/m\mathbb{Z}$. Together, these two operations make $\mathbb{Z}/m\mathbb{Z}$ a ring. Even our previous notation $(\mathbb{Z}/m\mathbb{Z})^\times$ agrees with our definition of the unit group of a ring! This is not a coincidence.

The ring $\mathbb{Z}/m\mathbb{Z}$ may or may not be a field. For example, when $m = 6$ the element $[3] \in \mathbb{Z}/m\mathbb{Z}$ does not have an inverse: if there existed $[d] \in \mathbb{Z}/m\mathbb{Z}$ with

$$[3] \cdot [d] = [1],$$

then we would have

$$3d \equiv 1 \pmod 6,$$

and reducing mod 3 would imply that

$$0 \equiv 1 \pmod 3,$$

which is absurd. However, we can characterize which rings $\mathbb{Z}/m\mathbb{Z}$ are fields.

Theorem 2.11.6. [NZM91, Theorem 2.50] *For an integer $m > 0$, the set $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m is a ring under reduced addition and multiplication mod m. One has that $\mathbb{Z}/m\mathbb{Z}$ is a field if and only if m is prime.*

Proof. The example above showed that $\mathbb{Z}/m\mathbb{Z}$ is a ring. For the second part: if $\mathbb{Z}/m\mathbb{Z}$ is a field, then for any nonzero $[a] \in \mathbb{Z}/m\mathbb{Z}$ there exists $[b] \in \mathbb{Z}/m\mathbb{Z}$ with

$$[a] \cdot [b] = [1],$$

i.e.,

$$ab \equiv 1 \pmod m.$$

Such a congruence implies $\gcd(a, m) = 1$. Thus m is coprime to all integers in $[1, m]$, which forces m to be prime. For the other direction: if $m = p$ is prime, then all nonzero elements $[a] \in \mathbb{Z}/p\mathbb{Z}$ satisfy $\gcd(a, p) = 1$. In particular $[a]$ is invertible. We conclude that $\mathbb{Z}/p\mathbb{Z}$ is a field. □

Similar to the case of groups, one can understand rings by understanding maps between them which "preserve the ring structure."

Definition 2.11.5. Given two rings $(R, \oplus_R, \odot_R)$ and $(S, \oplus_S, \odot_S)$, a **ring homomorphism** from R to S is a set map

$$\varphi \colon R \to S$$

such that:

1. φ is an additive group homomorphism: for elements $a, b \in R$ one has
$$\varphi(a \oplus_R b) = \varphi(a) \oplus_S \varphi(b).$$
2. φ respects multiplication: for $a, b \in R$ one has
$$\varphi(a \odot_R b) = \varphi(a) \odot_S \varphi(b).$$
3. $\varphi(1_R) = 1_S$.

If φ is also a bijection, then we say that φ is a **ring isomorphism,** and that R and S are **isomorphic.** We write

$$\varphi \colon R \xrightarrow{\sim} S$$

and

$$R \cong S$$

(or $(R, \oplus_R, \odot_R) \cong (S, \oplus_S, \odot_S)$).

Example 2.11.5. For any integer $m > 0$, we have the **mod-m reduction map**

$$\pi_m \colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z},$$

where for $a \in \mathbb{Z}$ we set

$$\pi_m(a) := [r]$$

with

$$a \equiv r \pmod{m}.$$

This is the same *map* as the group homomorphism $\pi_m \colon (\mathbb{Z}, +) \to (\mathbb{Z}/m\mathbb{Z}, +)$ from Example 2.10.4, but we now see that the group homomorphism can be interpreted as a *ring homomorphism.*

We now move on to our final topics, which are products of groups and rings, and the Chinese Remainder Theorem for $\mathbb{Z}/m\mathbb{Z}$ and $(\mathbb{Z}/m\mathbb{Z})^\times$.

Definition 2.11.6. Given two groups G and H, we define the **direct product of G and H** as the Cartesian product

$$G \times H := \{(g, h) : g \in G, h \in H\}.$$

Then $G \times H$ is a group, with a coordinatewise group operation:

$$(g_1, h_1) \cdot (g_2, h_2) := (g_1 g_2, h_1 h_2).$$

This generalizes to a product of $n \geq 1$ groups: consider the Cartesian product

$$G_1 \times G_2 \times \ldots \times G_n := \{(g_1, g_2, \ldots, g_n) : \forall 1 \leq i \leq n, g_i \in G_i\}.$$

Then the group law on $G_1 \times G_2 \times \ldots \times G_n$ is

$$(g_1, g_2, \ldots, g_n) \cdot (g_1', g_2', \ldots, g_n') := (g_1 g_1', g_2 g_2', \ldots, g_n g_n').$$

Similarly, given two rings R and S, the Cartesian product

$$R \times S := \{(r, s) : r \in R, s \in S\}$$

is a ring via coordinatewise addition and multiplication: for $(r_1, s_1), (r_2, s_2) \in R \times S$ we have

$$(r_1, s_1) + (r_2, s_2) := (r_1 + r_2, s_1 + s_2)$$

and

$$(r_1, s_1) \cdot (r_2, s_2) := (r_1 r_2, s_1 s_2).$$

This generalizes to a ring structure on a Cartesian product of finitely many rings.

We can reinterpret the Chinese Remainder Theorem as a result on the "factorization" of the ring $\mathbb{Z}/m\mathbb{Z}$ and its unit group $(\mathbb{Z}/m\mathbb{Z})^\times$.

Theorem 2.11.7 (CRT for the ring $\mathbb{Z}/m\mathbb{Z}$)**.** *Let $m > 1$ be an integer with prime factorization*

$$m = \prod_{i=1}^{r} p_i^{e_i}.$$

Then the natural reduction map

$$\begin{aligned}\mathbb{Z}/m\mathbb{Z} &\to \mathbb{Z}/p_1^{e_1} \times \mathbb{Z}/p_2^{e_2} \times \ldots \times \mathbb{Z}/p_r^{e_r} \\ [a] &\mapsto ([a_1], [a_2], \ldots, [a_r]),\end{aligned}$$

is a ring isomorphism. *Thus*

$$\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/p_1^{e_1} \times \mathbb{Z}/p_2^{e_2} \times \ldots \times \mathbb{Z}/p_r^{e_r}.$$

From this, we deduce the group isomorphisms

$$(\mathbb{Z}/m\mathbb{Z}, +) \cong (\mathbb{Z}/p_1^{e_1}, +) \times (\mathbb{Z}/p_2^{e_2}, +) \times \ldots \times (\mathbb{Z}/p_r^{e_r}, +)$$

and

$$(\mathbb{Z}/m\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{e_1})^\times \times (\mathbb{Z}/p_2^{e_2})^\times \times \ldots \times (\mathbb{Z}/p_r^{e_r})^\times.$$

We omit the proof of this theorem, but it is essentially identical to the proof of CRT for residues systems that we proved in §2.3 (see Theorem 2.3.2 ([NZM91, Theorem 2.19]) and Remark 2.3.2). The unit group isomorphism has an interesting alternative proof, from noting that a ring isomorphism

$$R \xrightarrow{\sim} S$$

implies a unit group isomorphism

$$R^\times \xrightarrow{\sim} S^\times,$$

and then applying this inductively. You will prove this Exercise 2.11.3.

Example 2.11.6. By CRT for the ring $\mathbb{Z}/60\mathbb{Z}$, one has a natural ring isomorphism

$$\mathbb{Z}/60\mathbb{Z} \cong (\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/5\mathbb{Z})$$

which implies that

$$(\mathbb{Z}/60\mathbb{Z})^\times \cong (\mathbb{Z}/4\mathbb{Z})^\times \times (\mathbb{Z}/3\mathbb{Z})^\times \times (\mathbb{Z}/5\mathbb{Z})^\times.$$

This allows us to study the algebraic structure of $\mathbb{Z}/m\mathbb{Z}$ and $(\mathbb{Z}/m\mathbb{Z})^\times$ via factorization. For example $(\mathbb{Z}/4\mathbb{Z})^\times$ and $(\mathbb{Z}/3\mathbb{Z})^\times$ are cyclic groups of order 2, and $(\mathbb{Z}/5\mathbb{Z})^\times$ is cyclic of order 4. Therefore $(\mathbb{Z}/60\mathbb{Z})^\times$ is a product of three cyclic groups, of orders $2, 2$ and 4, respectively.

Exercises. From [NZM91, §2.11], pages 126–127: $\#1 - 6, 12, 14 - 16, 19 - 21$.

Exercise 2.11.1. Let G be a group.

a) For an element $g \in G$, prove that if for some $k \in \mathbb{Z}^+$ one has
$$g^k = e,$$
then g has finite order and $|g| \mid k$.

b) Show that if $g \in G$ has finite order, then for all $k \in \mathbb{Z}^+$ one has
$$|g^k| = \frac{|g|}{\gcd(|g|, k)}.$$
Deduce that $|g^k| = |g|$ if and only if $\gcd(|g|, k) = 1$.

c) Prove that for each integer $m \in \mathbb{Z}^+$, the additive group $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic with $\varphi(m)$ generators.

d) Assuming that the unit group $(\mathbb{Z}/m\mathbb{Z})^\times$ is cyclic, prove it has $\varphi(\varphi(m))$ generators.

Exercise 2.11.2. Let G be an abelian group.

a) Show that for any elements $a, b \in G$, one has for each $n \in \mathbb{Z}$ that
$$(ab)^n = a^n b^n.$$

b) Show that if $a, b \in G$ have finite orders, then so does ab, and
$$|ab| \mid \operatorname{lcm}(|a|, |b|).$$

c) Show that if $a, b \in G$ have finite *coprime* orders, then
$$|ab| = |a| \cdot |b|.$$

(*Hint:* for parts b) and c), you will want to apply Exercise 2.11.1.)

For the next two exercises, we let $(R, +, \cdot)$ be a ring with additive identity $0 := 0_R$ and multiplicative identity $1 := 1_R$.

Exercise 2.11.3.

a) Show that for all $r \in R$ one has
$$r \cdot 0 = 0 \cdot r = 0.$$

b) Prove that the *unit group of R,* denoted as
$$R^\times := \{r \in R : \exists s \in R \text{ with } rs = sr = 1\},$$
is a group under multiplication.

c) Show that for rings $R_1, R_2, \ldots, R_n$, one has
$$(R_1 \times R_2 \times \cdots \times R_n)^\times \cong R_1^\times \times R_2^\times \times \cdots \times R_n^\times.$$

Let S be another ring, and let $\varphi\colon R \to S$ be a ring homomorphism.

d) Show that φ induces a group homomorphism on unit groups,
$$\varphi\colon R^\times \to S^\times.$$

Exercise 2.11.4. Say that an element $r \in R$ is a **zero divisor** if for some nonzero $s \in R$ one has
$$rs = 0.$$
We call R an **integral domain** if it has no nonzero zero divisors.

a) Show that an integral domain R satisfies the *cancellation property:* for $r, s, t \in R$ with $r \neq 0$, if
$$rs = rt$$
then $s = t$. Similarly, show that if $sr = tr$ then $s = t$.

b) Show that a field is an integral domain.

c) Give an example of a ring which is not an integral domain, and an integral domain which is not a field.

Bonus Exercise 2.11.5. In this exercise, let G and G' be finite groups and $\phi\colon G \to G'$ a homomorphism.

a) Given an element $g \in G$, show the order divisibility $|\phi(g)| \mid |g|$.
b) Show that if ϕ is surjective, then for all $g' \in G'$ one has $|g'| \mid |G|$.
c) Give an example of a nontrivial surjective group homomorphism $\phi\colon G \to G'$ where, for some $g \in G$, one has $|\phi(g)| < |g|$.
d) Show that if ϕ is injective, then $|\phi(g)| = |g|$. In particular, injective homomorphisms preserve orders of elements.

Bonus Exercise 2.11.6. This exercise explores the *characteristic* of an integral domain.

a) Prove that for a ring R, there exists exactly one ring homomorphism
$$\iota\colon \mathbb{Z} \to R.$$
b) Show that in part a), one has that ι is injective if and only if 1_R has infinite additive order.
c) Show that in part a), if ι is not injective, then *assuming that R is an integral domain* the additive order of 1_R is prime.

When ι is injective, we say that R has **characteristic zero.** When ι is not injective, we say that the integral domain R has **positive characteristic** p**,** where p is the additive order of 1_R. As it turns out, any field with characteristic zero contains $\mathbb{Q}$, and any field with positive characteristic p contains $\mathbb{Z}/p\mathbb{Z}$.

Bonus Exercise 2.11.7 (Examples of rings)**.** For each set R below, determine whether:

1. R is a ring;
2. R is commutative;
3. R is an integral domain;
4. R is a field.

If R is a ring, then determine its group of units $R^\times$ if possible.

a) The set $\mathbb{Z}[x]$ of polynomials with integer coefficients.
b) The set $C([0,1])$ of continuous real-valued functions $f\colon [0,1] \to \mathbb{R}$.
c) For $n \in \mathbb{Z}^+$, the set $\mathrm{Mat}_{n\times n}(\mathbb{R})$ of $n \times n$ matrices with real entries.
d) For $n \in \mathbb{Z}^+$, the set $\{c_0 + c_1x + \ldots + c_nx^n : c_i \in \mathbb{Z}\}$ of degree $\leq n$ polynomials over $\mathbb{Z}$.
e) The set of Gaussian integers $\mathbb{Z}[i] := \{a + bi : a, b \in \mathbb{Z}\}$.
f) The set of squares of rational numbers, $\{\frac{a^2}{b^2} : a, b \in \mathbb{Z}, b \neq 0\}$.
g) The set of real-valued functions $f\colon \mathbb{R} \to \mathbb{R}$ with $\lim_{x\to 0} f(x) = 0$.

Bonus Exercise 2.11.8. Prove that for any prime $p > 2$, writing
$$1 + \frac{1}{2^3} + \ldots + \frac{1}{(p-1)^3} = \frac{a}{b}$$
where $a, b \in \mathbb{Z}$, one has $p \mid a$. (*Hint:* interpret this sum modulo p, and use the identity $\sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4}$.)

2.8. **Primitive Roots and Power Residues.** In our last section of Chapter 2, we will study *primitive roots,* which are generators of the multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$. Primitive roots can be used to find solutions to a special class of polynomials modulo p (see below), and have connections to the *discrete logarithm problem* in cryptography (which someone will hopefully present on at the end of the semester!). For integers a, n and p with $n > 0$ and p prime, we will also study *n'th power residues modulo p,* which are solutions to the congruence

$$x^n - a \equiv 0 \pmod{p}.$$

Here are our main goals for this section.

- Review when primitive roots exist modulo m, and determine how to lift them from prime powers to higher-level powers.
- Prove *Euler's Criterion* for n'th power residues.

As we observed in §2.10, the additive group $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic, and can be generated by [1]. However, the multiplicative group $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ need not be cyclic. For example, in the last section we saw the group isomorphism

$$(\mathbb{Z}/60\mathbb{Z})^\times \cong (\mathbb{Z}/4\mathbb{Z})^\times \times (\mathbb{Z}/3\mathbb{Z})^\times \times (\mathbb{Z}/5\mathbb{Z})^\times.$$

One can use Exercise 2.11.2 to check that all elements of $(\mathbb{Z}/60\mathbb{Z})^\times$ have order at most 8. Since $|(\mathbb{Z}/60\mathbb{Z})^\times| = \varphi(60) = 16$, we conclude from this that $(\mathbb{Z}/60\mathbb{Z})^\times$ is not cyclic.

Definition 2.8.1. Fix an integer $m > 0$. Then for any integer $1 \leq g < m$, if $[g]$ generates $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ then we say that g is a **primitive root modulo** m**.**

Remark 2.8.1. By Proposition 2.11.3, we see that $1 \leq g < m$ is a primitive root modulo m if and only if the multiplicative order of $[g]$ in $(\mathbb{Z}/m\mathbb{Z})^\times$ is equal to $|(\mathbb{Z}/m\mathbb{Z})^\times|$, which is $\varphi(m)$.

Example 2.8.1. Determining the multiplicative orders of elements in $\mathbb{Z}/m\mathbb{Z}$ is usually harder than calculating the additive order. However, knowing the structure of $(\mathbb{Z}/m\mathbb{Z})^\times$ will help tell us when this can or cannot happen. For example, we just noted that no element of $(\mathbb{Z}/60\mathbb{Z})^\times$ has order 16. However, for

$$(\mathbb{Z}/7\mathbb{Z})^\times = \{[1], [2], [3], [4], [5], [6]\}$$

we can check that [3] has order $\varphi(7) = 6$:

$$\begin{aligned}
[3]^2 &= [2];\\
[3]^3 &= [6];\\
[3]^4 &= [4];\\
[3]^5 &= [5];\\
[3]^6 &= [1].
\end{aligned}$$

We conclude that 3 is a primitive root mod 7.

Abstract Algebra Digression 2.8.2. Let us note that primitive roots modulo m are the "finite ring" analog of primitive n'th roots of unity. Recall that a *primitive n'th root (of unity)* is a number $\zeta_n \in \mathbb{C}^\times$ with multiplicative order n, i.e.,

$$\zeta_n^n = 1$$

but $\zeta_n^d \neq 1$ for each proper divisor d of n. In comparison, a primitive root modulo m is an integer $1 \leq g < m$ with

$$g^m \equiv 1 \pmod{m}$$

but $g^d \not\equiv 1 \pmod{m}$ for each proper divisor d of n. Both types of primitive roots are particular roots of a polynomial of the form $x^n - 1$, either over $\mathbb{C}$ or $\mathbb{Z}/m\mathbb{Z}$.

As a consequence of Exercise 2.11.1, if there exists a primitive root modulo m then there are exactly $\varphi(\varphi(m))$ such primitive roots. This begs the question: *how do we know if there exists a primitive root modulo m?* As it turns out, there is a precise condition on m for primitive roots to exist.

Theorem 2.8.1. [NZM91, Theorem 2.41] *For an integer $m \in \mathbb{Z}^+$, there exists a primitive root modulo m if and only if $m = 1, 2, 4, p^k$ or $2p^k$, where p is an odd prime.*

The proof in [NZM91] is quite involved, and involves an analysis of factoring polynomials modulo p which we have not touched on. Instead, Bonus Exercise 2.8.6 sketches an alternate proof.

We now expand on the concept of "lifting" primitive roots modulo a prime power.

Theorem 2.8.2. [NZM91, Theorem 2.40] *If p is an odd prime and g is a primitive root modulo p^2, then g is a primitive root modulo p^k for all $k \geq 2$.*

Proof. For an integer $k \geq 1$, we let r_k denote the order of $[g]$ in $(\mathbb{Z}/p^k\mathbb{Z})^\times$. For $k \geq 2$, to show that g is a primitive root modulo p^k it is equivalent to show that

$$r_k = \varphi(p^k) = p^{k-1}(p-1).$$

We proceed by induction on k.

The base case $k = 2$ is our initial assumption. Assume then that the result is true for $k < n$. To show that $r_n = \varphi(p^n) = p^{n-1}(p-1)$, we first note that $r_{n-1} \mid r_n$: this follows from the fact that

$$g^{r_n} \equiv 1 \pmod{p^n},$$

which implies

$$g^{r_n} \equiv 1 \pmod{p^{n-1}},$$

and so $r_{n-1} \mid r_n$ by Exercise 2.11.1. By the inductive hypothesis, we have $r_{n-1} = \varphi(p^{n-1}) = p^{n-2}(p-1)$, whence we deduce that

$$p^{n-2}(p-1) \mid r_n.$$

However, by Lagrange's Theorem we also know that

$$r_n \mid \varphi(p^n).$$

Since $\varphi(p^n) = p^{n-1}(p-1)$, we deduce that

$$p^{n-2}(p-1) \mid r_n \mid p^{n-1}(p-1),$$

from which we conclude that r_n equals either $p^{n-2}(p-1)$ or $p^{n-1}(p-1)$, i.e.,

$$r_n \in \{\varphi(p^{n-1}), \varphi(p^n)\}. \tag{16}$$

Let us recap: to show that g is a primitive root modulo p^n, we must show that $r_n = \varphi(p^n)$. By (16), it suffices to show that $r_n \neq \varphi(p^{n-1})$, i.e.,

$$g^{\varphi(p^{n-1})} \not\equiv 1 \pmod{p^n}.$$

We first observe that from

$$g^{\varphi(p^{n-2})} \equiv 1 \pmod{p^{n-2}}$$

we can write

$$g^{\varphi(p^{n-2})} = 1 + bp^{n-2} \tag{17}$$

for some $b \in \mathbb{Z}$. We necessarily have

$$p \nmid b, \tag{18}$$

since otherwise $g^{\varphi(p^{n-2})} \equiv 1 \pmod{p^{n-1}}$, contradicting that $\varphi(p^{n-1})$ is the order of $[g]$ in $(\mathbb{Z}/p^{n-1}\mathbb{Z})^\times$. We use (17) to expand $g^{\varphi(p^{n-1})}$ and show it is not congruent to 1 mod p^n, thus showing by (16) that the order of $[g]$ in $(\mathbb{Z}/p^n\mathbb{Z})^\times$ must be $\varphi(p^n)$.

$$\begin{aligned}
g^{\varphi(p^{n-1})} &= g^{p\varphi(p^{n-2})} && \text{(since } n \geq 2\text{)} \\
&= (g^{\varphi(p^{n-2})})^p \\
&= (1 + bp^{n-2})^p && \text{(by (17))} \\
&= \sum_{k=0}^{p} \binom{p}{k} (bp^{n-2})^k \cdot 1^{p-k} && \text{(Binomial Theorem for } (bp^{n-2}+1)\text{)} \\
&= 1 + p \cdot bp^{n-2} + \frac{p(p-1)}{2} \cdot b^2 p^{2(n-2)} + \dots.
\end{aligned}$$

Observe that for the terms indexed by $k \geq 2$ above, one has $k(n-2)+1 \geq n$ since $n \geq 3$, and so $p^n \mid p^{k(n-2)+1}$; thus, reducing this equation modulo p^n gives

$$g^{\varphi(p^{n-1})} \equiv 1 + bp^{n-1} \pmod{p^n}.$$

We noted in (18) that $p \nmid b$; thus $g^{\varphi(p^{n-1})} \not\equiv 1 \pmod{p^n}$. We deduce that the order r_n of $[g]$ in $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is not $\varphi(p^{n-1})$, whence we conclude by (16) that $r_n = \varphi(p^n)$. Therefore g is a primitive root modulo p^n. This concludes our proof by induction. □

We have just shown that a primitive root modulo p^2 is also a primitive root mod p^3, p^4 and so on. However, this still leaves the problem of determining primitive roots modulo p and p^2. One can adapt our proof above to prove the following result.

Proposition 2.8.3. *If p is an odd prime and g is a primitive root modulo p such that*

$$g^{p-1} \not\equiv 1 \pmod{p^2},$$

then g is a primitive root modulo p^2, and thus modulo p^3, p^4 and so on.

Proof. Following our notation from the previous proof, we can show that (16) still holds, and that the multiplicative order r_2 of g modulo p^2 is either $p-1=\varphi(p)$ or $p(p-1)=\varphi(p^2)$. By our assumption that $g^{p-1}\not\equiv 1 \pmod{p^2}$, we conclude that $r_2=\varphi(p^2)$. □

Remark 2.8.2. It is worth noting that our proof of Theorem 2.8.2 ([NZM91, Theorem 2.40]) can be adapted (using our Binomial Theorem expansion) to prove that primitive roots "descend:" for any prime $p>2$, if g is a primitive root modulo p^n for some $n\geq 2$, then g is a primitive root mod p^k for $1\leq k\leq n$. In particular g is a primitive root modulo *all* powers of p.

Similar to finding solutions to congruences modulo p, the above proposition allows us to sometimes reduce the problem of finding primitive roots modulo odd prime powers p^k to finding a primitive root modulo p. However, there is no general formula for finding primitive roots modulo p. Nonetheless, there are techniques you can employ to search for them – a few exercises at the end of this section will address this.

Remark 2.8.3. Given that primitive roots modulo m exist if and only if $m=1,2,4,p^k$ or $2p^k$ where p is an odd prime (Theorem 2.8.1 ([NZM91, Theorem 2.41])), after Theorem 2.8.2 ([NZM91, Theorem 2.40]) one can ask what is known about primitive roots modulo $2p^k$. There is a tight connection between primitive roots mod p^k and $2p^k$: if g is a primitive root mod p^k, then either g or $g+p^k$ is a primitive root mod $2p^k$. Conversely, if g is a primitive root mod $2p^k$, then it is a primitive root mod p^k once reduced to an integer in $[1,p^k]$. Prove this for yourself! (Note that $\varphi(2p^k)=\varphi(p^k)$.)

Next, we turn our attention to studying n'th roots modulo p – this will have a close connection to studying primitive roots modulo p.

Definition 2.8.2. Let a,n and p be integers with p prime and $\gcd(a,p)=1$. We say that a is an n**'th power residue modulo** p if the congruence

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

has a solution, i.e., if x^n-a has a root modulo p.

In other words, an n'th power residue modulo p is analogous to taking an n'th root of a real or complex number.

Example 2.8.3. We have studied n'th power residues already in this chapter: for example, we studied the existence of solutions to x^2+1 modulo p, and have characterized them in terms of congruence classes modulo 4, as well as sums of squares (see Remark 2.1.7).

There is a theorem attributed to Euler that describes when x^n-a has solutions modulo p, and characterizes them.

Theorem 2.8.4 (Euler's Criterion (for n'th Power Residues))**.** [NZM91, Theorem 2.37] *Let a,n and p be integers with p prime and* $\gcd(a,p)=1$. *Then the congruence*

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

has a solution if and only if

$$a^{\frac{p-1}{\gcd(n,p-1)}} \equiv 1 \pmod{p}.$$

In this case, it has exactly $\gcd(n, p-1)$ *solutions.*

To prove this, we will repeatedly apply our Linear Congruence Theorem from §2.2 "in our exponents." We recall it here.

Theorem (Linear Congruence Theorem)**.** [NZM91, Theorem 2.17] *Fix integers* a *and* b*, and* $m > 0$*. Then the congruence*

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

has a solution if and only if

$$\gcd(a, m) \mid b.$$

In this case, there are $\gcd(a, m)$ *distinct solutions modulo* m*, given by*

$$c = (a')^{-1} \cdot \frac{b}{\gcd(a, m)} + m' \cdot k,$$

where $0 \leq k < \gcd(a, m)$*,* $a' := \frac{a}{\gcd(a,m)}$*,* $m' := \frac{m}{\gcd(a,m)}$ *and* $(a')^{-1}$ *is an integer representative of the multiplicative inverse of* a' *mod* m'*.*

Proof of Euler's Criterion. Fix a primitive root g modulo p; then $[g]$ generates $(\mathbb{Z}/p\mathbb{Z})^\times$. First, we prove the forward direction. We suppose that

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

has a solution: so there exists $x_0 \in \mathbb{Z}$ with

$$x_0^n \equiv a \pmod{p}.$$

We check that

$$\begin{aligned} a^{\frac{p-1}{\gcd(n,p-1)}} &\equiv (x_0^n)^{\frac{p-1}{\gcd(n,p-1)}} \pmod{p} \\ &= (x_0^{p-1})^{\frac{n}{\gcd(n,p-1)}} \\ &= 1 \pmod{p} \qquad \text{(since } x_0^{p-1} \equiv 1 \pmod{p} \text{ by e.g. Euler's Theorem).} \end{aligned}$$

For the other direction, suppose that

$$a^{\frac{p-1}{\gcd(n,p-1)}} \equiv 1 \pmod{p}.$$

Since $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$, we can write

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

for some $\ell \in \mathbb{Z}$. Then our assumption above says that

$$g^{\ell \cdot \frac{p-1}{\gcd(n,p-1)}} \equiv 1 \pmod{p}.$$

Since g has order $p-1$ in $(\mathbb{Z}/p\mathbb{Z})^\times$, this implies by Lagrange's Theorem that

$$p-1 \mid \ell \cdot \frac{p-1}{\gcd(n, p-1)}.$$

Dividing both sides by $\frac{p-1}{\gcd(n,p-1)}$ shows that

$$\gcd(n, p-1) \mid \ell.$$

By the Linear Congruence Theorem, this implies that the congruence

$$xn \equiv \ell \pmod{p-1}$$

has $\gcd(n, p-1)$ solutions. Pick any solution $c \in \mathbb{Z}$; then

$$cn \equiv \ell \pmod{p-1}.$$

We check that $x^n - a$ has a solution modulo p from this:

$$\begin{aligned} a &\equiv g^\ell \pmod{p} \\ &\equiv g^{cn} \pmod{p} && \text{(by Exercise 2.1.5)} \\ &= (g^c)^n. \end{aligned}$$

We deduce that a is congruent to x_0^n modulo p where $x_0 := g^c$. In particular a is an n'th power residue mod p.

We conclude our proof by noting that when $x^n \equiv a \pmod{p}$ has solutions, it has exactly $\gcd(n, p-1)$ solutions, each of the form g^c where c ranges over the $\gcd(n, p-1)$-number of solutions for the linear congruence

$$ny \equiv \ell \pmod{p-1},$$

where ℓ is from $a \equiv g^\ell \pmod{p}$. □

Euler's Criterion is particularly amenable to calculations. Let us write our formulas down: when $\gcd(a, p) = 1$, the congruence

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

has a solution if and only if

$$a^{\frac{p-1}{\gcd(n,p-1)}} \equiv 1 \pmod{p}.$$

In such a case, there are exactly $\gcd(n, p-1)$ solutions. Fixing any primitive root g modulo p and writing $a = g^\ell$, these solutions are given by

$$x = g^c$$

where c is any of the $\gcd(n, p-1)$ solutions to the "linear exponent congruence"

$$ny \equiv \ell \pmod{p-1},$$

which have an explicit characterization as

$$c = (n')^{-1} \cdot \frac{\ell}{\gcd(n, p-1)} + (p-1)' \cdot k \pmod{p-1}$$

with $0 \le k < \gcd(n, p-1)$. Here, we have written $n' := \left(\frac{n}{\gcd(n,p-1)}\right)$ and $(p-1)' := \frac{p-1}{\gcd(n,p-1)}$, and $(n')^{-1}$ as the multiplicative inverse of n' modulo $(p-1)'$. (Note that $\gcd(n, p-1) \mid \ell$.)

Example 2.8.4. Suppose one wants to determine solutions to the congruence

$$x^5 \equiv 6 \pmod{101}.$$

Note that 101 is prime, so Euler's Criterion applies; here, we have $a = 6, n = 5$ and $p = 101$. To have solutions, we must have

$$6^{\frac{101-1}{\gcd(5,101-1)}} \equiv 1 \pmod{101},$$

i.e.,

$$6^{\frac{100}{\gcd(5,100)}} \equiv 1 \pmod{101},$$

which simplifies to checking

$$6^{20} \equiv 1 \pmod{101}.$$

Check for yourself that this is true, e.g. using the techniques in Example 2.1.5. Therefore, Euler's Criterion implies that we have exactly $\gcd(5, 100) = 5$ solutions modulo 101. Fixing a primitive root g mod 101 and writing $6 \equiv g^\ell \pmod{101}$, the solutions to

$$x^5 \equiv 6 \pmod{101}$$

have the form

$$x = g^c$$

where

$$c = \left(\frac{5}{\gcd(5,100)}\right)^{-1} \cdot \frac{\ell}{\gcd(5,100)} + \frac{100}{\gcd(5,100)} \cdot k$$

for $0 \leq k < 5$. This simplifies to

$$c = \frac{\ell}{5} + 20k$$

for $k = 0, 1, 2, 3, 4$. Thus, our solutions to $x^5 \equiv 6 \pmod{101}$ are

$$x = g^{\frac{\ell}{5}}, g^{\frac{\ell}{5}+20k}, g^{\frac{\ell}{5}+40k}, g^{\frac{\ell}{5}+60k}, g^{\frac{\ell}{5}+80k},$$

where g is *any* fixed primitive root mod 101 and $g^\ell \equiv 6 \pmod{101}$. For example, one can show that 2 is a primitive root mod 101, and that $2^{70} \equiv 6 \pmod{101}$. Thus, one can take $g = 2$ and $\ell = 70$, and conclude that these solutions are

$$2^{14}, 2^{34}, 2^{54}, 2^{74}, 2^{94}.$$

As noted previously, finding primitive roots modulo p can be difficult, but there are some techniques to do so. For example, Exercise 2.8.1 provides one such way.

Remark 2.8.4. Fixing a primitive root g mod m, let us write $\log_g(a) := \log_g^m(a)$ for the *discrete log of a modulo m with base g:* Then the formula for the exponents of solutions to $x^n \equiv a \pmod{p}$ can be condensed as

$$c = (n')^{-1} \cdot \frac{\log_g(a)}{\gcd(n, p-1)} + (p-1)' \cdot k$$

where $k = 0, 1, \ldots, \gcd(n, p-1) - 1$. Exercise 2.8.3 will explore the discrete logarithm.

One last thing to mention in this chapter is an important case of Euler's Criterion: the *quadratic case,* where $n = 2$. This is commonly referred to as **the** Euler's Criterion. It determines when an integer a coprime to a prime p is a *square* modulo p.

Theorem 2.8.5. [NZM91, Corollary 2.38] *Let p be an odd prime and a an integer with $p \nmid a$. Then a is a square modulo p, i.e., the congruence*

$$x^2 \equiv a \pmod{p}$$

has a solution, if and only if

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

In this case, it has exactly two solutions.

In the above theorem, if a is a square modulo p then we say that a is a **quadratic residue modulo p.** Otherwise, we say that a is a **quadratic nonresidue modulo p.** We will study quadratic residues in detail in Chapter 3.

Exercises. From [NZM91, §2.8], page 106: $\#1-3, 7-9, 12-14$.

Exercise 2.8.1. Let G be a finite group.

a) Show that an element $g \in G$ is a generator of G if and only if for all proper divisors d of $|G|$ one has $g^d \neq e$.
b) After proving part a), show that an element $g \in G$ is a generator of G iff for all primes $p \mid |G|$ one has $g^{\frac{|G|}{p}} \neq e$.
c) Without a calculator, use part b) to show that 2 is a primitive root modulo 29, and is *not* a primitive root modulo 31.

Exercise 2.8.2.

a) Use Exercise 2.8.1 to show that 3 is a primitive root modulo 43.
b) Without a calculator, determine with proof the solutions to $x^6 \equiv -2 \pmod{43}$ as powers of 3, if they exist.

Exercise 2.8.3. Let g be a primitive root modulo m. Let us define the **discrete logarithm (modulo m with base g)** as follows. For each element $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$, we know there exists unique $0 \leq k < \varphi(m)$ with

$$[g]^k = [a],$$

i.e.,

$$g^k \equiv a \pmod{m}.$$

We define the *discrete logarithm of a* as

$$\log_g(a) := k.$$

(Sometimes $\log_g^m(a)$ is used to emphasize the modulus m.)

a) Show that for any $a, b \in \mathbb{Z}$ coprime to m, one has

$$\log_g(a \cdot b) \equiv \log_g(a) + \log_g(b) \pmod{\varphi(m)}.$$

(*Hint:* Exercise 2.11.1 will help.)
b) Prove the following *change of base formula* for discrete logarithms: given another primitive root h modulo m, show that $\log_h(g)$ is coprime to $\varphi(m)$, and that for all $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$ one has

$$\log_g(a) \equiv \frac{\log_h(a)}{\log_h(g)} \pmod{\varphi(m)},$$

where $\frac{1}{\log_h(g)} := \log_h(g)^{-1}$ is the multiplicative inverse of $\log_h(g)$ modulo $\varphi(m)$. (*Hint:* Exercise 2.11.1 will help.)

c) Assume 3 and 5 are primitive roots modulo $4802 = 2 \cdot 7^4$, and that the discrete logarithm $\log_3^{4802}(5)$ is equal to 911. Use part b) to compute the discrete logarithm $\log_5^{4802}(81)$.

Exercise 2.8.4.

a) Use Exercise 2.8.1 to create a `Sage` function which for positive integers g and m, takes as input (m, g) and returns `True` if g is a primitive root modulo m. Have it return `False,` and print a message, if m has no primitive roots, or if g is not a primitive root mod m.

Run output for this to check whether $g = 2$ is a primitive root modulo p, for primes $p \leq 163$.

b) Based on part a), can you come up with a conjecture on the primes p which have 2 as a primitive root? You can extend your search over larger p if necessary. (One point)

c) Use part a) to create a `Sage` function which for positive integers a, g and m with g a primitive root modulo m, takes as input (a, m, g) and outputs the discrete logarithm $\log_g(a) := \log_g^m(a)$, which is the integer $0 \leq \log_g(a) < m$ with

$$g^{\log_g(a)} \equiv a \pmod{m}.$$

Have it return a message if any of the following hold: m does not have primitive roots; g is not a primitive root mod m; or a is not coprime to m.

Run output for this function for $g = 2$, $a = 3$ and over primes $p \leq 163$.

Bonus Exercise 2.8.5. This exercise studies **Fermat numbers,** which are integers of the form $2^n + 1$ for $n \geq 0$. The first few Fermat numbers are listed here: https://oeis.org/A000215.

A prime number which is a Fermat number is called a **Fermat prime.** More information about them can be found here: https://oeis.org/A019434.

a) Show that if a Fermat number is prime, then it is of the form $2^{2^k} + 1$ for some $k \geq 0$. (*Hint:* consider how to factorize the difference$x^a - y^a$ of odd a'th powers of two numbers x and y.)

b) Show that 2 is a primitive root modulo any Fermat prime p.

c) More generally, show that if a is not a square modulo a Fermat prime p (so $x^2 \equiv a \pmod{p}$ has no solutions), then a is a primitive root modulo p.

The only known Fermat primes are $3, 5, 17, 257$ and 65537.

Bonus Exercise 2.8.6. The following exercise outlines a proof of [NZM91, Theorem 2.41] on when primitive roots modulo m exist.

a) Let G and H be finite abelian groups. Show that the order of any element $(g, h) \in G \times H$ is equal to $\text{lcm}(|g|, |h|)$. (*Hint:* Exercise 2.11.2 should help.)

b) Explain how part a) should generalize to a product $G_1 \times G_2 \times \ldots \times G_n$ of finite abelian groups.

c) Let an integer $m > 1$ have prime factorization

$$m = \prod_{i=1}^{r} p_i^{e_i}.$$

Show that for any element $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$, one has

$$|[a]|^\times = \operatorname{lcm}\{|[\pi_{p_i^{e_i}}(a)]|^\times\}_{i=1}^r.$$

(Here, we use $|[a]|^\times$ to denote the multiplicative order of $[a]$ in $(\mathbb{Z}/m\mathbb{Z})^\times$, and for $d \mid m$ we let $\pi_d \colon (\mathbb{Z}/m\mathbb{Z})^\times \to (\mathbb{Z}/d\mathbb{Z})^\times$ denote the mod-d reduction map.)

d) Finally, use the previous parts to give an alternate proof for the existence of primitive roots:

Theorem. [NZM91, Theorem 2.41] *For an integer $m \in \mathbb{Z}^+$, there exists a primitive root modulo m if and only if $m = 1, 2, 4, p^k$ or $2p^k$, where p is an odd prime.*

3. Quadratic Reciprocity

In this chapter, we are interested in studying the existence of solutions modulo primes p to polynomials of the form $x^2 - a$ where $a \in \mathbb{Z}$. As we saw in §2.7 (and in Remark 2.1.7), there is a concise characterization for when $x^2 + 1$ has roots modulo p: when p is odd, this is equivalent to $p \equiv 1 \pmod 4$, which is equivalent to p being a sum of two perfect squares. As we will see, there is a more general characterization for the existence of solutions to $x^2 - a$ modulo p, i.e., to determine when a is a square mod p. When this happens, we will say that a is a *quadratic residue modulo p.*

The information of an integer a being a quadratic residue modulo p is captured in the *Legendre symbol of a modulo p.* As we will see, the Legendre symbol is very amenable to calculations. It will also satisfy a property called the *Law of Quadratic Reciprocity,* a remarkable symmetry which will help make calculations with the Legendre symbol easier. There are currently 345 documented proofs of Quadratic Reciprocity, see here. At the end of this chapter, we will also discuss a more general *Jacobi symbol,* which will help us determine whether integers are quadratic residues modulo p without having to verify the entire factorization of integers in intermediate steps.

3.1. **Quadratic Residues.** Here are the goals for this section.

- Define quadratic residues modulo primes p, and Legendre symbols mod p.
- Prove some algebraic properties of Legendre symbols.

A significant portion of Chapter 2 was dedicated to determining solutions to a congruence of the form

$$f(x) \equiv 0 \pmod m$$

for polynomials $f(x) \in \mathbb{Z}[x]$ and moduli $m > 0$. We saw through the Chinese Remainder Theorem and Hensel's Lemma that this problem can often be reduced to studying solutions to

$$f(x) \equiv 0 \pmod p.$$

However, as noted in §2.7 there are usually no good techniques to analyze roots of $f(x)$ mod p beyond plugging in values. Nonetheless, in §2.8 we proved Euler's Criterion (for n'th Power Residues), which helps us find solutions to

$$x^n \equiv a \pmod p.$$

Here is Euler's Criterion restated.

Theorem 3.1.1 (Euler's Criterion (for n'th Power Residues))**.** [NZM91, Theorem 2.37] *Let a, n and p be integers with p prime and $\gcd(a, p) = 1$. Then the congruence*

$$x^n \equiv a \pmod p$$

has a solution if and only if

$$a^{\frac{p-1}{\gcd(n,p-1)}} \equiv 1 \pmod p.$$

In this case, it has exactly $\gcd(n, p-1)$ solutions. (Fixing a primitive root g modulo p and $\ell \in \mathbb{Z}$ with $g^\ell \equiv a \pmod p$, these solutions are determined by the linear congruence $ny \equiv \ell \pmod{p-1}$.)

We ended Chapter 2 with a special case, often known as THE Euler's Criterion.

Theorem 3.1.2 (Euler's Criterion)**.** [NZM91, Corollary 2.38] *Let p be an odd prime and a an integer with $p \nmid a$. Then a is a square modulo p, i.e., the congruence*

$$x^2 \equiv a \pmod{p}$$

has a solution, if and only if

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

In this case, it has exactly two solutions.

Let us introduce some notation for recording when $x^2 - a$ has a (nonzero) square modulo p.

Definition 3.1.1. For positive coprime integers a and m, we say that a is a **quadratic residue** modulo m (or **QR mod** m) if $x^2 - a$ has a root modulo m. Otherwise, we say that a is a **quadratic nonresidue** modulo m (or **QNR mod** m).

Remark 3.1.1. In the definition of quadratic residues and nonresidues, we assume that $\gcd(a, m) = 1$. This means we must exclude e.g. 0 from being a quadratic residue mod m, despite it being a square in $\mathbb{Z}$. This is a standard choice, since it will make our analysis of quadratic residues more straightforward. For example, we will eventually see that the product of a quadratic residue and nonresidue mod m must be a nonresidue; on the other hand, the product of a square and nonsquare mod m can be either a square or nonsquare mod m (to see this, try $m = 15$, and consider ab with $a = 5$ and $b = 6$).

It is worth noting that we will work primarily with residues modulo **prime** p; thus, the nonzero squares mod p are precisely the quadratic residues mod p.

Example 3.1.1.

- For any integer $m > 0$, we see that 1 is a quadratic residue modulo m.
- When is -1 a quadratic residue mod m? This is the same as asking whether $x^2 + 1$ mod m has a root. When $m := p$ is an odd prime, this is true if and only if $p \equiv 1 \pmod 4$, iff p is a sum of two perfect squares; see the remarks at the start of this chapter.
- Is 5 a quadratic residue modulo $m = 7$? Let us compute all the nontrivial squares mod 7:

$$\begin{aligned} 1^2 &= 1; \\ 2^2 &= 4; \\ 3^2 = 9 &\equiv 2 \pmod{7}; \\ 4^2 \equiv (-3)^2 &\equiv 2 \pmod{7}; \\ 5^2 \equiv (-2)^2 &= 4 \pmod{7}; \\ 6^2 \equiv (-1)^2 &= 1 \pmod{7}. \end{aligned}$$

Thus 5 is a quadratic *non*residue mod 7.

- One finds that 10000000000000000001 (ten sextillion and one, which has twenty-one zeros) is a quadratic residue modulo 5: this is because
$$10000000000000000001 \equiv 1 \pmod 5,$$
and 1 is a quadratic residue mod 5.
- In the exercises at the end of §3.3, you will show that 1011 is a quadratic residue modulo 9907. At the moment, this is far from obvious!

Remark 3.1.2. In general, for integers a and b, if $a \equiv b \pmod m$ and a is a quadratic residue modulo m, then so is b. In practice, we often change our congruence class representative to be between 0 and m; we did this in the previous example with ten sextillion and one.

Definition 3.1.2. For an odd prime p and an integer a, we define the **Legendre symbol of a modulo p** as follows:
$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{if } a \text{ is a quadratic residue modulo } p, \\ -1 & \text{if } a \text{ is a quadratic nonresidue modulo } p, \\ 0 & \text{if } p \mid a. \end{cases}$$

It is important to note that our Legendre symbol notation is *not* a fraction.

Example 3.1.2. Based on our examples above:
- For any integer $m > 0$, one has $\left(\frac{1}{m}\right) = 1$.
- For any prime $p > 2$, one has $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \pmod 4$, iff p is a sum of two perfect squares.
- $\left(\frac{5}{7}\right) = -1$.
- $\left(\frac{10000000000000000001}{5}\right) = \left(\frac{1}{5}\right) = 1$.
- $\left(\frac{1011}{9007}\right) = 1$, to be proven in the exercises at the end of §3.3!

Here are some useful properties from the "algebra" of Legendre symbols.

Theorem 3.1.3. [NZM91, Theorem 3.1] *For prime $p > 2$, one has the following for all $a, b \in \mathbb{Z}$.*

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

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

(3) if $a \equiv b \pmod p$ *then* $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(4) if $\gcd(a, p) = 1$ *then* $\left(\frac{a^2}{p}\right) = 1$, *and thus* $\left(\frac{a^2 b}{p}\right) = \left(\frac{b}{p}\right)$.

(5) $\left(\frac{1}{p}\right) = 1$, *and*
$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4, \\ -1 & \text{if } p \equiv -1 \pmod 4. \end{cases}$$

Proof.

(1) If $p \mid a$, then both $a^{\frac{p-1}{2}}$ and $\left(\frac{a}{p}\right)$ are congruent to 0 modulo p. Suppose then that $\gcd(a, p) = 1$.
   (a) If $x^2 - a$ has a root modulo p, then by Euler's Criterion

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

   We let b be a root of $x^2 - a \bmod p$. Then $b^2 - a \equiv 0 \pmod{p}$, i.e. $a \equiv b^2 \pmod{p}$, so that $\left(\frac{a}{p}\right) = 1$ by definition. Thus $a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) = 1 \pmod{p}$.
   (b) If $x^2 - a$ has no roots mod p, then a is a quadratic nonresidue modulo p, so that $\left(\frac{a}{p}\right) = -1$. Additionally, by Euler's Criterion we have $a^{\frac{p-1}{2}} \not\equiv 1 \pmod{p}$. However $a^{\frac{p-1}{2}}$ is a root of $x^2 - 1 \bmod p$, since

$$\begin{aligned}(a^{\frac{p-1}{2}})^2 &= a^{p-1} \\ &\equiv 1 \pmod{p} \qquad \text{(by Euler's Theorem).}\end{aligned}$$

   Since $x^2 - 1$ has roots $x = -1, 1$, this forces $a^{\frac{p-1}{2}}$ to be the root $-1 \bmod p$, i.e.,

$$a^{\frac{p-1}{2}} \equiv -1 \pmod{p}.$$

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

The other parts follow from part (1):

(2) We check that

$$\begin{aligned}\left(\frac{ab}{p}\right) &\equiv (ab)^{\frac{p-1}{2}} \pmod{p} \qquad \text{(from part (1))} \\ &= a^{\frac{p-1}{2}} \cdot b^{\frac{p-1}{2}} \\ &\equiv \left(\frac{a}{p}\right) \cdot \left(\frac{b}{p}\right) \pmod{p} \qquad \text{(by another application of part (1)).}\end{aligned}$$

   Thus

$$p \mid \left(\left(\frac{ab}{p}\right) - \left(\frac{a}{p}\right) \cdot \left(\frac{b}{p}\right)\right),$$

   and since $p > 2$ this forces

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

(3) Since $a \equiv b \pmod{p}$, we see that a is zero/a quadratic residue/a quadratic nonresidue modulo p if and only if b is zero/a quadratic residue/a quadratic nonresidue modulo p, respectively.

(4) We check that

$$\begin{aligned}\left(\frac{a^2}{p}\right) &\equiv (a^2)^{\frac{p-1}{2}} \pmod{p} && \text{(by part (1))}\\ &= a^{p-1}\\ &\equiv 1 \pmod{p} && \text{(by Euler's Theorem).}\end{aligned}$$

This implies that $\left(\frac{a^2}{p}\right) = 1$. More generally, we check that

$$\begin{aligned}\left(\frac{a^2b}{p}\right) &\equiv (a^2b)^{\frac{p-1}{2}} \pmod{p} && \text{(by part (1))}\\ &= (a^2)^{\frac{p-1}{2}} \cdot b^{\frac{p-1}{2}}\\ &= a^{p-1} \cdot b^{\frac{b-1}{2}}\\ &\equiv b^{\frac{p-1}{2}} \pmod{p} && \text{(by Euler's Theorem)}\\ &\equiv \left(\frac{b}{p}\right) \pmod{p} && \text{(by part (1)),}\end{aligned}$$

which forces $\left(\frac{a^2b}{p}\right) = \left(\frac{b}{p}\right)$.

(5) The first part is clear. The second part follows from part (1) and [NZM91, Theorem 2.12] (see also our remarks at the start of this chapter). □

We can calculate much more Legendre symbols with the above properties. Here are a few.

Example 3.1.3.

- $\left(\frac{75}{11}\right) = \left(\frac{3\cdot 5^2}{11}\right) = \left(\frac{3}{11}\right)$. Since $5^2 \equiv 3 \pmod{11}$, we deduce that $\left(\frac{3}{11}\right) = 1$. Thus 75 is a quadratic residue modulo 11. (Alternatively, one could have noted $75 \equiv 9 \pmod{11}$ at the start, which is clearly a square.)
- $\left(\frac{10000000000000000000001}{5}\right) = \left(\frac{1}{5}\right) = 1$.
- $\left(\frac{96}{97}\right) = \left(\frac{-1}{97}\right) = -1$, since $97 \equiv -1 \pmod{4}$.

In the next section, we will prove Quadratic Reciprocity, and see how this simplifies more complicated Legendre symbol calculations.

Exercises. From [NZM91, §3.1], pages $135 - 136$: $\#4 - 5, 7 - 10$.

Exercise 3.1.1. The following exercise can be useful for simplifying Legendre symbol calculations. Show that for any integer $k \in \mathbb{Z}^+$, one has the following.

a) k is congruent to its unit digit a_0 modulo 2, and thus $2 \mid k$ if and only if $a_0 = 0, 2, 4, 6$ or 8.
b) k is congruent to the sum of its digits modulo 3, and thus $3 \mid k$ if and only if 3 divides this sum.
c) k is congruent to its unit digit a_0 modulo 5, and thus $5 \mid k$ if and only if $a_0 = 0, 5$.

(*Hint:* for each of these parts, work with the base 10 expansion of k: write $k = a_0 + a_1 \cdot 10 + a_2 \cdot 10^2 + \ldots + a_r \cdot 10^r$, where each $0 \leq a_i < 10$ and $a_r \neq 0$.)

Exercise 3.1.2. Prove that for a prime $p > 2$, if $a, b \in \mathbb{Z}$ are not squares modulo p then ab is a square modulo p.

Exercise 3.1.3. Let $p > 2$ be prime, and suppose that g is a primitive root modulo p.

a) Prove that for any $a \in \mathbb{Z}$ coprime to p, one has that a is a quadratic residue modulo p if and only if $a \equiv g^{2k} \pmod{p}$ for some $k \in \mathbb{Z}$.
b) Use part a) to give a complete list of quadratic residues modulo 19; **write them as positive integers in** $[1, 18]$**, listed in increasing order.**
c) How many quadratic residues are there modulo any odd prime p? What about quadratic nonresidues?

Bonus Exercise 3.1.4. Let a be an integer, and set $f(x) := x^2 - a$. Show that for any prime p, the number of solutions to $f(x) \equiv 0 \pmod{p}$ is $1 + \left(\frac{a}{p}\right)$.

3.2. **Quadratic Reciprocity.** The main goal of this section is to prove the classical theorem of Quadratic Reciprocity, which relates the Legendre symbols of two odd primes with respect to one another. As mentioned at the start of this chapter, there are at least 345 proofs of Quadratic Reciprocity – we've opted to go with a proof that's a balance of being "minimally technical" and not too long.

Theorem 3.2.1. [NZM91, Theorem 3.4] *Let p and q be distinct odd primes. Then one has*

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

i.e.,

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

In other words,

$$\left(\frac{q}{p}\right) = \begin{cases} \left(\frac{p}{q}\right) & \text{if } p \equiv 1 \pmod 4 \textbf{ or } q \equiv 1 \pmod 4 \\ -\left(\frac{p}{q}\right) & \text{if } p \equiv 3 \pmod 4 \textbf{ and } q \equiv 3 \pmod 4. \end{cases}$$

Quadratic Reciprocity can be very useful when calculating Legendre symbols, by "cutting down" large prime moduli. For example, calculating $\left(\frac{3}{101}\right)$ may seem tough at first glance, but we have by Quadratic Reciprocity that

$$\begin{aligned} \left(\frac{3}{101}\right) &= \left(\frac{101}{3}\right) && \text{(by Quadratic Reciprocity, since } 101 \equiv 1 \pmod 4\text{)} \\ &= \left(\frac{2}{3}\right) \\ &= -1. \end{aligned}$$

We conclude that 3 is not a square modulo 101. This is a much better way to determine this than to calculate all of the squares mod 101 by hand!

Proof of Quadratic Reciprocity. This proof is credited to George Rousseau [Rou91]. It uses the Chinese Remainder Theorem, Wilson's Theorem and Euler's Criterion.

The plan of this proof is to split $(\mathbb{Z}/pq\mathbb{Z})^\times$ by a subset H_1, where every element $[k] \in (\mathbb{Z}/pq\mathbb{Z})^\times$ is such that either $[k] \in H_1$ or $-[k] \in H_1$. We will then split the copy of $(\mathbb{Z}/q\mathbb{Z})^\times$ in $(\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$ by a subset H_2, and then use CRT to relate these two halves. This will manifest into a product of elements in H_1 and H_2 (this is Equation (19)). Finally, once we simplify these products, the Legendre symbols $\left(\frac{p}{q}\right)$ and $\left(\frac{q}{p}\right)$ will appear, and we will deduce Quadratic Reciprocity.

Let us define the sets

$$H_1 := \left\{ [k] \in (\mathbb{Z}/pq\mathbb{Z})^\times : 1 \leq k < \frac{pq}{2} \right\}$$

and

$$H_2 := \left\{ ([a], [b]) \in (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times : 1 \leq b < \frac{q}{2} \right\}.$$

By CRT for $\mathbb{Z}/pq\mathbb{Z}$, we have a unit group isomorphism

$$\Phi\colon (\mathbb{Z}/pq\mathbb{Z})^\times \xrightarrow{\sim} (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times.$$

By this isomorphism, for all elements $([a],[b]) \in (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$ there exists a unique $1 \leq k < pq$ with

$$\Phi([k]) = ([a],[b]),$$

i.e.,

$$\begin{aligned} k &\equiv a \pmod{p},\\ k &\equiv b \pmod{q}. \end{aligned}$$

(1) If $1 \leq k < \frac{pq}{2}$ then
$$\Phi([k]) = ([a],[b]).$$
(2) If $\frac{pq}{2} < k < pq$ then $1 \leq pq - k < \frac{pq}{2}$, and
$$\Phi([pq-k]) = \Phi([-k]) = -\Phi([k]) = -([a],[b]).$$
(3) Note that the case $k = \frac{pq}{2}$ cannot happen, since both p and q are odd, and thus $\frac{pq}{2} \notin \mathbb{Z}$.

We conclude that for every $([a],[b]) \in H_2$ there exists a unique $1 \leq k < pq$ with $\Phi([k]) = ([a],[b])$ such that either $k \in H_1$ or $pq - k \in H_1$. We will denote such a k as $k_{a,b}$ below.

We have the following equality of products in $(\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$:

$$\prod_{([a],[b])\in H_2} ([a],[b]) = \epsilon \cdot \prod_{[k]\in H_1} ([k],[k]), \tag{19}$$

where $\epsilon \in \{\pm 1\}$. This follows from the calculations

$$\begin{aligned}
\prod_{([a],[b])\in H_2} ([a],[b]) &= \prod_{([a],[b])\in H_2} \Phi([k_{a,b}])\\
&= \prod_{\substack{([a],[b])\in H_2:\\ k_{a,b}<\frac{pq}{2}}} \Phi([k_{a,b}]) \cdot \prod_{\substack{([a],[b])\in H_2:\\ \frac{pq}{2}<k_{a,b}}} \Phi([k_{a,b}])\\
&= \prod_{\substack{([a],[b])\in H_2:\\ k_{a,b}<\frac{pq}{2}}} \Phi([k_{a,b}]) \cdot (\pm 1) \cdot \prod_{\substack{([a],[b])\in H_2:\\ \frac{pq}{2}<k_{a,b}}} \Phi([pq-k_{a,b}])\\
&=: \epsilon \cdot \prod_{\substack{([a],[b])\in H_2:\\ k_{a,b}<\frac{pq}{2}}} \Phi([k_{a,b}]) \cdot \prod_{\substack{([a],[b])\in H_2:\\ \frac{pq}{2}<k_{a,b}}} \Phi([pq-k_{a,b}])\\
&= \epsilon \cdot \prod_{\substack{([a],[b])\in H_2:\\ k_{a,b}<\frac{pq}{2}}} ([k_{a,b}],[k_{a,b}]) \cdot \prod_{\substack{([a],[b])\in H_2:\\ \frac{pq}{2}<k_{a,b}}} ([pq-k_{a,b}],[pq-k_{a,b}]) \quad \text{(by definition of } \Phi)\\
&= \epsilon \cdot \prod_{[k]\in H_1} ([k],[k]);
\end{aligned}$$

combined with the fact that the integers $k_{a,b}$ and $pq - k_{a,b}$ in the respective products above range over all elements of H_1 exactly once. (Convince yourself of this! For example, if $([k_{a,b}], [k_{a,b}]) = ([pq - k_{c,d}], [pq - k_{c,d}])$, then $[d] = [q - b]$, which is impossible if $b, d < \frac{q}{2}$.)

We will simplify each side of (19) individually, and then equate them to deduce Quadratic Reciprocity. We first simplify the left-hand side. Let us set $P := \frac{p-1}{2}$ and $Q := \frac{q-1}{2}$. After each step of our calculation, we will make a comment explaining it. To ease notation, we will also drop the bracket notation for elements mod p and mod q:

$$\prod_{(a,b)\in H_2} (a, b) = \prod_{\substack{1\le a<p,\\ 1\le b<\frac{q}{2}}} (a, b) = ((p-1)!^Q, Q!^{p-1}),$$

noting that $Q = \frac{q-1}{2}$ is the greatest positive integer below $\frac{q}{2}$ (the exponents come from the number of choices for b and a in these pairs);

$$= ((p-1)!^Q, Q!^{2P}),$$

by definition of P;

$$\equiv ((-1)^Q, Q!^{2P}),$$

by Wilson's Theorem modulo p;

$$\equiv ((-1)^Q, ((q-1)!(-1)^Q)^P)),$$

by the claim that $(\left(\frac{q-1}{2}\right)!)^2 \equiv (-1)^{\frac{q-1}{2}} \cdot (q-1)! \pmod q$ (this will be a bonus exercise at the end of this section);

$$= ((-1)^Q, ((-1)^{Q+1})^P),$$

by Wilson's Theorem modulo q;

$$= ((-1)^Q, (-1)^{PQ+P}).$$

We have just shown that

$$\prod_{(a,b)\in H_2} (a, b) \equiv ((-1)^Q, (-1)^{PQ+P}). \tag{20}$$

We now move on to simplifying the right-hand side of (19), namely

$$\epsilon \cdot \prod_{[k]\in H_1} ([k], [k]).$$

We we will do this one coordinate at a time. **Modulo p,** we check that

$$\prod_{[k]\in H_1} k = \prod_{\substack{1\le k<\frac{pq}{2}:\\ \gcd(k,pq)=1}} k \equiv \left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ p\nmid k}} k\right) \cdot \left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1},$$

the latter congruence from the fact that the set of integers in $[1, \frac{pq}{2}]$ coprime to pq is the set of integers in $[1, \frac{pq}{2}]$ coprime to p with the multiples of q removed;

$$= \left(\prod_{0<k<p} k \cdot \prod_{p<k<2p} k \cdot \prod_{2p<k<3p} k \cdots \prod_{(Q-1)p<k<Qp} k \cdot \prod_{Qp<k<\frac{pq}{2}} k \right) \cdot \left(\prod_{\substack{1 \leq k < \frac{pq}{2}: \\ q \mid k}} k \right)^{-1},$$

where we have explicitly written out our product of all integers in $[1, \frac{pq}{2}]$ coprime to p – such integers are precisely the non-multiples of p. We note that the blue product might have less terms than the other products of $p-1$ integers, as it only goes up to $\frac{pq}{2}$;

$$\equiv \left(\underbrace{(p-1)! \cdot (p-1)! \cdot (p-1)! \cdots (p-1)!}_{Q \text{ times}} \cdot \prod_{Qp<k<\frac{pq}{2}} k \right) \cdot \left(\prod_{\substack{1 \leq k < \frac{pq}{2}: \\ q \mid k}} k \right)^{-1},$$

since each product of the form $\prod_{jp<k<(j+1)p} k$ above is congruent to $(p-1)!$ mod p;

$$= \left((p-1)!^Q \cdot \prod_{Qp<k<\frac{pq}{2}} k \right) \cdot \left(\prod_{\substack{1 \leq k < \frac{pq}{2}: \\ q \mid k}} k \right)^{-1} \equiv \left((p-1)!^Q \cdot P! \right) \cdot \left(\prod_{\substack{1 \leq k < \frac{pq}{2}: \\ q \mid k}} k \right)^{-1},$$

which follows from the fact that every integer $Qp < k < \frac{pq}{2}$ has the form $k = \ell + Qp$ with $1 \leq \ell \leq \frac{p-1}{2}$ – this is also a bonus exercise;

$$\equiv \frac{(p-1)!^Q \cdot P!}{q \cdot 2q \cdot 3q \cdots Pq},$$

since $\prod_{\substack{1 \leq k < \frac{pq}{2}: \\ q \mid k}} k \equiv q \cdot 2q \cdot 3q \cdots Pq \pmod{p}$ – part of this is also a bonus exercise – and we have written the inverse as a denominator;

$$= \frac{(p-1)!^Q \cdot P!}{q^P \cdot P!} = \frac{(p-1)!^Q}{q^P} \equiv \frac{(-1)^Q}{q^P},$$

by Wilson's Theorem modulo p;

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

by Euler's Criterion for the Legendre symbol, see Theorem 3.1.3 ([NZM91, Theorem 3.1]);

$$= (-1)^Q \cdot \left(\frac{q}{p}\right).$$

We conclude by our hard work above that

$$\prod_{[k]\in H_1} k \equiv (-1)^Q \cdot \left(\frac{q}{p}\right) \pmod{p}.$$

By a symmetric argument (where we switch the p's and q's in our work above, as well as the P's and Q's), we can also conclude that

$$\prod_{[k]\in H_1} k \equiv (-1)^P \cdot \left(\frac{p}{q}\right) \pmod{q}.$$

We can now substitute our conclusions above into (19): the equation

$$\prod_{([a],[b])\in H_2} ([a],[b]) = \epsilon \cdot \prod_{[k]\in H_1} ([k],[k])$$

now becomes (again dropping the bracket notation)

$$((-1)^Q, (-1)^{PQ+P}) \equiv \left(\epsilon \cdot (-1)^Q \cdot \left(\frac{q}{p}\right), \epsilon \cdot (-1)^P \cdot \left(\frac{p}{q}\right)\right).$$

From the first coordinate equality, we have

$$(-1)^Q \equiv \epsilon \cdot (-1)^Q \cdot \left(\frac{q}{p}\right) \pmod{p},$$

i.e.,

$$1 \equiv \epsilon \cdot \left(\frac{q}{p}\right) \pmod{p}.$$

Thus

$$p \mid \left(1 - \epsilon \cdot \left(\frac{q}{p}\right)\right);$$

since $\left|1 - \epsilon \cdot \left(\frac{q}{p}\right)\right| \leq 2 < p$, this forces $1 - \epsilon \cdot \left(\frac{q}{p}\right) = 0$, and so

$$1 = \epsilon \cdot \left(\frac{q}{p}\right),$$

whence we have

$$\epsilon = \left(\frac{q}{p}\right).$$

A similar argument on the second coordinate equality shows that

$$\epsilon = (-1)^{PQ} \cdot \left(\frac{p}{q}\right).$$

Equating these two expressions for ϵ, we conclude that

$$\left(\frac{q}{p}\right) = (-1)^{PQ} \cdot \left(\frac{p}{q}\right),$$

i.e.,

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

This proves Quadratic Reciprocity. □

Remark 3.2.1. If you did not enjoy this proof, feel free to choose from 344 other proofs here.

Example 3.2.1. As an application of Quadratic Reciprocity, we determine whether the polynomial $f(x) := x^2 - 7$ has a root over $\mathbb{Z}/43\mathbb{Z}$. Since $7 \equiv 43 \equiv 3 \pmod 4$, we have by Quadratic Reciprocity that

$$\left(\frac{7}{43}\right) = -\left(\frac{43}{7}\right).$$

We then check that

$$-\left(\frac{43}{7}\right) = -\left(\frac{1}{7}\right) = -1.$$

We conclude that 7 is not a square modulo 43, i.e. $x^2 - 7$ has no roots in $\mathbb{Z}/43\mathbb{Z}$.

We almost have a formula to compute $\left(\frac{a}{p}\right)$ for any integer a. When a is a product of odd primes, we can split up the Legendre symbol "in the numerator" and use Quadratic Reciprocity as necessary, along with the formula for $\left(\frac{-1}{p}\right)$. However, what do we do when a is even? We have the following result, often referred to as a supplementary law for Quadratic Reciprocity. We will not prove it.

Theorem 3.2.2. [NZM91, Theorem 3.3] *For prime $p > 2$, one has*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

For convenience, we list one more supplementary law, which we saw in §3.1.

Theorem. [NZM91, Theorem 3.1.(5)] *One has for prime $p > 2$ that*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4, \\ -1 & \text{if } p \equiv -1 \pmod 4. \end{cases}$$

Example 3.2.2. Let us determine whether 15 is a quadratic residue modulo 29. We have

$$\left(\frac{15}{29}\right) = \left(\frac{3}{29}\right) \cdot \left(\frac{5}{29}\right);$$

since $29 \equiv 1 \pmod 4$, Quadratic Reciprocity shows that

$$\begin{aligned} \left(\frac{3}{29}\right) \cdot \left(\frac{5}{29}\right) &= \left(\frac{29}{3}\right) \cdot \left(\frac{29}{5}\right) \\ &= \left(\frac{2}{3}\right) \cdot \left(\frac{4}{5}\right) \\ &= (-1) \cdot 1 = -1. \end{aligned}$$

We conclude that 15 is a quadratic *non*residue modulo 29.

In the next section, we will define a generalization of the Legendre symbol, called the *Jacobi symbol*, which will allow us to potentially check for quadratic residues without knowing the complete factorization of the numerator, and perform calculations when the denominator is not necessarily known to be prime.

Exercises. From [NZM91, §3.2], pages $140 - 141$: $\#1 - 11$.

Exercise 3.2.1.

a) List the squares modulo 7, and then the non-squares.
b) Determine all primes p such that $x^2 - 7$ has a root modulo p. Your final answer should include several different congruence conditions on p.

Exercise 3.2.2. Using Quadratic Reciprocity and/or its supplemental laws, prove that for any prime $p > 2$ one has

$$\left(\frac{-2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 3 \pmod 8 \\ -1 & \text{if } p \equiv 5, 7 \pmod 8. \end{cases}$$

Bonus Exercise 3.2.3. This exercise fills in some steps from our proof of Quadratic Reciprocity. Following the notation from the proof, let $p, q > 2$ be distinct primes, and set $P := \frac{p-1}{2}$ and $Q := \frac{q-1}{2}$.

a) Show that
$$\left(\frac{q-1}{2}\right)!^2 \equiv (-1)^{\frac{q-1}{2}} \cdot (q-1)! \pmod q.$$
(*Hint:* for each integer $1 \leq k < q$, one has $k \leq \frac{q-1}{2}$ if and only if $\frac{q-1}{2} < q - k$.)
b) Show that
$$\prod_{Qp < k < \frac{pq}{2}} k \equiv P! \pmod p.$$
(*Hint:* observe that each term of this product has the form $k = \ell + Qp$ where $\ell \geq 1$.)
c) Show that Pq is the greatest multiple of q below $\frac{pq}{2}$.

Bonus Exercise 3.2.4. Show that $5^{1500} \equiv 1 \pmod{3001}$. (*Hint:* you may assume that 3001 is prime.)

3.3. **The Jacobi Symbol.** The main goal of this section is to define the *Jacobi symbol* and prove some of its properties. Previously, we defined the Legendre symbol $\left(\frac{a}{p}\right)$ to determine whether a is a square modulo *an odd prime* p. This time, we will define an analogous symbol to make calculations involving the Legendre symbol simpler, where we do not have to know the complete prime factorization of a.

We know that if an integer $n > 1$ has a prime factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, then we can simplify the Legendre symbol as

$$\left(\frac{n}{p}\right) = \prod_{i=1}^{r} \left(\frac{p_i}{p}\right)^{e_i}.$$

However, if we wish to compute e.g. $\left(\frac{100101}{5001}\right)$, how do we even begin to factor 100101 by hand?

Definition 3.3.1. Let $m > 1$ be an odd number; write its prime factorization as $m = \prod_{i=1}^{r} p_i^{e_i}$. Then for $a \in \mathbb{Z}$, we define the **Jacobi symbol of** a **modulo** m as

$$\left(\frac{a}{m}\right) := \prod_{i=1}^{r} \left(\frac{a}{p_i}\right)^{e_i}.$$

We also define $\left(\frac{a}{1}\right) := 1$.

In particular, the Jacobi symbol is a product of Legendre symbols *based on the factorization of the modulus.* Note that when the modulus is prime, the Jacobi symbol is equivalent to the Legendre symbol.

Example 3.3.1. We make a few Jacobi symbol calculations.

- $\left(\frac{7}{15}\right) = \left(\frac{7}{3}\right) \cdot \left(\frac{7}{5}\right) = \left(\frac{1}{3}\right) \cdot \left(\frac{2}{5}\right) = 1 \cdot (-1) = -1.$
- $\left(\frac{3}{35}\right) = \left(\frac{3}{5}\right) \cdot \left(\frac{3}{7}\right) = -1 \cdot (-1) = 1.$
- $\left(\frac{5}{375}\right) = \left(\frac{5}{3}\right) \cdot \left(\frac{5}{5}\right)^3 = 0.$

While the Legendre symbol $\left(\frac{a}{p}\right)$ tells us whether a is a square modulo p, the Jacobi symbol $\left(\frac{a}{m}\right)$ does not necessarily if m is composite. For example, we see that $\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right) \cdot \left(\frac{2}{5}\right) = (-1) \cdot (-1) = 1$, but the squares modulo 15 are $0, 1, 4, 6, 9, 10$. What, then, is the utility of the Jacobi symbol? As we will soon see, it satisfies similar laws that the Legendre symbol follows without having to factorize the numerator.

First, let us highlight some algebraic properties of the Jacobi symbol.

Theorem 3.3.1. [NZM91, Theorem 3.6] *For odd numbers $m, n > 1$, one has the following for all $a, b \in \mathbb{Z}$.*

(1) $\left(\frac{ab}{m}\right) = \left(\frac{a}{m}\right) \cdot \left(\frac{b}{m}\right)$.
(2) $\left(\frac{a}{mn}\right) = \left(\frac{a}{m}\right) \cdot \left(\frac{a}{n}\right)$.
(3) If $a \equiv b \pmod{m}$ *then* $\left(\frac{a}{m}\right) = \left(\frac{b}{m}\right)$.
(4) If $\gcd(a, m) = 1$ *then* $\left(\frac{a^2}{m}\right) = 1$ *and* $\left(\frac{a}{m^2}\right) = 1$.
(5) If $\gcd(ab, mn) = 1$ *then* $\left(\frac{ab^2}{mn^2}\right) = \left(\frac{a}{m}\right)$.

Proof. Each of these follows from the definition of the Jacobi symbol (which is completely multiplicative in the denominator), as well as the corresponding properties for the Legendre symbol. For example, let us write the factorization $m = \prod_{i=1}^{r} p_i^{e_i}$. To prove (1), we check that

$$\begin{aligned}\left(\frac{ab}{m}\right) &:= \prod_{i=1}^{r}\left(\frac{ab}{p_i}\right)^{e_i} \\ &= \prod_{i=1}^{r}\left(\frac{a}{p_i}\right)^{e_i}\cdot\left(\frac{b}{p_i}\right)^{e_i} \qquad \text{(Legendre symbol, see Theorem 3.1.3 ([NZM91, Theorem 3.1]))} \\ &= \prod_{i=1}^{r}\left(\frac{a}{p_i}\right)^{e_i}\cdot\prod_{i=1}^{r}\left(\frac{b}{p_i}\right)^{e_i} \\ &= \left(\frac{a}{m}\right)\cdot\left(\frac{b}{m}\right).\end{aligned}$$

The proofs for the other parts are similar, utilizing the algebraic properties of the Legendre symbol. □

A remarkable fact about the Jacobi symbol is that it satisfies a Quadratic Reciprocity law similar to the Legendre symbol, along with the supplemental laws.

Theorem 3.3.2. [NZM91, Theorem 3.7 and 3.8] *Let m and n be odd, positive, coprime integers. Then one has*

$$\begin{aligned}\left(\frac{m}{n}\right) &= \left(\frac{n}{m}\right)\cdot(-1)^{\frac{m-1}{2}\cdot\frac{n-1}{2}} \\ &= \begin{cases}\left(\frac{n}{m}\right) & \text{if } m\equiv 1 \pmod 4 \ \textbf{or}\ n\equiv 1 \pmod 4 \\ -\left(\frac{n}{m}\right) & \text{if } m\equiv 3 \pmod 4 \ \textbf{and}\ n\equiv 3 \pmod 4.\end{cases}\end{aligned}$$

Furthermore, one has

$$\left(\frac{-1}{m}\right) = (-1)^{\frac{m-1}{2}} = \begin{cases}1 & \text{if } m\equiv 1 \pmod 4 \\ -1 & \text{if } m\equiv -1 \pmod 4\end{cases}$$

and

$$\left(\frac{2}{m}\right) = (-1)^{\frac{m^2-1}{8}} = \begin{cases}1 & \text{if } m\equiv \pm 1 \pmod 8 \\ -1 & \text{if } m\equiv \pm 3 \pmod 8.\end{cases}$$

We will forgo the proofs of the Quadratic Reciprocity laws for the Jacobi symbol. It is worth noting, however, that these have proofs by induction which utilize the previous laws for Legendre symbols.

Example 3.3.2. For our final example in this chapter, we wish to determine whether 851 is a square modulo the prime 1013. We could compute the Legendre symbol $\left(\frac{851}{1013}\right)$ to determine this – but we would need to factor 851 first. However, with the Jacobi

symbol, we can do the following:

$$\begin{aligned}
\left(\frac{851}{1013}\right) &= \left(\frac{1013}{851}\right) && \text{(QR for Jacobi, since } \gcd(1013, 851) = 1; 1013 \text{ is asssumed prime!)} \\
&= \left(\frac{162}{851}\right) \\
&= \left(\frac{2}{851}\right) \cdot \left(\frac{81}{851}\right) \\
&= -\left(\frac{81}{851}\right) && \text{(supplementary law for Jacobi: } 851 \equiv 3 \pmod 8) \\
&= -\left(\frac{9}{851}\right)^2 && \text{(since } 3 \nmid 851, \text{ by Exercise 3.1.1)} \\
&= -1.
\end{aligned}$$

We conclude that 851 is a quadratic nonresidue modulo 1013, i.e. $x^2 - 851$ has no roots mod 1013.

Remark 3.3.1. With these properties of the Jacobi symbol in mind, you should try and compute $\left(\frac{100101}{5001}\right)$ by hand!

Remark 3.3.2. There is a generalization of the Jacobi symbol that allows one to compute $\left(\frac{a}{m}\right)$ where m is allowed to be even, called the *Kronecker symbol.* In particular, it allows us to define $\left(\frac{a}{2}\right)$ for any $a \in \mathbb{Z}$, extending the Legendre symbol to all primes. For this reason, the Kronecker symbol shows up in many places, especially the study of *Dirichlet characters* and *L-series.*

Exercises. From [NZM91, §3.3], page 147: $\#2 - 4$.

Exercise 3.3.1. Compute the following Legendre/Jacobi symbols.

a) $\left(\frac{51}{71}\right)$.
b) $\left(\frac{-35}{97}\right)$.
c) $\left(\frac{1011}{9907}\right)$, where 9907 is prime.

Exercise 3.3.2. Determine with proof whether the polynomial $x^4 - 36$ has a root modulo the prime $p = 5077$.

5. Some Diophantine Equations

In this course so far, we have focused primarily on techniques for finding solutions to equations modulo m. Questions about solutions modulo m can be hard, but are always "finitely determinable," since there are only finitely many solutions to any congruence in any number of variables. Such solutions are called **local solutions.** For the remainder of this class, we are primarily interested in understanding **global solutions,** particularly solutions to polynomials over $\mathbb{Z}$ which lie in $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$ rather than $\mathbb{Z}/m\mathbb{Z}$. We will see that we can leverage our knowledge of local solutions to help us determine global solutions.

The goal of this chapter is to introduce the problem of studying global solutions to *Diophantine equations.* We are particularly interested in understanding solutions in $\mathbb{Z}$ and $\mathbb{Q}$ to *plane curves.* This will culminate in us studying *elliptic curves,* and understanding their group structure.

5.0. **A Dictionary for Diophantine Geometry.** The main goal of this section is to define some of the terms we will use throughout this chapter.

Recall that for a set X, for each integer $n > 0$ we write

$$X^n := \underbrace{X \times X \times \ldots \times X}_{n \text{ times}}.$$

We will attempt to understand solutions to polynomial equations in n variables, and are thus led to study points in X^n for specific X. Towards this, let us recall the following distinguished sets.

- $\mathbb{Z}$ is the ring of integers.
- $\mathbb{Q}$ is the field of rational numbers.
- $\mathbb{R}$ is the field of *real numbers* (such as $\pi, 2, e^1, \ln(2), \frac{1}{3}$).
- $\mathbb{C}$ is the field of *complex numbers* (such as $i, e^1, \pi + 3i, 7, e^{2\pi i/7}$).

One has $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

Definition 5.0.1. A **Diophantine equation** is a polynomial equation over $\mathbb{Z}$. Such equations can be written as

$$f(x_1, x_2, \ldots, x_n) = 0$$

for some polynomial $f \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$. For such an equation, a **rational solution** is a point $(a_1, a_2, \ldots, a_n) \in \mathbb{Q}^n$ with

$$f(a_1, a_2, \ldots, a_n) = 0.$$

If each a_i is an integer, then we call this point an **integral/integer solution to f,** or an **integral/rational point on f.** In general, if each a_i is contained in a ring R (such as $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$), we say that the point is a *solution over R.*

Example 5.0.1. Here are some examples of Diophantine equations.

- Define $f(x, y) \in \mathbb{Z}[x, y]$ by

$$f(x, y) := x^2 + y^2 - 1.$$

  This defines the unit circle in the x, y-plane $\mathbb{R}^2$.

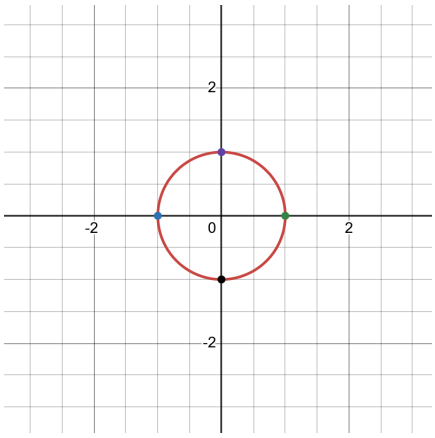


FIGURE 1. The unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$.

What are its integral solutions? They are integer points $(a, b) \in \mathbb{Z}^2$ with

$$f(a, b) = 0,$$

i.e.,

$$a^2 + b^2 = 1.$$

We have that $(\pm 1, 0)$ and $(0, \pm 1)$ are the only integral solutions – try and convince yourself of this! The unit circle also has an *infinite* amount of rational solutions, which will be a consequence of our work in §5.3.

- Define the polynomial

$$g(x, y) := y - (2x + 7).$$

This defines a line in the real plane.

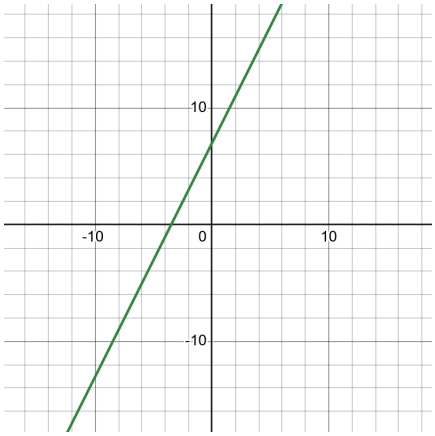


FIGURE 2. The line $y = 2x + 7$ in $\mathbb{R}^2$.

One can easily show there are infinitely many integral points on this line: they are of the form $(n, 2n + 7)$ where $n \in \mathbb{Z}$.

- Let
$$h(x_1, x_2, \ldots, x_n) := x_1 \cdot x_2 \cdots x_n - 1.$$
If $(a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n$ is an integral solution, then
$$a_1 a_2 \cdots a_n = 1,$$
which forces each $a_i = \pm 1$, as well as an even number of negative a_i's.
- Consider
$$k(x, y) := y^2 - (x^3 - x).$$
This equation defines an *elliptic curve*.

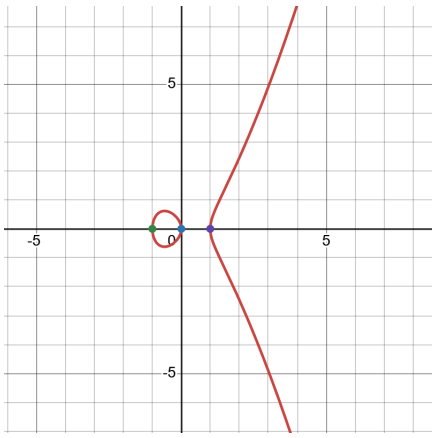


FIGURE 3. The elliptic curve $y^2 = x^3 - x$ in $\mathbb{R}^2$.

An integral solution (a, b) satisfies
$$b^2 = a^3 - a.$$
We have the integral solutions $(0, 0), (1, 0)$ and $(-1, 0)$. In fact, these are the only *rational* solutions – this is not obvious! However, real solutions are abundant: for any $r \in \mathbb{R}$ with $r^3 - r > 0$, the points $(r, \pm\sqrt{r^3 - r}) \in \mathbb{R}^2$ are solutions.

We are most interested in Diophantine equations in two variables; these will define *plane curves over* $\mathbb{R}$.

Definition 5.0.2. Given a polynomial $f(x, y) \in \mathbb{R}[x, y]$, we say that f defines a **plane curve over** $\mathbb{R}$**.** Denoting this curve by C (sometimes C_f), we also use notation
$$C : f(x, y) = 0$$
and realize C as the *graph* of f over $\mathbb{R}$. Thus, the curve C is the set of real solutions to $f(x, y)$.

When f is defined over $\mathbb{Z}$, we write $C/\mathbb{Q}$, and can consider integral and rational points on C. We write its set of integral and rational points as $C(\mathbb{Z})$ and $C(\mathbb{Q})$, respectively.

Example 5.0.2. Let us revisit some of our examples from a moment ago.

- We have the unit circle
$$S : x^2 + y^2 = 1.$$
We saw that $S(\mathbb{Z}) = \{(\pm 1, 0), (0, \pm 1)\}$, and claimed $\#S(\mathbb{Q}) = \infty$.
- We have a line
$$L : y = 2x + 7.$$
We convinced ourselves that $\#L(\mathbb{Z}) = \#L(\mathbb{Q}) = \infty$.
- We have an elliptic curve
$$E : y^2 = x^3 - x.$$
We saw that $E(\mathbb{Z})$ contains $(0, 0), (\pm 1, 0)$. It can be proven that $E(\mathbb{Q}) = E(\mathbb{Z})$ (which is not obvious!).

A polynomial $f \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$ always has a finite amount of solutions modulo $m > 0$. However, it can have *infinitely* many integral solutions. Despite this, there is no "one size fits all" algorithm to determine whether f has an integral solution; this is related to *Hilbert's Tenth Problem.* Similarly, there is no known algorithm to determining rational solutions to a general f.

In this chapter, we will study several different kinds of Diophantine equations, including Fermat equations and those which define plane curves, and study their integral and rational points; this will culminate in us spending considerable time studying elliptic curves. Sometimes, we will be able to use our "local" techniques (modular arithmetic) to say something about our "global" solutions (in particular, integral and rational points).

Let us end this section with an example of using local techniques to determine whether integral solutions to a curve exist.

Example 5.0.3. [NZM91, Theorem 5.6] Consider the curve
$$C : 15x^2 - 7y^2 = 9.$$
This defines a hyperbola in $\mathbb{R}^2$.

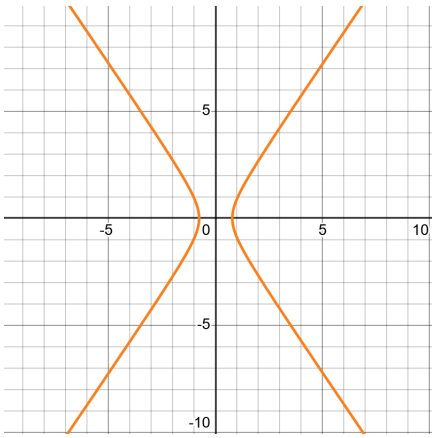


FIGURE 4. The hyperbola $C : 15x^2 - 7y^2 = 9$ in $\mathbb{R}^2$.

We will show that C has no integral points, i.e. $C(\mathbb{Z}) = \emptyset$. For the sake of contradiction, suppose there exists $(a, b) \in C(\mathbb{Z})$; then

$$15a^2 - 7b^2 = 9. \tag{21}$$

Reducing this equation modulo 9 gives

$$-7b^2 \equiv 0 \pmod 9,$$

i.e.,

$$2b^2 \equiv 0 \pmod 9.$$

Thus $9 \mid 2b^2$, which implies that $9 \mid b^2$. Then reducing (21) modulo 9 shows that

$$15a^2 \equiv 0 \pmod 9,$$

so that $9 \mid 15a^2$. This implies that $3 \mid a^2$, and thus $3 \mid a$, so that $9 \mid a^2$. Therefore 9 divides both a^2 and b^2, so dividing (21) by 9 gives

$$15c^2 - 7d^2 = 1$$

where $c := \frac{a^2}{9}$ and $d := \frac{b^2}{9}$ are integers. However, we reduce this new equation modulo 3 and find that

$$2d^2 \equiv 1 \pmod 3,$$

which is impossible since $d^2 \equiv 0, 1 \pmod 3$. We conclude that $C(\mathbb{Z}) = \emptyset$.

Exercises. From [NZM91, §5.4], page 239: $\#1 - 2, 5 - 6$.

Exercise 5.0.1. Show that the Diophantine equation

$$x^2 + y^2 = 9z + 6$$

has no integral solutions.

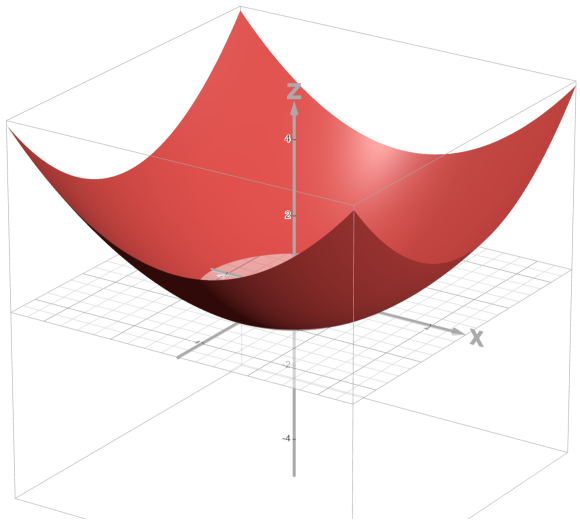


FIGURE 5. The *quadric surface* $x^2 + y^2 = 9z + 6$, pictured in $\mathbb{R}^3$.

Exercise 5.0.2. Show that the Diophantine equation

$$x^8 + 1 = 7y$$

has no integral solutions. However, demonstrate that it has infinitely many rational solutions.

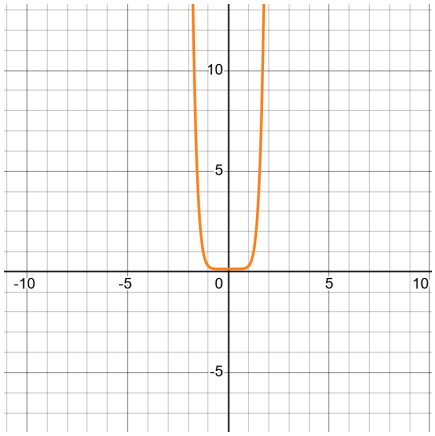


FIGURE 6. The plane curve $C : x^8 + 1 = 7y$, pictured in $\mathbb{R}^2$.

Exercise 5.0.3. Determine all primes p such that the equation

$$x^2 - y^2 = p$$

has integral solutions. For those p which it has integral solutions, determine how many such solutions there are.

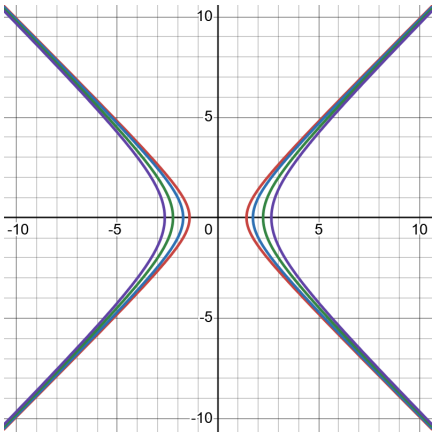


FIGURE 7. The hyperbolas $C_p : x^2 - y^2 = p$ for $p = 2, 3, 5$ and 7, pictured in $\mathbb{R}^2$.

Bonus Exercise 5.0.4. Show that the polynomial $f(x) := (x^2 - 2)(x^2 - 17)(x^2 - 34)$ has no solution over $\mathbb{Z}$, but a solution over $\mathbb{Z}/n\mathbb{Z}$ for each $n \in \mathbb{Z}^+$. (*Hint:* use CRT and Hensel's Lemma to reduce this problem to finding solutions over $\mathbb{Z}/p\mathbb{Z}$ for prime p, and then study quadratic residues.)

5.1. **The Equation $ax + by = c$.** Our main goal for this section is to prove the *Linear Diophantine Theorem,* which characterizes integral solutions to a line over $\mathbb{Z}$. The simplest nontrivial case of a Diophantine plane curve is where the defining equation $f(x, y) \in \mathbb{Z}[x, y]$ is a *line:*

$$L : ax + by = c$$

where $a, b, c \in \mathbb{Z}$.

The existence of integral solutions to lines depend on the equation. Here are some examples to highlight this:

Example 5.1.1.

- Consider the line
$$L_1 : x + y = 1.$$
What are its integral points? The defining equation can be written as
$$y = 1 - x,$$
so there are clearly infinitely many integral solutions, given in the form $(n, 1 - n) \in \mathbb{Z}^2$.
- Consider the line
$$L_2 : 2x + 4y = 5.$$
This line has *no* integral points: if such a solution $(a, b) \in \mathbb{Z}^2$ exists, then
$$2a + 4b = 5,$$
but reducing mod 2 then shows that $0 \equiv 1 \pmod 2$, which is impossible. Here, an obstruction to an integral solution was the fact that $\gcd(2, 4) \nmid 5$.

In light of the above, it turns out that decidability of integral solutions to planar lines is completely determinable, by the *Linear Diophantine Theorem.*

Theorem 5.1.1 (Linear Diophantine Theorem)**.** [NZM91, Theorem 5.1] *Fix integers a, b and c where $a \neq 0$ or $b \neq 0$. Then the line*

$$L : ax + by = c$$

has an integral solution if and only if $\gcd(a, b) \mid c$. When this happens, the line has infinitely many integral points. Furthermore, if $(x_1, y_1) \in \mathbb{Z}^2$ is any solution, then all other integral solutions are of the form

$$(x_2, y_2) = \left(x_1 + k \cdot \frac{b}{\gcd(a, b)}, y_1 - k \cdot \frac{a}{\gcd(a, b)} \right)$$

where $k \in \mathbb{Z}$.

Remark 5.1.1. Observe the similarities between this theorem and the Linear Congruence Theorem from §2.2 ([NZM91, Theorem 2.17]), where $ax \equiv b \pmod m$ has a solution if and only if $\gcd(a, m) \mid b$.

Proof. $\Rightarrow$: suppose that $(x_1, y_1) \in \mathbb{Z}^2$ satisfies

$$ax_1 + by_1 = c.$$

Then $\gcd(a, b)$ divides the left hand side, and thus $\gcd(a, b) \mid c$.

$\Leftarrow$: suppose that $\gcd(a,b) \mid c$; let us write $c = \gcd(a,b) \cdot k$ for some $k \in \mathbb{Z}$. Then we have

$$k = \frac{c}{\gcd(a,b)}.$$

Recall from §1.2 that $\gcd(a,b)$ can be expressed as a $\mathbb{Z}$-linear combination of a and b: let us write

$$ax_0 + by_0 = \gcd(a,b)$$

for some $x_0, y_0 \in \mathbb{Z}$. Multiply both sides by $k = \frac{c}{\gcd(a,b)}$ to get

$$a(kx_0) + b(ky_0) = \gcd(a,b)k = c.$$

We deduce that $(kx_0, ky_0) \in \mathbb{Z}^2$ is a point on L. This proves the "if and only if."

Next, suppose that (x_1, y_1) is a fixed integral solution to L. If (x_2, y_2) is another integral solution, then we have both

$$ax_1 + by_1 = c$$

and

$$ax_2 + by_2 = c,$$

and thus

$$a(x_1 - x_2) + b(y_1 - y_2) = 0,$$

i.e.,

$$a(x_1 - x_2) = -b(y_1 - y_2).$$

Dividing both sides by $\gcd(a,b)$, we get

$$\frac{a}{\gcd(a,b)} \cdot (x_1 - x_2) = -\frac{b}{\gcd(a,b)} \cdot (y_1 - y_2). \tag{22}$$

Thus

$$\frac{a}{\gcd(a,b)} \mid \frac{b}{\gcd(a,b)} \cdot (y_1 - y_2);$$

since $\frac{a}{\gcd(a,b)}$ and $\frac{b}{\gcd(a,b)}$ are coprime, this implies that

$$\frac{a}{\gcd(a,b)} \mid (y_1 - y_2),$$

and thus

$$y_1 - y_2 = k \cdot \frac{a}{\gcd(a,b)}$$

for some $k \in \mathbb{Z}$, which can be written as

$$y_2 = y_1 - k \cdot \frac{a}{\gcd(a,b)}. \tag{23}$$

On the other hand, plugging this into (22), we see that

$$\frac{a}{\gcd(a,b)} \cdot (x_1 - x_2) = -\frac{b}{\gcd(a,b)} \cdot k \cdot \frac{a}{\gcd(a,b)},$$

hence dividing both sides by $\frac{a}{\gcd(a,b)}$ gives

$$x_1 - x_2 = -k \cdot \frac{b}{\gcd(a,b)},$$

i.e.,

$$x_2 = x_1 + k \cdot \frac{b}{\gcd(a,b)}.$$

Combining this with (23), we conclude that our formula for (x_2, y_2) holds. □

Following the Linear Diophantine Theorem, given a line

$$L : ax + by = c$$

where $\gcd(a,b) \mid c$, an obvious question is how to find the "first integral solution" (x_1, y_1) on L to characterize the other integral solutions. Based on our proof, we have an algorithm for this.

1. Find a solution to the "GCD line"
$$ax + by = \gcd(a,b);$$
we denote this GCD line solution by
$$(x_0, y_0).$$
2. Plug (x_0, y_0) into the GCD line, and multiply both sides by $\frac{c}{\gcd(a,b)}$ to get
$$a\left(x_0 \cdot \frac{c}{\gcd(a,b)}\right) + b\left(y_0 \cdot \frac{c}{\gcd(a,b)}\right) = c.$$
Thus, a solution to the original line L is
$$(x_1, y_1) := \left(x_0 \cdot \frac{c}{\gcd(a,b)}, y_0 \cdot \frac{c}{\gcd(a,b)}\right).$$
3. The Linear Diophantine Theorem then implies that all other integral solutions $(x_2, y_2) \in L(\mathbb{Z})$ have the form
$$\begin{aligned}(x_2, y_2) &= \left(x_1 + k \cdot \frac{b}{\gcd(a,b)}, y_1 - k \cdot \frac{a}{\gcd(a,b)}\right) \\ &= \left(x_0 \cdot \frac{c}{\gcd(a,b)} + k \cdot \frac{b}{\gcd(a,b)}, y_0 \cdot \frac{c}{\gcd(a,b)} - k \cdot \frac{a}{\gcd(a,b)}\right),\end{aligned}$$
where $k \in \mathbb{Z}$.

The upshot to our algorithm is that the problem of finding integral points on a Diophantine line $L : ax + by = c$ when $\gcd(a,b) \mid c$ reduces to finding solutions on the associated GCD line $ax + by = \gcd(a,b)$, which can be done by writing the GCD as a $\mathbb{Z}$-linear combination of a and b using e.g. the Euclidean Algorithm or Blankinship's Algorithm.

Example 5.1.2. We would like to determine all integral solutions to the line

$$L : 6x + 9y = 21,$$

if they exist. Here $a := 6, b := 9$ and $c := 21$, so that $\gcd(a,b) = 3$. Since $3 \mid 21$, by the Linear Diophantine Theorem we conclude there are infinitely many integral solutions.

Let us follow our algorithm to characterize all integral solutions. Our associated GCD line is

$$6x + 9y = 3.$$

We can check that $(x_0, y_0) := (2, -1)$ lies on the GCD line. Therefore, a solution to the original line $L : 6x + 9y = 21$ is given by

$$(x_1, y_1) = \left(x_0 \cdot \frac{c}{\gcd(a,b)}, y_0 \cdot \frac{c}{\gcd(a,b)} \right) = (2 \cdot 7, -1 \cdot 7) = (14, -7).$$

Therefore, every integral solution on L has the form

$$\begin{aligned}(x_2, y_2) &= \left(x_1 + k \cdot \frac{b}{\gcd(a,b)}, y_1 - k \cdot \frac{a}{\gcd(a,b)} \right) \\ &= (14 + 3k, -7 - 2k)\end{aligned}$$

for any $k \in \mathbb{Z}$.

Example 5.1.3. How many integral points does the line

$$L : 216x + 135y = 100$$

have? By the Linear Diophantine Theorem, there are either zero integral solutions, or infinitely many; this depends on whether $\gcd(135, 216)$ divides 100. We can use Blankinship's Algorithm to calculate the GCD:

$$\begin{aligned}\left[\begin{array}{c|cc} 216 & 1 & 0 \\ 135 & 0 & 1 \end{array}\right] &\xrightarrow{A \mapsto A - B \cdot 1} \left[\begin{array}{c|cc} 81 & 1 & -1 \\ 135 & 0 & 1 \end{array}\right] \\ &\xrightarrow{B \mapsto B - A \cdot 1} \left[\begin{array}{c|cc} 81 & 1 & -1 \\ 54 & -1 & 2 \end{array}\right] \\ &\xrightarrow{A \mapsto A - B \cdot 1} \left[\begin{array}{c|cc} 27 & 2 & -3 \\ 54 & -1 & 3 \end{array}\right].\end{aligned}$$

Since $27 \mid 54$, we deduce that $\gcd(216, 135) = 27 = 3^3$. Since $3 \nmid 100$, we conclude that $L(\mathbb{Z}) = \emptyset$.

Remark 5.1.2. One can generalize our Linear Diophantine Theorem to characterize solutions to Diophantine *planes* in $\mathbb{R}^n$, i.e., linear equations in $n > 2$ variables. This is done in [NZM91, §5.2]; we will not cover this in our course.

Exercises. From [NZM91, §5.1], pages $218 - 219$: $\#2 - 12$.

Exercise 5.1.1. Determine whether the following lines have integral solutions. If they do, then give a complete description of them.

a) $L_1 : 10x - 7y = 17$.
b) $L_2 : 903x + 731y = 60$.
c) $L_m : mx + (m+1)y = 10$, where $m > 0$ is a fixed integer.
d) $L_n : (n-1)x + (n+1)y = 4573$, where $n > 1$ is a fixed odd integer.

Exercise 5.1.2.

a) Show that the line

$$L : ax + by = c$$

has an integral point if and only if for any $n \in \mathbb{Z}$ the line

$$L_n : ax + by = na + c$$

has an integral point. Briefly argue that this still holds if we replace na with nb in L_n.

b) Use part a) to show that the line

$$L : 100001x + 1010010001000001y = 1000010$$

has infinitely many integral solutions.

5.3. **Pythagorean Triangles.** The main goal of this section is to characterize the integral solutions to the "Pythagorean equation" $x^2+y^2=z^2$. As we continue studying integral and rational solutions to Diophantine equations, the next natural class of curves are those defined by quadratic equations; these are called *conics.* The conic we study is the familiar equation

$$F_2 : x^2 + y^2 = z^2.$$

This is in three variables, but can be realized as a plane curve by dividing both sides by z:

$$\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 = 1.$$

Thus, we treat this equation on equal footing as plane curves.

The equation $x^2 + y^2 = z^2$ naturally appears when studying right triangles, via the *Pythagorean Theorem:* for $a, b, c \in \mathbb{R}$ which are the side lengths of a right triangle with c as the hypotenuse, one has

$$a^2 + b^2 = c^2;$$

thus (a, b, c) is a solution to F_2. The integral solutions to $F_2(x, y, z)$ where $a, b, c > 0$ are called **Pythagorean triples.** Some examples of Pythagorean triples include $(3, 4, 5), (5, 12, 13)$, etc.

It is easy to construct Pythagorean triples: given any Pythagorean triple (a, b, c), one has for each $k \in \mathbb{Z}^+$ that (ak, bk, ck) is also a Pythagorean triple. Determining Pythagorean triples becomes much more interesting when one imposes coprimality conditions on a, b and c.

Definition 5.3.1. A Pythagorean triple (a, b, c) is called **primitive** if $\gcd(a, b, c) = 1$.

Remark 5.3.1. It is worth noting that a primitive Pythagorean triple (a, b, c) must have that a, b and c are pairwise coprime, as per the equation $a^2 + b^2 = c^2$.

As it turns out, primitive Pythagorean triples can be completely characterized.

Theorem 5.3.1. [NZM91, Theorem 5.5] *A point $(x, y, z) \in \mathbb{Z}^3$ where y is even is a primitive Pythagorean triple if and only if*

$$x = r^2 - s^s, y = 2rs, z = r^2 + s^2$$

for some coprime $r, s \in \mathbb{Z}^+$ with opposite parity such that $r > s$.

Remark 5.3.2. One can show that for any primitive Pythagorean triple (x, y, z), the parity of x and y must be different: if x and y are both odd, then $x^2 \equiv y^2 \equiv 1 \pmod 4$, and so $z^2 \equiv 2 \pmod 4$, which is impossible. If both x and y are even, then so is z, which contradicts primitivity. Thus x and y have different parity. Since the equation $x^2 + y^2 = z^2$ is symmetric in x and y, to simplify the statement of our result we assume that y is even.

Before we prove this theorem, let us prove one lemma.

Lemma 5.3.2. [NZM91, Lemma 5.4] *If coprime integers u and v are such that uv is a perfect square, then so are u and v.*

Proof. Let us write $uv = a^2$ for some $a \in \mathbb{Z}^+$. We will show that if a prime power $p^e \parallel u$, then e is even. We know that $p^e \mid uv = a^2$. Since $\gcd(u, v) = 1$, we have $p \nmid v$, and thus $p^e \parallel a^2$. This forces e to be even. An identical argument shows that v is a perfect square. □

Proof of Theorem 5.3.1. Assume that (x, y, z) is a primitive Pythagorean triple (so $x, y, z > 0$) where y is even. Then from

$$x^2 + y^2 = z^2,$$

we find that

$$x \equiv z \pmod 2,$$

so that x and z have the same parity. Since x and y have different parity (see Remark 5.3.2) x and z must both be odd. Thus $z + x$ and $z - x$ are even.

Next, from

$$x^2 + y^2 = z^2$$

we see that

$$y^2 = z^2 - x^2 = (z + x) \cdot (z - x),$$

and so

$$\left(\frac{y}{2}\right)^2 = \left(\frac{z+x}{2}\right) \cdot \left(\frac{z-x}{2}\right).$$

If d divides both $\frac{z+x}{2}$ and $\frac{z-x}{2}$, then it divides their sum z and their difference x, and thus $d \mid \gcd(x, z)$. Then $d \mid \gcd(x, y, z)$ by Remark 5.3.1, and thus $d = 1$. Therefore $\frac{z+x}{2}$ and $\frac{z-x}{2}$ are coprime. Since their product $\left(\frac{y}{2}\right)^2$ is a perfect square, they are both perfect squares by Lemma 5.3.2 above ([NZM91, Lemma 5.4]). Thus, we can write

$$\frac{z+x}{2} = r^2$$

and

$$\frac{z-x}{2} = s^2$$

for some $r, s \in \mathbb{Z}^+$.

Let us collect some facts about r and s:

- $x = 2r^2 - z$ and $z = 2s^2 + x$, so that
$$x = 2r^2 - 2s^2 - x$$
and thus
$$x = r^2 - s^2,$$
as well as
$$z = r^2 + s^2.$$
Since $y^2 = z^2 - x^2 = 4r^2s^2$, one takes square roots to get $y = 2rs$, as desired.
- Since $\frac{z+x}{2}$ and $\frac{z+x}{2}$ are coprime, so are r and s.
- Since $z + x > z - x > 0$, we have $r > s$.
- Since $z = r^2 + s^2$ is odd, we have $r^2 + s^2 \equiv 1 \pmod 2$, i.e. $r + s \equiv 1 \pmod 2$, so that r and s have opposite parity.

Finally, one can show that any pair $(r, s) \in (\mathbb{Z}^+)^2$ where $r > s$ are coprime and have opposite parity, will induce a primitive Pythagorean triple $(r^2 - s^2, 2rs, r^2 + s^2)$. $\square$

Example 5.3.1. It is now easy to produce lots of primitive Pythagorean triples. For example, if we take $r := 10$ and $s := 9$, then r and s are coprime, have opposite parity and $r > s$. The theorem then implies we have the primitive Pythagorean triple

$$(x, y, z) = (19, 180, 181).$$

We can double-check that $19, 180$ and 181 have simultaneous GCD equal to 1, and

$$19^2 + 180^2 = 181^2.$$

Let us remark that the equation

$$F_2 : x^2 + y^2 = z^2$$

is part of a special class of Diophantine equations called the *Fermat equations*

$$F_n : x^n + y^n = z^n,$$

where $n \in \mathbb{Z}^+$. We just showed that this has infinitely many (primitive) integral solutions when $n = 2$. When $n = 1$, this is a linear equation, which we have determined has solutions by §5.1.

However, when $n \geq 3$ the equation F_n has no solutions $(a, b, c) \in (\mathbb{Z}^+)^3$ – this is known as **Fermat's Last Theorem,** which is one of the most important results in mathematics. A proof of this theorem came out in 1994, which is 357 years after it was originally stated in 1637. In 1637, Pierre de Fermat claimed he had a proof of this result "too large to fit in the margin" of his copy of *Arithmetica* (a book written by Diophantus). It was likely that the proof he had in mind was incorrect. The proof from 1994 uses high level *arithmetic geometry,* which would take many years of study to properly understand. A key idea in the proof is to show that the existence of an integral solution $(a, b, c) \in (\mathbb{Z}^+)^3$ to F_n implies the existence of a certain *elliptic curve* $E_{a,b,c}$ which has self-contradicting properties.

Remark 5.3.3. There is an interesting idea in the proof of Theorem 5.3.1 that generalizes to other Diophantine equations: factoring solutions over the real or complex numbers. In our proof, we turned an equation

$$x^2 + y^2 = z^2$$

for $x, y, z \in \mathbb{Z}^+$ into

$$y^2 = z^2 - x^2 = (z + x)(z - x).$$

For another example, an integral equation

$$x^2 + 2 = y^3$$

with $x, y \in \mathbb{Z}$ and $xy \neq 0$ implies that over $\mathbb{C}$ one has

$$y^3 = (x + \sqrt{-2})(x - \sqrt{-2}).$$

Thus, there is a solution which lies inside the *algebraic number ring*

$$\mathbb{Z}[\sqrt{-2}] := \{a + b\sqrt{-2} : a, b \in \mathbb{Z}\};$$

with additional work, one can show that this implies $(x, y) = (3, \pm 5)$ (see [NZM91, §9.9]). This general technique/philosophy of lifting solutions over $\mathbb{Z}$ to solutions over algebraic number rings R is extremely useful, and is used in the proof of Fermat's Last Theorem, for example.

Exercises. From [NZM91, §5.3], page 233: $\#1, 4 - 8$.

Exercise 5.3.1. Show that every Pythagorean triple (x, y, z) is such that 3 divides (at least) one of x, y, z and 5 divides (at least) one of x, y, z.

Bonus Exercise 5.3.2. $99.\overline{9}\%$ of people cannot solve this one! For  $\in \mathbb{Z}$ with  ≥ 3, find all  ,  , $\in \mathbb{Z}^+$ with

(*Hint:* FLT.)

5.6. **Rational Points on Curves.** In this section, we will begin the study of determining rational points on Diophantine curves. While this is technically "easier" than finding integral points, it is still generally difficult – in fact, a large part of modern number theory is devoted to this problem.

Here are the main goals for this section:

- Define what a *curve* is.
- Understand the algorithm for *parametrizing* rational points on *conics,* i.e., *quadratic* curves.
- See how the algorithm for conics generalizes to *cubic* curves, via the *chord and tangent method.*
- Towards understanding properties of curves, define intersection multiplicity, singular points, the projective plane and irreducibility.

Definition 5.6.1. A **plane curve,** or **algebraic curve** (often simply called a **curve**) is the subset $C \subseteq \mathbb{R}^2$ of points (x, y) that are solutions to some fixed polynomial $f(x, y) \in \mathbb{R}[x, y]$. In particular C is defined by the equation

$$f(x, y) = 0.$$

Note that $\mathbb{R} \subseteq \mathbb{C}$; we will write $C(\mathbb{R})$ for the set of *real* solutions, and $C(\mathbb{C})$ for the set of *complex* solutions. When we write $P \in C$, we usually mean $P \in C(\mathbb{C})$.

When C is defined over $\mathbb{Q}$, we write $C(\mathbb{Q})$ for the *rational* solutions; and if C is defined by a polynomial $f \in \mathbb{Z}[x, y]$, we write $C(\mathbb{Z})$ for the *integral* solutions. We sometimes write $C/\mathbb{Q}$, $C/\mathbb{R}$ or $C/\mathbb{C}$ to emphasize the field f is defined over. More generally, for a field F we will write C/F when $f \in F[x, y]$; we also call a point P in $C(F)$ an **F-rational point;** thus, "F-rational" means "defined over F." As before, when we just say "rational," without qualification, we will mean $\mathbb{Q}$-rational.

Remark 5.6.1. Since this is a number theory class, it is almost always the case that our polynomials $f(x, y)$ are defined over $\mathbb{Z}$ or $\mathbb{Q}$. For technical reasons from *algebraic geometry,* even if f is defined over $\mathbb{Z}$, we will write $C_f/\mathbb{Q}$ instead of $C_f/\mathbb{Z}$.

Example 5.6.1.

- The curve

$$S : x^2 + y^2 = 1$$

is the unit circle centered at the origin $(0, 0)$.

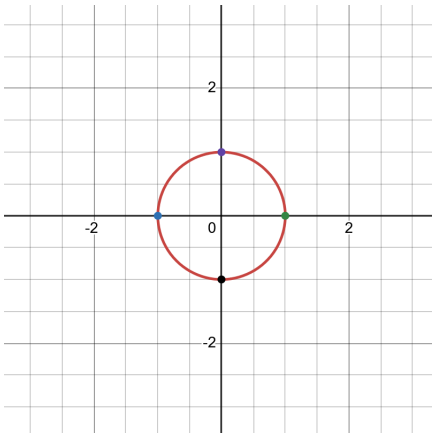


FIGURE 8. The unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$.

We saw last time that $S(\mathbb{Z}) = \{(0,1),(1,0),(-1,0),(0,-1)\}$. However, we can also convince ourselves that $\#S(\mathbb{Q}) = \infty$.

- The curve
$$C : x^2 + y^2 = -1$$
has no real points, i.e. $C(\mathbb{R}) = \emptyset$. However, we can check that $C(\mathbb{C})$ is an infinite set.
- The curve
$$E : y^2 = x^3 - 2x + 2$$
is an *elliptic curve.* We spot two integral points $(1, \pm 1)$ on it.

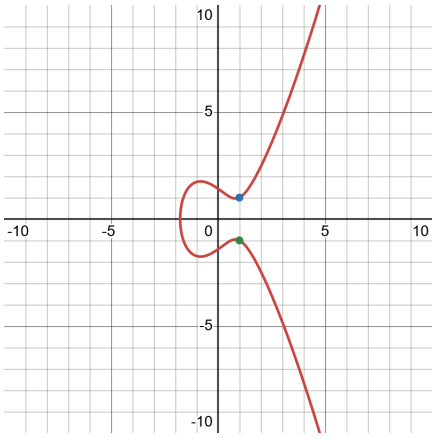


FIGURE 9. The elliptic curve $y^2 = x^3 - 2x + 2$ in $\mathbb{R}^2$.

Compared to our elliptic curve example in §5.0 (see Example 5.0.1), this curve has a single "connected component." Furthermore, the *only* integral points on E are $(1, \pm 1)$, and yet $\#E(\mathbb{Q}) = \infty$ – these are not obvious facts!

Definition 5.6.2. For a polynomial $f \in \mathbb{R}[x_1, \dots x_n]$, we can express it as a sum of monomials that are products of variables:

$$f(x_1, \dots, x_n) = \sum_{i_1,\dots,i_n \geq 0} a_{i_1,\dots,i_n} \cdot x_1^{i_1} \cdot x_2^{i_2} \cdots x_n^{i_n}$$

where each coefficient $a_{i_1,\dots,i_n} \in \mathbb{R}$. We define the **(total) degree of** f as the largest exponent $i_1 + \dots + i_n$ among monomials $x_1^{i_1} \cdot x_2^{i_2} \cdots x_n^{i_n}$ with $a_{i_1,\dots,i_n} \neq 0$. We write this degree as $\deg(f)$. For each $1 \leq i \leq r$, we also have the **x_i-degree of f,** which is the largest exponent that appears for x_i in f; we write this as $\deg_{x_i}(f)$.

Given a curve C defined by $f \in \mathbb{R}[x, y]$, we define its **degree** as $\deg(C) := \deg(f)$.

- If $\deg(C) = 1$, then we call C a *line.*
- If $\deg(C) = 2$, then we call C a *conic,* or a *quadratic curve.* Such curves include circles, ellipses, parabolas and hyperbolas.
- If $\deg(C) = 3$, then we call C a *cubic curve.* This includes *elliptic curves.*

Example 5.6.2.

- The degree of $f := x_1^2 + x_1x_2x_4 + x_3^2 \in \mathbb{R}[x_1, x_2, x_3, x_4]$ is $\deg(f) = 3$.
- The degree of the curve $C : xy^4 = x^3y^2 + 1$ is $\deg(C) = 5$.
- The curve $D : xy + y = x^3y$ has degree $\deg(D) = 4$.

In this section, we are interested in describing rational points on plane curves of low degree. For *conics* defined over $\mathbb{Q}$, we often have a particularly nice way to construct infinitely many rational points from a single rational point, comparable to the conclusion of the Linear Diophantine Theorem from §5.1. The idea is that if a point on a conic is rational, then any line through that point with a rational slope intersects the conic at another rational point. We illustrate this with an example.

Example 5.6.3. Let us consider the ellipse

$$C : x^2 + 5y^2 = 1.$$

We spot the point $P := (1, 0)$ on this curve (as well as $(-1, 0)$).

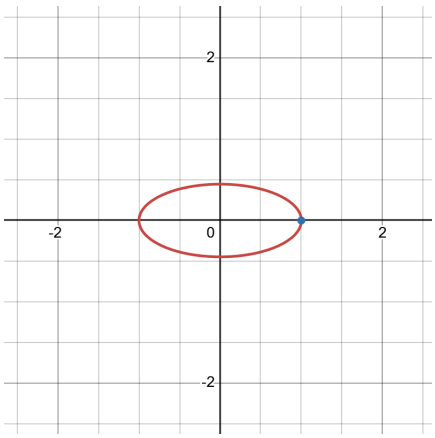


FIGURE 10. The ellipse $C : x^2 + 5y^2 = 1$ in $\mathbb{R}^2$.

We will use P to *completely describe* all other rational points on C. Let $m \in \mathbb{Q}$ be any rational number. Consider the line through P with slope m:

$$L : y - y_1 = m(x - x_1)$$

where $(x_1, y_1) = P$, i.e.,

$$L : y = m(x - 1).$$

Observe that P lies in the set intersection $L \cap C$. We claim that there exists another point in $L \cap C$; intuitively L has degree 1 and C has degree 2, so there "should" be two points in $L \cap C$. (We will discuss this more precisely later.)

Consider a point $Q = (x_2, y_2) \in L \cap C$. Then both

$$Q \in L \Rightarrow y_2 = m(x_2 - 1) \tag{24}$$

and

$$Q \in C \Rightarrow x_2^2 + 5y_2^2 = 1. \tag{25}$$

Let us plug Equation (24) into Equation (25) and solve for x_2. Writing $x = x_2$ to reduce notational clutter, we check that:

$$\begin{aligned} & x^2 + 5(m(x-1))^2 = 1 \\ \Rightarrow\ & x^2 + 5m^2(x-1)^2 - 1 = 0 \\ \Rightarrow\ & x^2 + 5m^2(x^2 - 2x + 1) - 1 = 0 \\ \Rightarrow\ & x^2 + 5m^2x^2 - 10m^2x + 5m^2 - 1 = 0 \\ \Rightarrow\ & (1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1) = 0. \end{aligned}$$

Thus x_2 is a root of

$$p(x) := (1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1).$$

However, we know that $x_1 = 1$ is *also* a root of $p(x)$, since $P = (1, 0)$ is in $L \cap C$ (and thus satisfies Equations (24) and (25)). We then deduce that

$$p(x) = (x - 1) \cdot [(5m^2 + 1)x - (5m^2 - 1)],$$

and so

$$x_2 = \frac{5m^2 - 1}{5m^2 + 1}.$$

This deduction is realized in several different ways:

- Write $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1) = (x - 1)(ax - b)$ for some $a, b \in \mathbb{Q}$, and solve for a and b in terms of m.
- Divide $x - 1$ into $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1)$ using long division.
- Use the Quadratic Formula to compute the roots of $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1)$.

To get the y-coordinate y_2 of the second point Q, we simply plug $x = x_2$ back into the Equation (24) for the line and solve for y_2:

$$y_2 = m(x_2 - 1) = m\left(\frac{5m^2 - 1 - (5m^2 + 1)}{5m^2 + 1}\right) = \frac{-2m}{5m^2 + 1}.$$

We conclude that for each rational number m, we can produce a rational point Q which lies on C, using both $P = (1, 0)$ and m:

$$Q = \left(\frac{5m^2 - 1}{5m^2 + 1}, \frac{-2m}{5m^2 + 1} \right).$$

In a converse direction, given any rational point $(x_2, y_2) \in C(\mathbb{Q})$ not equal to $(1, 0)$, the line through $(1, 0)$ and (x_2, y_2) has slope $m = \frac{y_2}{x_2 - 1} \in \mathbb{Q}$. We conclude that there exists a *bijection* between $\mathbb{Q}$ and $C(\mathbb{Q}) \smallsetminus \{(1, 0)\}$.

The curve in the example above is called a **parametrizable** curve: there is a formula for all but one point in terms of rational numbers. If in our example we write $m = \frac{r}{s}$, then the corresponding description of rational points (sans including $(1, 0)$) is

$$(x_2, y_2) = \left(\frac{5r^2 - s^2}{5r^2 + s^2}, \frac{-2rs}{5r^2 + s^2} \right)$$

where $r, s \in \mathbb{Z}$ with $(r, s) \neq (0, 0)$. This looks similar to our parametrization of primitive solutions to the Fermat curve $F_2 : x^2 + y^2 = z^2$ from §5.3.

Remark 5.6.2. In our example above, we can say that $(1, 0)$ corresponds to $m := \infty$, the slope of the *tangent line* T_P of C at P, which "passes through P twice" (more on this soon). Then our bijection $\mathbb{Q} \to C(\mathbb{Q}) \smallsetminus \{(1, 0)\}$ extends to a bijection

$$\mathbb{Q} \cup \{\infty\} \to C(\mathbb{Q}).$$

The technique in our example above for producing infinitely many rational points works because we were considering a *quadratic* curve. This meant that our intersections $L \cap C$ had at most two points in them; this led us to analyzing the quadratic polynomial $p(x) := f(x, mx+b)$ where $L : y = mx+b$. Since our starting point P is rational and lies in $L \cap C$, and since $p(x)$ has rational coefficients (from L and C both being rational), this forced *both* roots of $p(x)$ to be rational. The second root x_2 then gave us the rational point $Q = (x_2, y_2)$ where $y_2 = mx_2 + b$, from the line equation.

There are subtleties that can arise when attempting to generalize this technique to other curves, involving *intersection multiplicities* and *singularities*. Let us define these terms, and then create an algorithm from our example above.

Remark 5.6.3. One can skip the details on intersection multiplicity when first learning about rational points on curves. However, it will clarify why *nonsingular points* are simpler to work with, and prove especially useful when understanding *flex points* on elliptic curves in the next section.

Before defining intersection multiplicity, let us recall some facts about polynomial division.

Definition 5.6.3. Let R be a ring. For $f, g \in R[x]$, say that g **divides** f, and write $g \mid f$, if $f = g \cdot h$ for some $h \in R[x]$. For $r \in R$ and $k \in \mathbb{Z}^+$, we write $(x - r)^k \parallel f$ if $(x - r)^k \mid f$, and where writing $f = (x - r)^k \cdot h$ we have $h(r) \neq 0$. We call this k the **multiplicity of r in f.**

Proposition 5.6.1. *Let R be a commutative ring. Then for any polynomial $f \in R[x]$, an element $r \in F$ is a root of $f(x)$ if and only if $(x - r) \mid f(x)$.*

Definition 5.6.4. Consider a curve C defined by $f \in \mathbb{R}[x, y]$. Let $L : y = mx + b$ be a real line. Given a point $P = (x_1, y_1) \in L \cap C$, we have that x_1 is a root of

$$p(x) := f(x, mx + b).$$

Then the **intersection multiplicity of L and C at P** is the multiplicity of the root x_1 in $p(x)$.

If L is a vertical line, then writing $L : x = x_1$, the intersection multiplicity of L and C at P is instead defined from the multiplicity of the root $y = y_1$ in

$$q(y) := f(x_1, y).$$

The intersection multiplicity between a line and a curve at a point is the number of times the line "intersects" the curve at this point. The subtlety is that a line can intersect a curve at a point "more than once." For example, this always happens with tangent lines. We next define *singular points,* which have extra multiplicities with *all* lines.

Definition 5.6.5. Given a curve C defined by $f \in \mathbb{R}[x, y]$, a point $P = (x_1, y_1) \in C$ is **singular** if one has partial derivative evaluations

$$f_x(P) := \left.\frac{\partial f}{\partial x}\right|_P = 0$$

and

$$f_y(P) := \left.\frac{\partial f}{\partial y}\right|_P = 0.$$

Otherwise, the point P is said to be **nonsingular,** or **simple.** We call C a **singular curve** if it has at least one singular point. Otherwise C is called a **nonsingular curve.**

Singular curves often exhibit atypical behavior because of their singular points' "more-than-expected" multiplicities at singular points.

Example 5.6.4. Consider the curve $C : y^2 = x^3 - 3x + 2$. To check whether C is nonsingular, we calculate the partial derivatives of $f := y^2 - (x^3 - 3x + 2)$ and attempt to solve for points on C. We check that

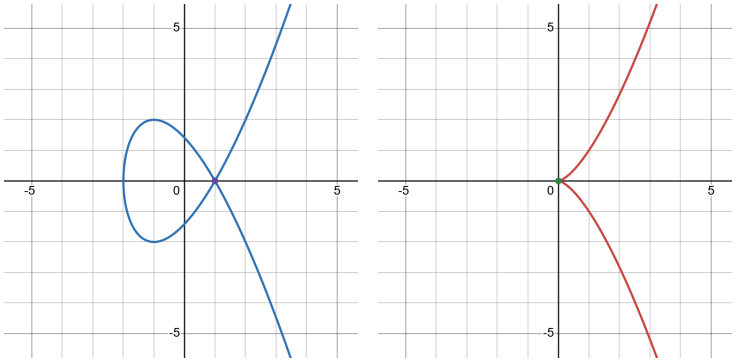
$$f_x = -3x^2 + 3,$$

and thus $f_x = 0$ implies $x = \pm 1$. On the other hand

$$f_y = 2y,$$

so that $f_y = 0$ implies $y = 0$. We deduce that the points $(\pm 1, 0)$ are singular on C *if they lie on* C. We check that $(1, 0) \in C$ but $(-1, 0) \notin C$. We conclude that C has exactly one singular point, namely $P = (1, 0)$. This is an example of a *nodal* cubic curve (see below).

Consider next the curve $D : y^2 = x^3$. Similar to the above, we can check that D has exactly one singular point, namely $P := (0, 0)$. This is an example of a *cuspidal* cubic curve.



(A) (B)

FIGURE 11. The singular curves $C : y^2 = x^3 - 3x + 2$ (nodal) and $D : y^2 = x^3$ (cuspidal).

Before returning to our procedure for creating rational points on a conic, we state some facts about intersection multiplicities and singular points for reference.

Proposition 5.6.2. *For a curve C of degree d and a nonsingular point $P \in C$, if T is the tangent line of P at C and $T \not\subseteq C$, then the intersection multiplicity k of L and C at P satisfies*

$$2 \leq k \leq d.$$

Furthermore T is the unique line which intersects C at P with intersection multiplicity ≥ 2.

Proposition 5.6.3. *For a curve C and singular point $P \in C$, the intersection multiplicity of any line L with C at P is ≥ 2.*

Let us revisit our example for producing infinitely many rational points on a conic $C_f/\mathbb{Q}$. We start with a rational point $P = (x_1, y_1) \in C(\mathbb{Q})$, and then for an arbitrary rational number $m \in \mathbb{Q}$ construct the line $L : y = mx + b$ through P, and then the polynomial

$$p(x) := f(x, mx + b).$$

This polynomial is quadratic since C is quadratic, and is defined over $\mathbb{Q}$ since L and C are defined over $\mathbb{Q}$. The roots of this polynomial are the x-coordinates of points in $L \cap C$. Since $P \in L \cap C$, we have that x_1 is a root of $p(x)$, *which implies the second root x_2 must also be rational.* It follows that the point $Q := (x_2, y_2) := (x_2, mx_2 + b)$ is also a rational point on C. As long as P is nonsingular and L is not the tangent line of P at C, we know by Proposition 5.6.2 that the intersection multiplicity of L and C at P is 1, and thus $x_2 \neq x_1$, so that Q is a rational point on C not equal to P. With a bit more work, this implies the following theorem.

Theorem 5.6.4. [NZM91, Page 255] *Let $C/\mathbb{Q}$ be a nonsingular conic defined by $f \in \mathbb{Q}[x, y]$. If $C(\mathbb{Q}) \neq \emptyset$, then C is parametrizable; in particular $\#C(\mathbb{Q}) = \infty$.*

Let us state our observations as an algorithm.

1. Start with a nonsingular conic $C/\mathbb{Q}$ defined by $f \in \mathbb{Q}[x, y]$, and fix a point $P = (x_1, y_1) \in C(\mathbb{Q})$.
2. Let $m \in \mathbb{Q}$, and consider the line L through P with slope m. Write this as
$$L : y = mx + b.$$
3. In analyzing $L \cap C$, plug $y = mx + b$ into $f(x, y)$ to get a single-variable polynomial
$$p(x) := f(x, mx + b).$$
4. Then $p(x)$ is quadratic, with x_1 as one of the roots. *Thus $p(x)$ has another rational root x_2.*
5. We conclude that $Q := (x_2, y_2)$ is in $C(\mathbb{Q})$, where $y_2 := mx_2 + b$; and if L is not the tangent line to C at P, then $Q \neq P$.

Remark 5.6.4. If $C/\mathbb{Q}$ is a *singular* conic, then it is still possible to characterize all rational points on C, provided that $C(\mathbb{Q}) \neq \emptyset$. This will be explored in Exercise 5.6.2.

We move next to analyzing rational points on *cubic* equations. As we will see, our previous idea of using rational lines and points to produce new rational points will generalize in an appropriate way. For (nonsingular) cubics, we will need to start with *two* rational points to produce a third one, to "force" our intermediate cubic polynomial $p(x)$ to have *three* rational roots.

Like before, we lead our analysis with an example.

Example 5.6.5. Consider the cubic curve $E : y^2 = x^3 + 17$; this is defined by $f(x, y) := y^2 - (x^3 + 17)$. We check that there are several integral points: $(-1, \pm 4)$, $(-2, \pm 3)$ and $(2, \pm 5)$.

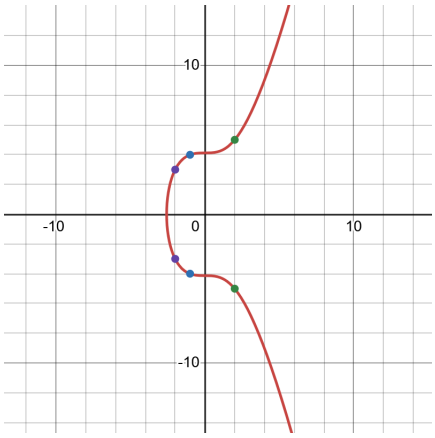


FIGURE 12. The cubic curve $E : y^2 = x^3 + 17$ in $\mathbb{R}^2$.

Check for yourself that this curve is nonsingular. Let us try and construct more rational points in analogy to our algorithm for conics. Our intermediate polynomial $p(x) := f(x, mx + b)$ from our intersection of a line $L : y = mx + b$ with this curve is a cubic with two known rational roots. Thus, it will have one more root, and it must be rational; this can be seen via long division, for example. This suggests that instead of

starting with a point $P \in E(\mathbb{Q})$ and slope $m \in \mathbb{Q}$, *we start with two points* $P, Q \in E(\mathbb{Q})$ *and take the line through them.*

Let us fix points $P = (-1, 4)$ and $Q = (2, 5)$ on E. Let L be the line through P and Q; then L has slope

$$m = \frac{5-4}{2-(-1)} = \frac{1}{3},$$

and thus an equation

$$L : y - 4 = \frac{1}{3}(x+1),$$

i.e.,

$$L : y = \frac{1}{3}x + \frac{13}{3}.$$

Consider the intersection $L \cap E$. The x-coordinates of points in this intersection correspond to the roots of

$$p(x) := f\left(x, \frac{1}{3}x + \frac{13}{3}\right).$$

We check that

$$\begin{aligned} p(x) &= \left(\frac{1}{3}x + \frac{13}{3}\right)^2 - (x^3 + 17) \\ &= -x^3 + \frac{1}{9}x^2 + \frac{26}{9}x + \frac{16}{9}. \end{aligned}$$

Observe that $p(x) = 0$ is equivalent to

$$-9p(x) = 9x^3 - x^2 - 26x - 16 = 0.$$

We know that $x = -1, 2$ are roots of $-9p(x)$ since $(-1, 4)$ and $(2, 5)$ are in $L \cap E$. Thus $(x+1)(x-2) = x^2 - x - 2$ divides $9x^3 - x^2 - 26x - 16$. We can use this to find the remaining linear factor, and thus the remaining root by Proposition 5.6.1. Long division shows that

$$9x^3 - x^2 - 26x - 16 = (x^2 - x - 2)(9x + 8).$$

Therefore, we have our third x-coordinate $x_3 := -\frac{8}{9}$. Plugging this back into L shows that we can take $y_3 := \frac{109}{27}$. We thus conclude that

$$R := \left(-\frac{8}{9}, \frac{109}{27}\right) \in E(\mathbb{Q}).$$

This method is called the **chord method** for producing rational points on a cubic, since we use the chord (line) between two rational points to produce a third rational point.

There is also another way to construct rational points *from just one point.* It is called the **tangent method,** since it uses the tangent line at a point instead of the chord through two points. We will apply the tangent method to $P := (-1, 4) \in E(\mathbb{Q})$. The tangent slope m_0 of E at P is

$$m_0 := \left.\frac{dy}{dx}\right|_P$$

(note the difference between this derivative and the partial derivative $\frac{\partial f}{\partial x}$!). From $y^2 = x^3 + 17$, we see that

$$\frac{dy}{dx} = \frac{3x^2}{2y},$$

and thus

$$m_0 = \frac{3}{8}.$$

Then the tangent line to E at P is

$$T : y - y_1 = m_0(x - x_1),$$

i.e.,

$$T : y = \frac{3}{8}x + \frac{35}{8}.$$

Plugging this expression for y into $f(x, y)$, we get the polynomial

$$p(x) := f\left(x, \frac{3}{8}x + \frac{35}{8}\right).$$

By construction, one root of $p(x)$ is $x_1 = -1$; since $p(x)$ was constructed from the *tangent line* to E at P, from Proposition 5.6.2 we have $(x+1)^2 \mid p(x)$, i.e. x_1 is a root of $p(x)$ with multiplicity two.

To find the other root x_3 of $p(x)$, we must analyze $p(x) = 0$. This simplifies to

$$64p(x) = 64x^3 - 9x^2 - 210x - 137 = 0.$$

Since $(x+1)^2 \mid 64p(x)$, we can apply long division to find the last linear factor:

$$64p(x) = 64x^3 - 9x^2 - 210x - 137 = (x+1)^2(64x - 137).$$

Thus, if we take $x_3 := \frac{137}{64}$, then setting $y_3 := \frac{3}{8}x_1 + \frac{35}{8} = \frac{2651}{512}$ we can conclude that

$$R := \left(\frac{137}{64}, \frac{2651}{512}\right) \in E(\mathbb{Q}).$$

The above curve is an example of an *elliptic curve.* Our methods for creating rational points on this curve are part of a **chord and tangent method;** with one additional step in this method, we can realize our elliptic curve as a *group.* We will talk more about this in the next section.

Remark 5.6.5. One way our construction of rational points on cubics differs from that of nonsingular conics, is that the procedure above for cubics does not necessarily produce infinitely many distinct rational points – even if the curve is nonsingular. For nonsingular conics $C/\mathbb{Q}$, one has by Theorem 5.6.4 that $C(\mathbb{Q}) \neq \emptyset \Rightarrow \#C(\mathbb{Q}) = \infty$, and in fact $C(\mathbb{Q})$ is parametrizable. In contrast, elliptic curves will always have $E(\mathbb{Q}) \neq \emptyset$ (in the *projective plane*), but $E(\mathbb{Q})$ can be either finite or infinite. We will discuss this in §5.7.

In the last part of this section, we will cover a few more geometric concepts which will be relevant to our study of elliptic curves.

1. The projective plane. The usual real plane $\mathbb{R}^2$ is an example of what is called an *affine plane.* As it turns out, curves in the affine plane can have points on them

which live in the ambient *projective plane;* such points are "invisible" over $\mathbb{R}^2$. Let us define the real projective plane.

Definition 5.6.6. We define an equivalence relation on $(\mathbb{R}^3)^\bullet := \mathbb{R}^3 \smallsetminus \{(0,0,0)\}$ where for points $P := (a,b,c), Q := (d,e,f) \in \mathbb{R}^3$, we have

$$P \sim Q$$

if there exists $\lambda \in \mathbb{R}$ with

$$P = \lambda Q := (\lambda d, \lambda e, \lambda f).$$

The **projective plane,** denoted $\mathbb{P}^2(\mathbb{R})$, is the resulting quotient set

$$\mathbb{P}^2(\mathbb{R}) := (\mathbb{R}^3)^\bullet / \sim .$$

We write $[a : b : c]$ for the equivalence class of an element $(a,b,c) \in (\mathbb{R}^3)^\bullet$.

One can view points in the projective plane as lines in $\mathbb{R}^3$ through the origin, sans including the origin. For example, in $\mathbb{P}^2(\mathbb{R})$ we have $[1:2:3] = [2:4:6] = [\pi : 2\pi : 3\pi]$, and we notice that $(1,2,3)$, $(2,4,6)$ and $(\pi, 2\pi, 3\pi)$ lie on same line in $\mathbb{R}^3$.

Definition 5.6.7. We can include $\mathbb{R}^2$ in $\mathbb{P}^2(\mathbb{R})$ via the inclusion

$$(a,b) \mapsto [a : b : 1].$$

Points in $\mathbb{P}^2(\mathbb{R}) \smallsetminus \mathbb{R}^2$ are called **points at infinity,** and have the form $[a : b : 0]$.

One of the key features of $\mathbb{P}^2(\mathbb{R})$ is that *parallel lines in* $\mathbb{R}^2$ *converge in* $\mathbb{P}^2(\mathbb{R})$, to a point at infinity. Here is an example of this for the parabola $y = x^2$.

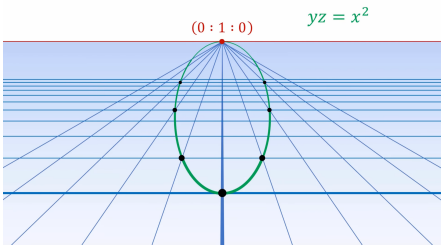


FIGURE 13. The parabola $y = x^2$ pictured in $\mathbb{P}^2(\mathbb{R})$. Picture from this video, titled "Putting Algebraic Curves in Perspective."

For more intuition on projective geometry, see the video in the caption above. For this class, the following is enough to know about the projective plane $\mathbb{P}^2(\mathbb{R})$.

- The definition of $\mathbb{P}^2(\mathbb{R})$ as a quotient set.
- The intuition that "parallel lines in $\mathbb{R}^2$ converge in $\mathbb{P}^2(\mathbb{R})$."
- The point at infinity $O := [0:1:0]$ lies on all vertical lines in $\mathbb{R}^2$ (see Remark 5.6.10).

Remark 5.6.6. As we will see, most elliptic curves have exactly one point at infinity, which is $O := [0:1:0]$, and this almost always serves as the identity element for their group law.

2. Homogeneous coordinates. To describe points on curves in the projective plane, the defining polynomials for these curves must be *homogeneous.*

Definition 5.6.8. A polynomial $F(X, Y, Z) \in \mathbb{R}[X, Y, Z]$ is **homogeneous** if every monomial in F has the same degree. ("Homogeneous" is synonymous with "same.")

Remark 5.6.7. Capital letters are often used for homogeneous polynomials and variables.

Example 5.6.6. We see that the polynomial $F(X, Y, Z) := XY^3 - Z^4$ is homogeneous, as $\deg(XY^3) = \deg(Z^4) = 4$. However $G(X, Y, Z) := XY + XZ - Y$ is not homogeneous.

Definition 5.6.9. Given a polynomial $f(x, y) \in \mathbb{R}[x, y]$ of degree $d \geq 1$, we can **homogenize** f and transform it into a homogeneous polynomial: the **homogenization of f,** denoted F, is

$$F(X, Y, Z) := f\left(\frac{X}{Z}, \frac{Y}{Z}\right) \cdot Z^d.$$

Conversely, given a homogeneous polynomial $G \in \mathbb{R}[X, Y, Z]$, one has the **dehomogenization of** G via

$$g(x, y) := G(x, y, 1).$$

Remark 5.6.8. Practically speaking, to compute a homogenization for a polynomial $f(x, y)$, you simply multiply all terms in f by an appropriate power of Z until all the monomial degrees are equal to $\deg(f(x, y))$.

Example 5.6.7.

- For $f := y^2 - x^3 - x - 1$, we have the homogenization
$$F(X, Y, Z) = Y^2Z - X^3 - XZ^2 - Z^3.$$
- For $g := xy - 1$, we have its homogenization
$$G(X, Y, Z) = XY - Z^2.$$
- For $H := Y^2Z - X^3 - 2XZ^2$, we have its dehomogenization
$$h(x, y) = y^2 - x^3 - 2x.$$

An upshot to the notion of homogenization is that an **affine curve** $C := C_f$ (so defined in $\mathbb{R}^2$) has a corresponding **projective curve**, written as C_H or C_F, which is the set of solutions in $\mathbb{P}^2(\mathbb{R})$ to

$$F(X, Y, Z) = 0.$$

Studying the homogenization of a curve lets us uncover the points at infinity on the curve.

Remark 5.6.9. It is important that in the definition of a projective curve, we only consider homogeneous polynomials. For a homogeneous polynomial $F \in \mathbb{R}[X, Y, Z]$, for any point $(a, b, c) \in \mathbb{R}^3$ one has for each $\lambda \in \mathbb{R}$ that

$$F(\lambda a, \lambda b, \lambda c) = \lambda^{\deg(F)} \cdot F(a, b, c),$$

and thus

$$F(a,b,c)=0$$

if and only if

$$F(\lambda a,\lambda b,\lambda c)=0.$$

Therefore, "being zero" at a point in $\mathbb{P}^2(\mathbb{R})$ makes sense for $F(X,Y,Z)$, so that the projective curve C_F, being the set of points in $\mathbb{P}^2(\mathbb{R})$ that are zeroes for $F(X,Y,Z)$, is well-defined.

Remark 5.6.10. The most common point at infinity is $O:=[0:1:0]$, which is seen on all elliptic curves in "Weierstrass form." For our eventual group law on an elliptic curve, it will be useful to note that this O "lies on all vertical lines in $\mathbb{R}^2$." To prove this claim, observe that for any vertical line

$$L: x=a,$$

it homogenization is

$$L_H: X=aZ.$$

Then we see that $O\in L_H$.

The converse of the above is also worth noting. That is, suppose we have a line

$$L: ax+by=c$$

which contains O in $\mathbb{P}^2$. This means that the homogenization

$$L_H: aX+bY=cZ$$

satisfies $O\in L_H$. This forces $b=0$, so that

$$L_H: aX=cZ.$$

Dehomogenizing then shows that

$$L: ax=c,$$

which is indeed a vertical line. (Note that $a\neq 0$.)

Remark 5.6.11. Given a polynomial $f\in\mathbb{R}[x,y]$ and its homogenization $F\in\mathbb{R}[X,Y,Z]$, the inclusion

$$\mathbb{R}^2\subseteq\mathbb{P}^2(\mathbb{R}),\qquad\qquad (a,b)\mapsto[a:b:1]$$

restricts to an inclusion

$$C_f\subseteq C_F.$$

This just says that $f(a,b)=0$ if and only if $F([a:b:1])=0$. Thus C_F includes points from C_f, as well as possible points at infinity $[a:b:0]$. By abuse of notation, when we refer to a curve C_f, we often implicitly mean its homogenization C_F, *and thus include its points at infinity in its description.*

Since we are now considering projective curves, let us define what it means for a point in the projective plane to be rational.

Definition 5.6.10. We say that a point $P=[a:b:c]\in\mathbb{P}^2(\mathbb{R})$ is **rational** if there exists $\lambda\neq 0\in\mathbb{R}$ such that $(\lambda a,\lambda b,\lambda c)\in\mathbb{Q}^3$. For a projective curve $C/\mathbb{Q}$, we write $C(\mathbb{Q})$ for the set of its rational points.

Example 5.6.8. The point $P := [\pi : 2\pi : 3\pi] \in \mathbb{P}^2(\mathbb{R})$ is rational since $P = [1 : 2 : 3]$. However, the point $Q := [1 : \sqrt{2} : 3] \in \mathbb{P}^2(\mathbb{R})$ is not rational – this amounts to proving that $\sqrt{2}$ is irrational.

The notion of singular points on affine curves extends to projective curves.

Definition 5.6.11. If C_F is a projective curve, then a point $P \in C_F$ is a **singular point** if all three partial derivatives of F evaluated at P are zero:

$$F_X(P) := \left.\frac{\partial F}{\partial X}\right|_P = 0,$$

$$F_Y(P) := \left.\frac{\partial F}{\partial Y}\right|_P = 0,$$

and

$$F_Z(P) := \left.\frac{\partial F}{\partial Z}\right|_P = 0,$$

Remark 5.6.12. For an affine curve C_f and its homogenization C_F, a point $(a, b) \in C_f$ is singular if and only if $[a : b : 1] \in C_F$ is singular. Thus C_f and C_F share *almost* the same set of singular points – sometimes, there may be singular points at infinity on C_F.

3. Irreducible curves. The last descriptor for an elliptic curve is tied to the fact that it cannot be "broken up" into smaller curves.

Definition 5.6.12. For a field F (such as $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$), we say that a nonconstant polynomial $f(x_1, \ldots, x_n) \in F[x_1, \ldots, x_n]$ is **irreducible over** F if it cannot be written as

$$f = g \cdot h$$

where $g, h \in F[x_1, \ldots, x_n]$ have strictly smaller degree. In such a case, we call the corresponding curve C_f an **irreducible curve over** F**.** This applies to both affine and projective curves.

Remark 5.6.13. The notion of an irreducible polynomial is in analogy to primality of an integer. In fact, these are both special instances of irreducible elements in a ring.

Example 5.6.9. Let us consider irreducibility of curves in $\mathbb{R}^2$.

- The polynomial $f(x, y) := x^2 - y^2$ is *reducible* over $\mathbb{R}$, as $f = (x + y)(x - y)$. Its curve C_f is the union of two lines which intersect at the origin.

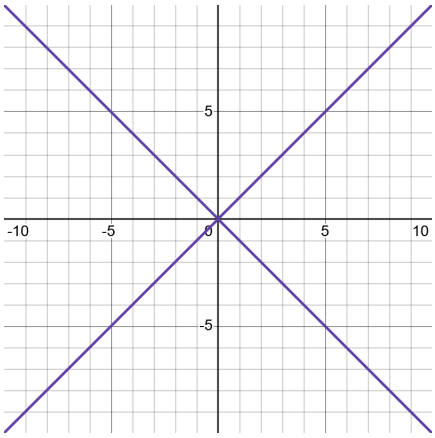


FIGURE 14. The conic $C : x^2 - y^2 = 0$ in $\mathbb{R}^2$.

- The polynomial $g(x) := x^2 + x + 1$ is *irreducible* over $\mathbb{R}$, since its roots are strictly complex. However, this shows us that it is reducible over $\mathbb{C}$.
- The polynomial $h(x, y) := x^2 + y^2$ is *irreducible* over $\mathbb{R}$; in fact, its curve C_h over $\mathbb{R}$ is empty. However, since $x^2 + y^2 = (x + iy)(x - iy)$, it is *reducible* over $\mathbb{C}$; it is a union of two complex lines.
- The polynomial $k(x, y, z) := x^3 + y^3 - z^3$ is *irreducible* over $\mathbb{C}$. Thus, the projective curve $E : X^3 + Y^3 = Z^3$ is irreducible.

Observe that if f and g are polynomials in $\mathbb{Q}[x, y]$ and $g \mid f$ over $\mathbb{R}$, then for any point $P \in \mathbb{C}^2$ one has

$$g(P) = 0 \qquad \Rightarrow \qquad f(P) = 0.$$

This implies that $C_g(\mathbb{Q}) \subseteq C_f(\mathbb{Q})$, as well as $C_g(\mathbb{R}) \subseteq C_f(\mathbb{R})$ and $C_g(\mathbb{C}) \subseteq C_f(\mathbb{C})$. Therefore, given a curve $C_f/\mathbb{Q}$, if you can find a divisor of f that is a smaller rational polynomial g, then rational/real/complex solutions of g are also rational/real/complex solutions of f. Thus *reducible curves reduce into smaller curves.* This gives one way to determine rational points on a reducible curve.

Remark 5.6.14. There will be a bonus exercise at the end of this section on checking that a polynomial of the form $y^2 - f(x)$, where $f(x)$ is a nonzero cubic polynomial, is irreducible over $\mathbb{C}$. This equation will account for most of the elliptic curves that we come across.

We are now ready to define an elliptic curve.

Definition 5.6.13. For a field F containing $\mathbb{Q}$ (such as $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$), a curve E/F is called an **elliptic curve over** F if it is a nonsingular cubic curve, irreducible over $\mathbb{C}$, with an F-rational point.

Exercises. From [NZM91, §5.6], page 260: $\#2, 5, 7 - 10$.

Exercise 5.6.1. Parametrize the rational points on the hyperbola

$$H : x^2 - 2y^2 = 1$$

using the point $(1,0) \in C(\mathbb{Q})$.

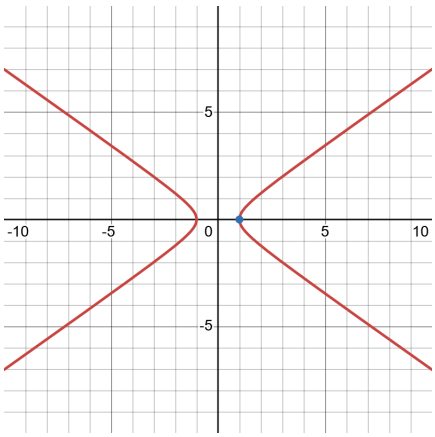


FIGURE 15. The hyperbola $H : x^2 - 2y^2 = 1$.

Exercise 5.6.2. Let $a, b, c \in \mathbb{Q}$ with $ab \neq 0$.

a) Show that if the conic

$$C : ax^2 + by^2 = c$$

is singular, then its only singular point is $(0,0)$, and $c = 0$.

b) Show that if a and b are squares in $\mathbb{Q}$, then the conic

$$C : ax^2 - by^2 = 0$$

is the union of two rational lines through the origin. Use this to characterize all rational points on C.

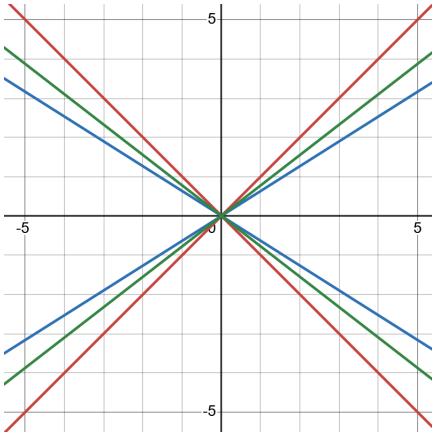


FIGURE 16. The (singular) conics $x^2 - y^2 = 0$, $2x^2 - 5y^2 = 0$ and $3x^2 - 5y^2 = 0$.

Exercise 5.6.3.

a) Show that for a polynomial $f(x) \in \mathbb{R}[x]$ and an integer $n \geq 2$, the curve

$$C : y^n = f(x)$$

has a singular point if and only if $f(x)$ has a *repeated root* in $\mathbb{R}$, i.e., there exists $x_0 \in \mathbb{R}$ with $f(x_0) = 0$ and $f'(x_0) = 0$.

b) Given a curve

$$C/\mathbb{R} : y^2 = (x - \alpha)(x - \beta)(x - \gamma)$$

where α, β and γ are real or complex numbers, the **discriminant** of C is

$$\Delta_C := [(\alpha - \beta)(\alpha - \gamma)(\beta - \gamma)]^2.$$

Prove that $\Delta_C = 0$ if and only if C is singular.

When such a curve $C : y^n = f(x)$ is nonsingular, it is called a *hyperelliptic curve.* When $n = 2$ and $\deg(f) \in \{3, 4\}$, this is an elliptic curve (which is more clear when $\deg(f) = 3$).

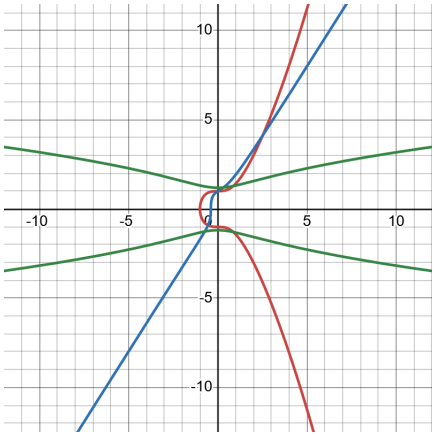


FIGURE 17. The hyperelliptic curves $y^2 = x^3 + 1$, $y^3 = 4x^3 + 2x + 1$ and $y^4 = x^2 + 1$.

Bonus Exercise 5.6.4. Prove that for a polynomial $p(x) \in \mathbb{Q}[x]$ of degree $n \geq 1$, if p has $n - 1$ rational roots, then it has n rational roots. (*Hint:* use the proposition from our notes that states $r \in \mathbb{Q}$ is a root of $p(x)$ if and only if $(x - r) \mid p(x)$.)

Bonus Exercise 5.6.5. Show that the following affine curves are nonsingular.

a) $F_n : x^n + y^n = 1$, where $n \geq 1$.
b) $C_1 : 5xy + y^2 = 2$.
c) $C_2 : y^5 = 4x^3 + 2x^2 - 2x - 1$.

Prove that the following projective curve is singular.

d) $C_3 : X^3 + X^2Z + X^2Y = Z^3$, where $C_3(\mathbb{R}) \subseteq \mathbb{P}^2(\mathbb{R})$.

(*Hint:* in some parts, Exercise 5.6.3 might help.)

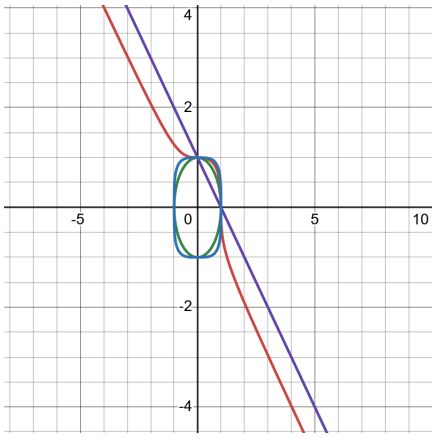


FIGURE 18. The Fermat curves F_1, F_2, F_3 and F_4.

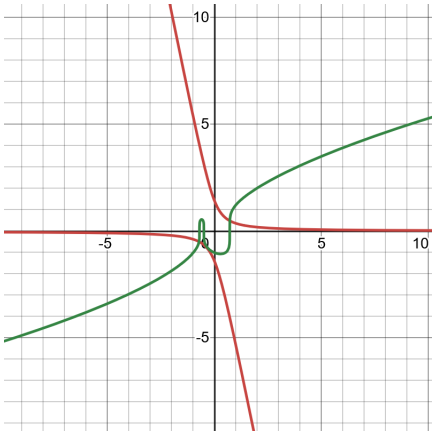


FIGURE 19. The curves C_1 (hyperbola) and C_2 (hyperelliptic).

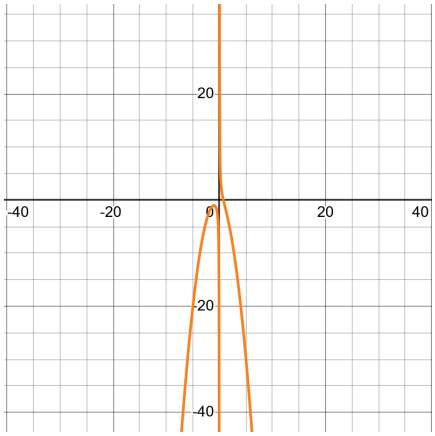


FIGURE 20. The curve C_3, dehomogenized and pictured in $\mathbb{R}^2$.

Bonus Exercise 5.6.6. Let $f(x) \in \mathbb{R}[x]$ be a cubic polynomial. Show that $y^2 - f(x) \in \mathbb{R}[x, y]$ is an irreducible polynomial over $\mathbb{C}$. (*Hint:* show that $y^2 - f(x)$ being reducible in $\mathbb{C}[x, y]$ means we can write $y^2 - f(x) = (y - g(x))(y - h(x))$ for some $g, h \in \mathbb{C}[x]$.)

Bonus Exercise 5.6.7. This exercise will explore the concept of the *genus* of a plane curve.

Suppose that $f(x, y) \in \mathbb{Q}[x, y]$ is an irreducible polynomial of degree d such that C_f is *nonsingular*. Then the **genus** of C, written as $g := g(C)$, is equal to $\frac{(d-1)(d-2)}{2}$.

The genus g of a curve $C/\mathbb{Q}$ is intimately connected to the number of rational points on C. When $g = 0$, either C has zero or infinitely many rational points; for example, conics are genus zero curves. When $g = 1$, we have that C is an elliptic curve. And when $g \geq 2$, a celebrated result of G. Faltings implies that C has *finitely* many rational points.

Determine whether the rational curves defined by the following equations have a finite or infinite amount of rational points (or that the information is inconclusive).

a) $C_1 : x^2 + y^2 = r^2$, where $r \neq 0 \in \mathbb{Q}$.
b) $C_2 : y^2 = x(x - 1)(x - 2)$.
c) $C_3 : y^5 = x(x - 1)(x - 3)(x - 5)(x - 7)$.
d) $F_n : x^n + y^n = 1$, where $n \in \mathbb{Z}^+$.

The genus also has a visual interpretation. A nonsingular irreducible curve $C/\mathbb{Q}$ with genus $g \geq 0$, when viewed as a *complex Riemann surface* in *projective space,* appears as a torus with g holes. Thus, an elliptic curve over $\mathbb{C}$ is a "complex donut," for example.

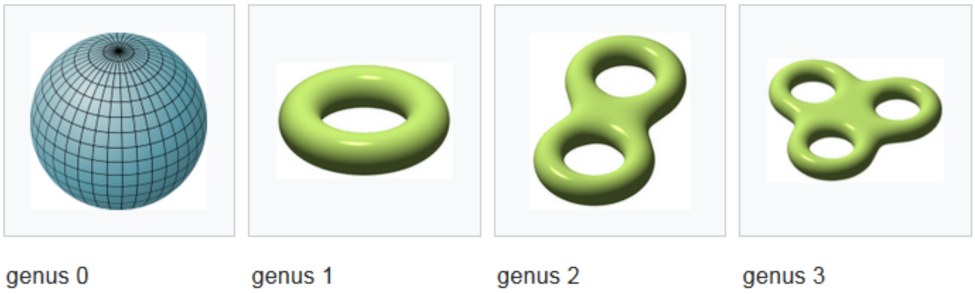


FIGURE 21. Pictures of g-holed tori in complex projective space, cf. Wikipedia.

5.7. **Elliptic Curves.** In this section, we will study the *group structure* of elliptic curves. Here are the main goals.

- Define chord and tangent method on elliptic curves, and prove it is a group law.
- State the Mordell-Weil Theorem.
- Prove the Collinearity Theorem for flex points.

Given a cubic curve $C/\mathbb{Q}$, we saw in the previous section that one can produce rational points on C either by using a line through two rational points $P, Q \in C(\mathbb{Q})$, or by using the tangent line from a point $P \in C(\mathbb{Q})$; this is called the "chord and tangent method." When $C = E$ is an elliptic curve, we can use this method to endow $E(\mathbb{Q})$ with a *group law,* which we will also call the chord and tangent method.

Recall our definition of an elliptic curve from the end of §5.6.

Definition 5.7.1. For a field F containing $\mathbb{Q}$ (such as $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$), a curve E/F is called an **elliptic curve (over F)** if it is a nonsingular cubic curve, irreducible over $\mathbb{C}$, with an F-rational point.

Remark 5.7.1. We will often take for granted that our curves are irreducible over $\mathbb{C}$. However, Bonus Exercise 5.6.6 shows that curves defined by $y^2 = f(x)$ for cubic $f \in \mathbb{R}[x]$ are irreducible over $\mathbb{C}$.

We now describe the group law on an elliptic curve. This group law, and many of results in this section, also apply to elliptic curves E/F where F is an arbitrary field. But to simplify matters, we state them for $F = \mathbb{Q}$.

The group law on an elliptic curve $E/\mathbb{Q}$. Fix a rational point $O \in E(\mathbb{Q})$. Then given two rational points $P, Q \in E(\mathbb{Q})$, define their **sum** $P \oplus Q$ as follows.

i) First, take the line $L_{P,Q}$ through P and Q; if $P = Q$, then take the tangent line $T_P := L_{P,P}$ to E at P instead. It will intersect the curve at a rational third point, which we denote by $P * Q$.
ii) Next, take the line $L_{P_1 * P_2, O}$ through $P_1 * P_2$ and O. It will intersect the curve at a rational third point, which we call the **sum** of P and Q, and denote as $P \oplus Q := (P * Q) * O$.

Remark 5.7.2. When E is given in *Weierstrass form* (which we define soon), this group law is described by a formula; see Exercises 5.7.1 and 5.7.2. When $E/\mathbb{Q}$ is given in Weierstrass form, the point $O := [0 : 1 : 0] \in E(\mathbb{Q})$ is our natural choice of identity element, which lends us simpler formulas for adding two points. For example, in *short Weierstrass form,* writing $P * Q := (x, y)$ one has $P \oplus Q = (x, -y)$. Unless otherwise stated, we will always assume $[0 : 1 : 0]$ is our identity when working in Weierstrass form.

Next we define a *flex point.*

Definition 5.7.2. For a curve $C/\mathbb{R}$, a nonsingular point $P \in C$ is called a **flex point,** or **inflection point,** if the intersection multiplicity of C and the tangent line T_P to C at P is ≥ 3.

If $C/\mathbb{R}$ is a nonsingular cubic curve, then $P \in C$ is a flex point if and only if $T_P \cap C = \{P\}$. As we will see, having a flex point identity will simplify the second step

of the group law for an elliptic curve. (Note that $[0:1:0]$ is always a flex point on an elliptic curve in Weierstrass form.)

Remark 5.7.3. It is worth noting that we have already seen flex points in Calculus 1! Recall that for a curve $C/\mathbb{R}$, a point $P \in C(\mathbb{R})$ is a flex point if and only if the *concavity* of C changes at P. Bonus Exercise 5.7.10 explores an example of this.

Here is a short list of facts that are useful to keep in mind when computing sums of points on an elliptic curve $E/\mathbb{Q}$ for the first time.

1. For a point $P \in E$, the tangent line T_P to E at P "contains P twice," i.e. $T_P \cap E$ has intersection multiplicity ≥ 2 at P. (See Proposition 5.6.2.)
2. Any line is uniquely determined by two points on it. Thus, any two lines L_1 and L_2 have $\#L_1 \cap L_2 \geq 2$ if and only if $L_1 = L_2$.
3. Consider an elliptic curve $E/\mathbb{Q}$ in **general Weierstrass form:**
$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$
where $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Q}$.
   - E has exactly one point at infinity, namely $O := [0:1:0]$, and O is rational.
   - O is a flex point on E, and thus its tangent line T_O to E satisfies $T_O \cap E = \{O\}$.
   - A line $L \subseteq \mathbb{R}^2$ is vertical if and only if $O \in L$; see Remark 5.6.10.

Let us give a proof that for an elliptic curve $E/\mathbb{Q}$, the chord and tangent method $\oplus$ described above indeed defines a group law on $E(\mathbb{Q})$. To simplify matters, we will assume our fixed identity $O \in E(\mathbb{Q})$ is a flex point. For example, when E is in general Weierstrass form, one has that $O := [0:1:0] \in E(\mathbb{Q})$ is automatically flex. As noted previously, flex point identities simplify the second step of the chord and tangent method. Another upshot is that there exists a *Collinearity Theorem for flex,* for sums of three points on the same line, which we will talk about later (Theorem 5.7.3).

Theorem 5.7.1 (Elliptic curve group law)**.** *Given an elliptic curve $E/\mathbb{Q}$ with a fixed point $O \in E(\mathbb{Q})$, the (two-step) chord and tangent method described above makes $E(\mathbb{Q})$ an abelian group.*

Remark 5.7.4. We sometimes write $(E(\mathbb{Q}), O)$ instead of $E(\mathbb{Q})$ to emphasize the choice of identity point O.

Proof. For simplicity, we assume that O is a flex point; the general case is left as an exercise for the reader. We need to check that $E(\mathbb{Q})$ satisfies the axioms for an abelian group, which are:

1. $E(\mathbb{Q})$ is **closed** under $\oplus$, i.e. $\oplus$ takes $E(\mathbb{Q})$ to itself.
2. The binary operation $\oplus$ is **abelian:** $\forall P, Q \in E(\mathbb{Q})$, one has $P \oplus Q = Q \oplus P$.
3. $E(\mathbb{Q})$ has an **identity element,** namely the fixed point $O \in E(\mathbb{Q})$.
4. $E(\mathbb{Q})$ has **inverses:** for $P \in E(\mathbb{Q})$, there exists $Q \in E(\mathbb{Q})$ with
$$P \oplus Q = O.$$
We will write the inverse additively, as $-P := Q$.

5. The binary operation $\oplus$ is **associative:** $\forall P, Q, R \in E(\mathbb{Q})$, one has
$$(P \oplus Q) \oplus R = P \oplus (Q \oplus R).$$

Let us check each axiom:

1. By definition of $\oplus$, it is a binary operation $E(\mathbb{Q}) \times E(\mathbb{Q}) \to E$, so it remains to show that it is in fact a map $E(\mathbb{Q}) \times E(\mathbb{Q}) \to E(\mathbb{Q})$. For $P, Q \in E(\mathbb{Q})$, since P and Q are rational, and since $L_{P,Q}$ and E are defined over $\mathbb{Q}$, the third point $P * Q$ in $L_{P,Q} \cap E$ is also rational. Since O is rational, and since $L_{P*Q,O}$ and E are defined over $\mathbb{Q}$, the sum $P \oplus Q$ is also rational.
2. For $P, Q \in E(\mathbb{Q})$, the line through P and Q is the same as the line through Q and P, so the first step of the chord and tangent method is the same. This shows that $P \oplus Q = Q \oplus P$.
3. Given a point $P \in E(\mathbb{Q})$, we must show that
$$P \oplus O = P.$$
The chord and tangent method is broken down into two steps:
   i) Construct the line $L_{P,O}$, which goes through E at a third point, written as $P * O \in E(\mathbb{Q})$.
   ii) Then construct the line $L_{P*O,O}$, whose third point with E is $P \oplus O := (P * O) * O$.

   However, this second line $L_{P*O,O}$ shares the points $P*O$ and O with $L_{P,O}$, which forces $L_{P*O,O} = L_{P,O}$. Thus their intersections with E are the same, so that the third point in $L_{P*O,O} \cap E$ is P, whence we deduce that $P \oplus O = P$.
4. We claim that $-P = P * O$, i.e. $P \oplus (P * O) = O$. To see this, we go through the chord and tangent method again:
   i) The first line is $L_{P,P*O}$, which has third point $P * (P * O)$ with E, by definition.
   ii) The second line is $L_{P*(P*O),O}$, which has third point $P \oplus (P * O)$, by definition.

   To reiterate, we want to show that $P \oplus (P * O) = O$. Consider our first line $L_{P,P*O}$: by definition of $P * O$, we know that the line $L_{P,O}$ contains $P * O$. Thus, these two lines $L_{P,P*O}$ and $L_{P,O}$ share the points P and $P * O$, which forces them to be equal. We deduce that $P * (P * O) = O$. Then our second line $L_{P*(P*O),O} = L_{O,O} = T_O$ is the tangent line to E at O. Since O is flex, this implies that $T_O \cap E = \{O\}$. We conclude by our steps above that $P \oplus (P*O) = O$, so that $-P = P * O$.
5. It can be tedious to check associativity, so here's an example picture of it:

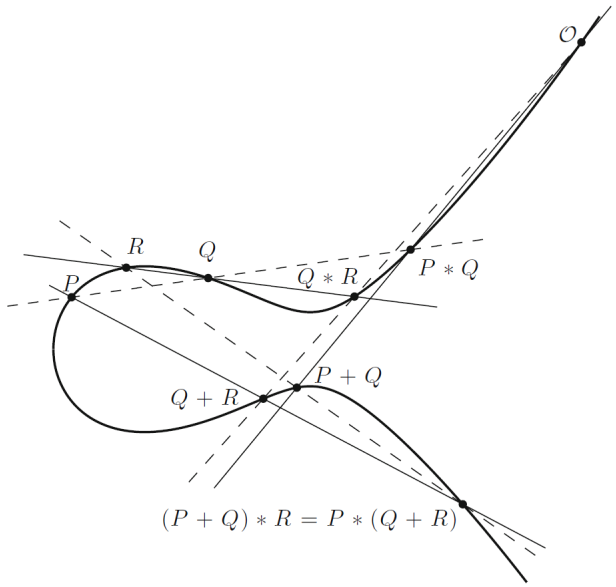


FIGURE 22. Example of associativity on an elliptic curve [ST15]. □

Remark 5.7.5. From here on out, given a point $P \in E(\mathbb{Q})$ and integer $n \in \mathbb{Z}^+$, we will write $nP := \underbrace{P \oplus P \oplus \ldots \oplus P}_{n \text{ times}}$. Similarly, for $n < 0$ we set $nP := \underbrace{-P \oplus -P \oplus \ldots \oplus -P}_{n \text{ times}}$, and $0P := O$.

Example 5.7.1. Let us do a full example of the group law on an elliptic curve. For this example, we will consider the cubic curve $E : y^2 = x^3 - 7x + 10$. We can check directly that it is a nonsingular curve, and is thus an elliptic curve. We can spot the integral points $P := (1, 2)$ and $Q := (3, 4)$ on E.

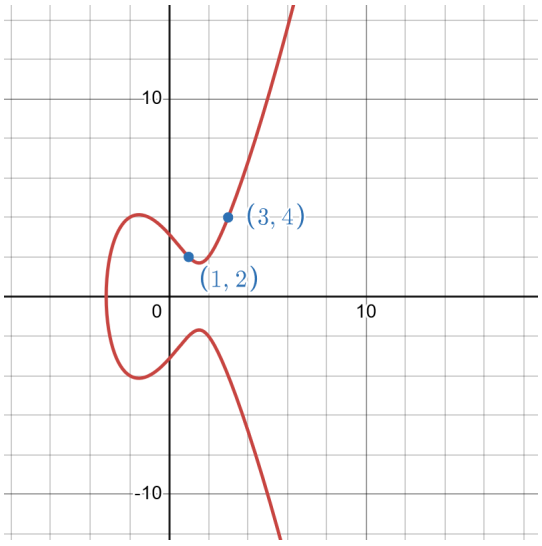


FIGURE 23. The elliptic curve $E : y^2 = x^3 - 7x + 10$.

The sum $P \oplus Q$ is determined by the chord and tangent method.

1. Let $L_1 := L_{P,Q}$ be the line through P and Q. Then its slope is $m = \frac{4-2}{3-1} = 1$; it has the equation
$$L_1 : y - y_1 = m(x - x_1),$$
i.e.,
$$L_1 : y = x + 1.$$
Let us analyze the intersection $L_1 \cap E$. To do this, we plug $y = x + 1$ into our equation for E:
$$\begin{aligned} y^2 = x^3 - 7x + 10 &\Rightarrow (x+1)^2 = x^3 - 7x + 10 \\ &\Rightarrow x^3 - x^2 - 9x + 9 = 0 \\ &\Rightarrow x^2(x-1) - 9(x-1) = 0 \\ &\Rightarrow (x^2 - 9)(x-1) = 0 \\ &\Rightarrow (x+3)(x-3)(x-1) = 0. \end{aligned}$$
Thus, there are 3 points in $L_1 \cap E$, each with x-coordinates $x = 1, 3$ and -3. We knew $x = 1$ and $x = -3$ were already roots of this polynomial, since $P, Q \in L_1 \cap E$. Let us set $x_3 := -3$. Taking $y_3 := x_3 + 1 = -2$, we conclude that $P * Q := (-3, -2)$ is on E and is collinear to P and Q.
2. Next, we consider the line through P and $O := [0 : 1 : 0]$. We know this is a vertical line by Remark 5.6.10, but we spell out the details one more time. We write this line as
$$L_2 := L_{R,O} : ax + by = c.$$
Homogenizing L_2 gives
$$L_{2,H} : aX + bY = cZ,$$

a line in $\mathbb{P}^2(\mathbb{R})$. Since $O \in L_2$, we have $a \cdot 0 + b \cdot 1 = c \cdot 0$, so that $b = 0$. Thus L_2 has the form

$$L_2 : ax = c,$$

which is a vertical line. Since $P * Q \in L_2$, we have $a \cdot -3 = c$, and so

$$L_2 : ax = -3a.$$

We can divide by a and get a new equation for this line:

$$L_2 : x = -3.$$

Let us analyze $L_2 \cap E$. Plug $x = -3$ into the equation for E and get the single-variable equation

$$y^2 = (-3)^3 - 7 \cdot (-3) + 10 = 4.$$

Thus, two points on $L_2 \cap E$ are $(-3, \pm 2)$ (note that the third point is the point at infinity, which can only been seen once the equation is homogenized). We conclude that $P \oplus Q := (-3, 2)$.

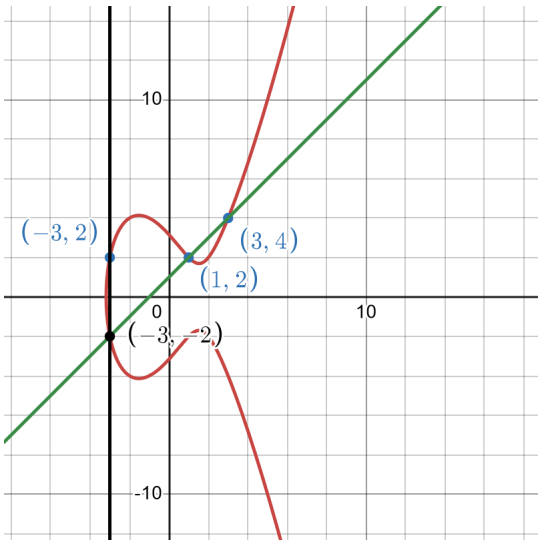


FIGURE 24. The elliptic curve $E : y^2 = x^3 - 7x + 10$, with the process for computing $P \oplus Q$ illustrated.

Next we add $P = (1, 2)$ to itself; this is written as $P \oplus P$, or simply $2P$. We will use the tangent method in the first step.

1. Let $T_3 := T_P$ be the tangent line to E at P. Then

$$T_3 : y = m(x - 1) + 2$$

where m is the tangent slope of E at P. We can compute it as follows:

$$\frac{d}{dx}[y^2 = x^3 - 7x + 10] \Rightarrow \frac{dy}{dx} = \frac{3x^2 - 7}{2y},$$

and so

$$m = \left.\frac{dy}{dx}\right|_{(1,2)} = \frac{-4}{4} = -1.$$

Thus

$$T_3 : y = 3 - x.$$

Plug $y = 3 - x$ into $y^2 = x^3 - 7x + 10$ and solve for x:

$$\begin{aligned} y^2 = x^3 - 7x + 10 &\Rightarrow (3-x)^2 = x^3 - 7x + 10 \\ &\Rightarrow x^3 - x^2 - x + 1 = 0 \\ &\Rightarrow x^2(x-1) - (x-1) = 0 \\ &\Rightarrow (x^2-1)(x-1) = 0 \\ &\Rightarrow (x-1)^2(x+1) = 0. \end{aligned}$$

As expected, the factor $(x-1)$ appears twice since $P = (1,2)$ has multiplicity two in $T_3 \cap E$. Thus, we can take $x_3 := -1$ and $y_3 := 3 - x_3 = 4$, and deduce that

$$P * P := (-1, 4) \in E.$$

2. Next, take the line $L_4 := L_{P*P,O}$ through $P * P$ and O. As noted before, this is the vertical line through $P * P$, so it has the form

$$L_4 : x = -1.$$

Immediately, the third point of intersection on L_4 must be $(-1, -4)$: simply note that $y = 4$ is already a root of $y^2 = (-1)^3 - 7(-1) + 10 = 16$, and thus the other root must be the negative of this. We conclude that $2P = (-1, -4)$.

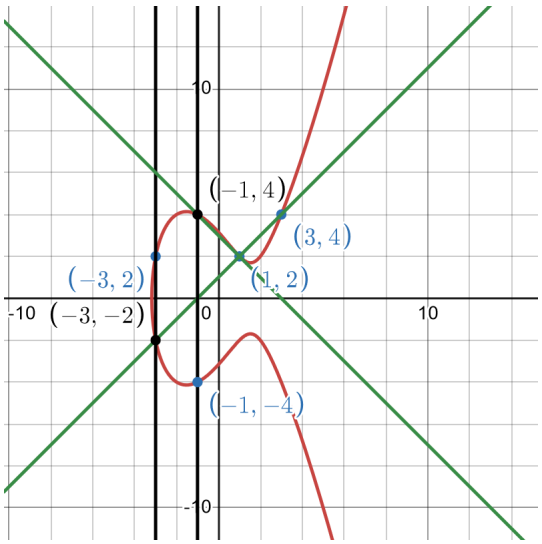


FIGURE 25. The elliptic curve $E : y^2 = x^3 - 7x + 10$, with the processes for computing $P \oplus Q$ and $2P$ illustrated.

The group $E(\mathbb{Q})$ is called the **Mordell-Weil group of E over $\mathbb{Q}$.** We know the Mordell-Weil group an abelian group. However, it admits additional structure: it is a *finitely generated* abelian group, i.e., there exists some finite set of rational points on E such that every rational point on E is a $\mathbb{Z}$-linear combination of these points.

Theorem 5.7.2 (The Mordell-Weil Theorem)**.** *For an elliptic curve $E/\mathbb{Q}$, the group $E(\mathbb{Q})$ is a* finitely generated abelian group: *there exist $P_1, P_2, \ldots, P_n \in E(\mathbb{Q})$ such that for any point $P \in E(\mathbb{Q})$, one has*

$$P = a_1P_1 \oplus a_2P_2 \oplus \ldots \oplus a_nP_n$$

for some $a_1, a_2, \ldots, a_n \in \mathbb{Z}$.

Proving the Mordell-Weil Theorem is beyond the scope of this course, so we will content ourselves with its statement. It is worth noting that this theorem holds more generally for elliptic curves over **number fields,** i.e., fields F containing $\mathbb{Q}$ such that the *degree* of the field extension $F/\mathbb{Q}$ is finite. We will talk about some consequences of the Mordell-Weil Theorem in the next section.

To conclude this section, let us prove a particularly nice result about collinear points on an elliptic curve with a flex identity.

Theorem 5.7.3 (Collinearity Theorem for flex points)**.** *Let $E/\mathbb{Q}$ be an elliptic curve, and assume that the identity $O \in E(\mathbb{Q})$ is a flex point. Then any three points $P, Q, R \in E(\mathbb{Q})$ are collinear (with multiplicity counted) if and only if*

$$P \oplus Q \oplus R = O.$$

Proof. Let us first note that for any points $P, Q \in E$, the three points P, Q and $P * Q$ are collinear, by definition of $P*Q$: they lie on the line $L_{P,Q}$. Let us also recall that any line containing e.g. both P and $P * Q$ must also contain Q, since the line containing P and $P * Q$ is unique, and thus must be $L_{P,Q}$.

We claim that

$$P \oplus Q \oplus (P * Q) = O.$$

To see this, we break down the chord and tangent method for adding $P \oplus Q$ and $P * Q$:

1. For the first step: the line $L_{P\oplus Q, P*Q}$ is equivalent to the line $L_{P*Q,O}$, as both contain $P * Q$ and $P \oplus Q$, by the definition $P \oplus Q := (P * Q) * O$. Thus $(P \oplus Q) * (P * Q) = O$ is the third point in $L_{P\oplus Q, P*Q} \cap E$.
2. For the second step: the line $L_{O,O} = T_O$ is the tangent line to E at O. Since O is flex, we have $T_O \cap E = \{O\}$.

We conclude that

$$P \oplus Q \oplus (P * Q) = O, \qquad \text{i.e.} \qquad P * Q = -(P \oplus Q). \tag{26}$$

The theorem then follows quickly, by uniqueness of the three points P, Q and $P * Q$ in $L_{P,Q} \cap E$.

$\Rightarrow$: if P, Q and R are collinear, then this forces $R = P*Q$ by uniqueness of $L_{P,Q} \cap E$, and thus $P \oplus Q \oplus R = O$ by (26).

$\Leftarrow$: if $P \oplus Q \oplus R = O$, then by (26) we have

$$R = -(P \oplus Q) = P * Q.$$

Thus P, Q and R are collinear. □

Exercises. From [NZM91, §5.7], pages $278 - 279$: $\#5, 6$.

Exercise 5.7.1. Let $E/\mathbb{Q}$ be an elliptic curve in **general Weierstrass form:**

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

where $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Q}$.

Fix points $P := (x_1, y_1)$ and $Q := (x_2, y_2)$ in $E(\mathbb{Q})$. Let L denote the line through P and Q, with slope $m \in \mathbb{Q} \cup \{\infty\}$. We set $O := [0 : 1 : 0] \in E(\mathbb{Q})$.

a) Prove that L is a vertical line if and only if P, Q and O are collinear, iff $x_1 = x_2$ (and thus $y_1 = -a_1x_1 - a_3 - y_2$). (*Hint:* for the second iff, you may need to complete a square.)

b) Show that if P, Q and O are not collinear, then

$$P * Q := (x_3, y_3) = (x_3, m(x_3 - x_1) + y_1)$$

with $x_3 = m^2 + a_1m - a_2 - x_1 - x_2$.

c) Continuing part b), show that

$$P \oplus Q := (x_4, y_4) = (x_3, -a_1x_3 - a_3 - y_3).$$

d) In contrast to parts b) and c), prove that if P, Q and O are collinear, then

$$Q = -P.$$

e) Prove that for a point $P = (x, y) \in E(\mathbb{Q})$, one has

$$-P = (x, -a_1x - a_3 - y).$$

f) Argue why these formulas should still hold if we replace $\mathbb{Q}$ with an arbitrary field F. (One point)

Exercise 5.7.2. Let $E/\mathbb{Q}$ be an elliptic curve in **short Weierstrass form:**

$$E : y^2 = x^3 + Ax + B$$

where $A, B \in \mathbb{Q}$.

a) Briefly explain how short Weierstrass form is a special case of general Weierstrass form.

b) Show that for two points $P := (x_1, y_1)$ and $Q := (x_2, y_2)$ in $E(\mathbb{Q})$ which are not collinear to $O := [0 : 1 : 0]$, one has the formula

$$P \oplus Q := (x_3, y_3) = (m^2 - x_1 - x_2, -m(x_3 - x_1) - y_1).$$

c) Show that for any point $P := (x, y) \in E(\mathbb{Q})$, one has

$$-P = (x, -y).$$

(*Hint for all parts:* use Exercise 5.7.1.)

Exercise 5.7.3. Consider the elliptic curve

$$E : y^2 = x^3 + 17$$

(we studied it in §5.6). Given points $P_1 := (-2, 3)$, $P_2 := (-1, 4)$ and $P_3 := (2, 5)$ in $E(\mathbb{Q})$, prove the following.

a) $-2P_1 = (8, 23)$.
b) $P_2 \oplus P_3 = \left(-\frac{8}{9}, -\frac{109}{27}\right)$ (which is $\approx (-0.88889, -4.03704)$).

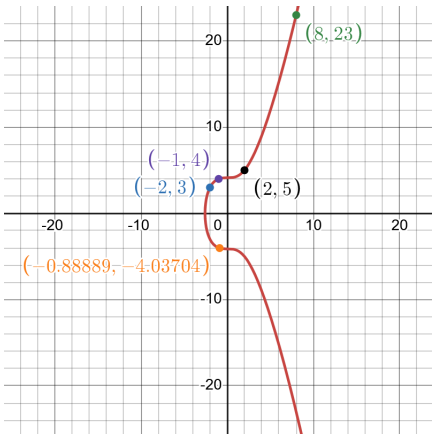


FIGURE 26. The elliptic curve $E : y^2 = x^3 + 17$.

Exercise 5.7.4. This exercise introduces some basic elliptic curve functions in `Sage`, as well as the *LMFDB database* [LMFDB]. A comprehensive manual on elliptic curve functions in `Sage` can be found here.

a) Using the elliptic curve $E : y^2 = x^3 + 17$ from Exercise 5.7.3, for the points $P_1 := (-2, 3)$, $P_2 := (-1, 4)$ and $P_3 := (2, 5)$ in $E(\mathbb{Q})$, write `Sage` code to compute the following.
   i) $-2P_1$.
   ii) $P_2 \oplus P_3$.
   iii) nP_1, for $0 \leq n \leq 30$. What do you observe from the output?
   (*Hint:* useful functions include `EllipticCurve([a1,a2,a3,a4,a6])`, as well as `E(a,b)` to realize a point (a, b) as a point on an elliptic curve E.)

A great site for elliptic curve data is the **LMFDB.** At this moment in time, it contains over 3.8 million elliptic curves over $\mathbb{Q}$, with dozens of numerical invariants listed for each of these curves. On each elliptic curve page, the 'Show commands' option in the top right corner lets you view these invariants as `Sage` code functions!

Parts b) - d) continue to focus on our elliptic curve $E : y^2 = x^3 + 17$ from part a).

b) Navigate the LMFDB and write down the *LMFDB label* of E.
c) Based on its LMFDB page, how many integral points does E have?
d) Based on its LMFDB page, what is the group structure of $E(\mathbb{Q})$, as an abstract abelian group?

The final part of this exercise focuses on patterns in elliptic curve *torsion groups.* For an elliptic curve $E/\mathbb{Q}$, we have its **torsion subgroup over** $\mathbb{Q}$**,** defined as

$$E(\mathbb{Q})[\text{tors}] := \{P \in E(\mathbb{Q}) : \exists n \in \mathbb{Z}^+, nP = O\}.$$

Thus $E(\mathbb{Q})[\text{tors}]$ is the subgroup of $E(\mathbb{Q})$ of points with finite order.

e) Write code to compute the *size* of torsion subgroups of all elliptic curves $E/\mathbb{Q}$: $y^2 = x^3 + Ax + B$ where $0 \leq A, B \leq 50$.

(*Hint:* you might find `E.torsion_subgroup()` useful for describing the structure of $E(\mathbb{Q})[\text{tors}]$. Every such torsion subgroup satisfies $E(\mathbb{Q})[\text{tors}] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ for some $m, n \in \mathbb{Z}$ with $m \mid n$; the function `E.torsion_subgroup().invariants()` returns () if $m = n = 1$, (n) if $n = 1$ and (m, n) otherwise.)

f) Based on your calculations in e), what observations can you make? (One point)

Exercise 5.7.5. This exercise will describe the 2-torsion and 3-torsion points on an elliptic curve $E/\mathbb{Q}$ in Weierstrass form. Recall that for an integer $n > 0$, a point $P \in E$ is **n-torsion** if $nP = O$.

First, assume that E is given in general Weierstrass form:

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

a) Prove that the 2-torsion points on E are precisely the points with vertical tangent lines.

b) Prove that the 3-torsion points are precisely the flex points of E.

For parts c) and d), assume that E is in short Weierstrass form:

$$E : y^2 = x^3 + Ax + B.$$

c) Show that the order two points on E are precisely the points $(\alpha, 0)$ where $\alpha \in \mathbb{C}$ is a root of $x^3 + Ax + B$.

d) Use part c) to prove that if $x^3 + Ax + B$ has a root over $\mathbb{Q}$, then $\#E(\mathbb{Q})[\text{tors}]$ is even.

Exercise 5.7.6. This exercise shows there are no integral points on the elliptic curve $E : y^2 = x^3 + 7$, using elementary techniques.

a) For the sake of contradiction, assume that $(a, b) \in E(\mathbb{Q})$ is an integral point. Show that a must be odd.

b) Show that $b^2 + 1 = (a + 2)(a^2 - 2a + 4)$.

c) Show that $a^2 - 2a + 4$ is congruent to 3 modulo 4. Then explain why there exists a prime divisor $p \mid (a^2 - 2a + 4)$ congruent to 3 modulo 4.

d) Reduce the original equation modulo p to derive a contradiction.

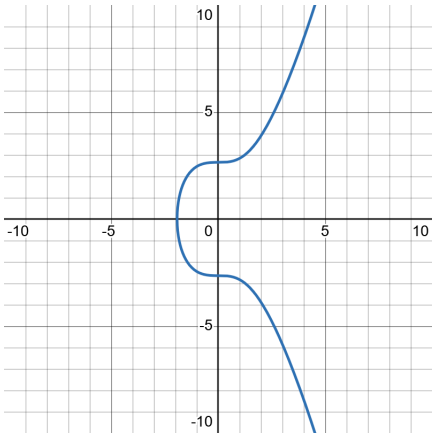


FIGURE 27. The elliptic curve $E : y^2 = x^3 + 7$.

Exercise 5.7.7. Consider the elliptic curve

$$E/\mathbb{Q} : y^2 = x^3 - 7x + 10,$$

which we did an example with in §5.7. We showed that for $P := (1, 2)$ and $Q := (3, 4)$, one has $P \oplus Q = (-3, 2)$ and $2P = (-1, -4)$. Compute $P \oplus' Q$ and $P \oplus' P$ in the group $(E(\mathbb{Q}), \oplus', Q)$, where Q is our new fixed identity.

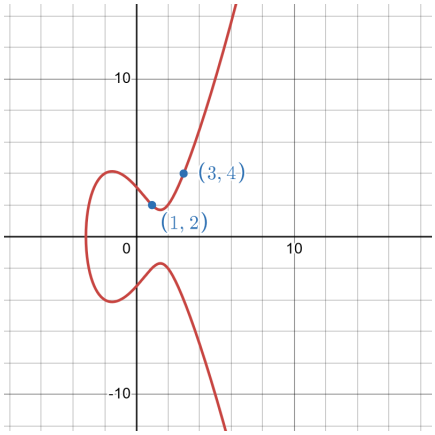


FIGURE 28. The elliptic curve $E : y^2 = x^3 - 7x + 10$.

Exercise 5.7.8. This exercise explores some arithmetic with an elliptic curve not in Weierstrass form.

Consider the cubic curve

$$E/\mathbb{Q} : x^3 + y^3 = 1.$$

a) Write down the homogenization E_H of E. Show that $O := [1 : -1 : 0]$ is the only real point at infinity on E. (Note that E has exactly three points at infinity over $\mathbb{C}$.)

b) Show that E_H is nonsingular. Assuming that E_H is irreducible, deduce that E_H is a projective elliptic curve.

c) Thus E is an elliptic curve over $\mathbb{Q}$; in particular $E(\mathbb{Q})$ is a group with identity $O := [1 : -1 : 0]$. Prove that for any point $P = (a, b) \in E(\mathbb{C})$ with $a \neq b$, the inverse of P is

$$-P = (b, a).$$

(You may assume that O is a flex point.)

d) For any point $P = (a, a) \in E$, show that P has order two.

e) (Extra credit) Explain why E has no positive rational points.

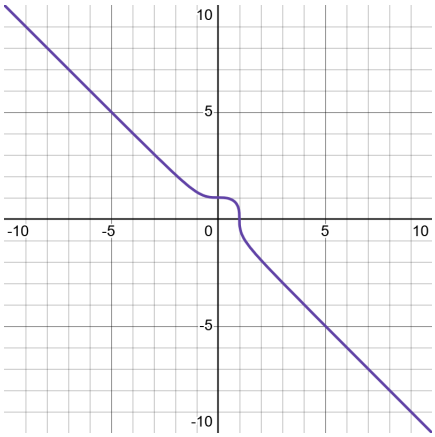


FIGURE 29. The elliptic curve $E : x^3 + y^3 = 1$.

Exercise 5.7.9. This exercise investigates the behavior of the number of points on the elliptic curve

$$E/\mathbb{Q} : y^2 = x^3 + x$$

modulo primes p. You can use https://graui.de/code/elliptic2/ to graph elliptic curves modulo p, as well as compute tables of point additions on them.

For a prime p, we will write $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. As we will discuss in §5.8, we can realize E as an elliptic curve over $\mathbb{F}_p$ for almost all primes p, which also admits a group law via the chord and tangent method. We will use $E(\mathbb{F}_p)$ to denote the group of points on E modulo p, which are simply solutions $(x, y) \in \mathbb{F}_p^2$ to the congruence

$$y^2 \equiv x^3 + x \pmod{p},$$

which includes $O := [0 : 1 : 0]$ once you homogenize this congruence.

a) For primes $p = 3, 7, 11$, compute by hand the set of points $(x_0, y_0) \in \mathbb{F}_p^2$ with $y_0^2 \equiv x_0^3 + x_0 \pmod{p}$.

b) Prove that for any prime $p \equiv 3 \pmod 4$, one has

$$\#E(\mathbb{F}_p) = p + 1.$$

(*Hint:* if $y_0^2 \equiv x_0^3 + x_0 \pmod{p}$, then $x_0^3 + x_0$ is a square modulo p. However -1 is not a quadratic residue modulo p since $p \equiv 3 \pmod 4$.)

c) Create `Sage` code that does the following: given a prime p and an elliptic curve $E : y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$ and $\Delta := -16(4A^3 + 27B^2) \not\equiv 0 \pmod{p}$, it returns the set of points in $E(\mathbb{F}_p)$, as well as the size of $\#E(\mathbb{F}_p)$ (you should include $[0 : 1 : 0]$). Run output for this for $E : y^2 = x^3 + x$ and $2 < p \leq 103$.

d) Based on your calculations in part c), make a conjecture for the size of $E(\mathbb{F}_p)$ when $E : y^2 = x^3 + x$ and $p \equiv 1 \pmod{4}$. (One point)

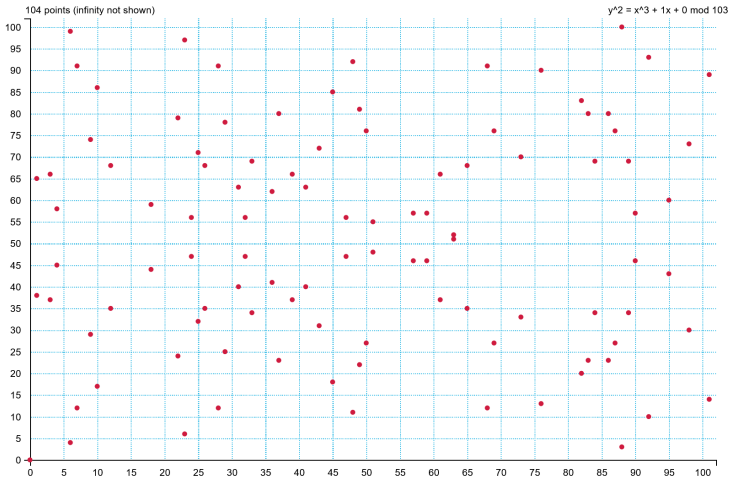


FIGURE 30. The elliptic curve $E : y^2 = x^3 + x$ modulo 103.

Bonus Exercise 5.7.10. Consider the elliptic curve

$$E : y^2 + y = x^3.$$

a. Using the picture below, guess the real flex points of E.

b. With proof, determine the real flex points of E.

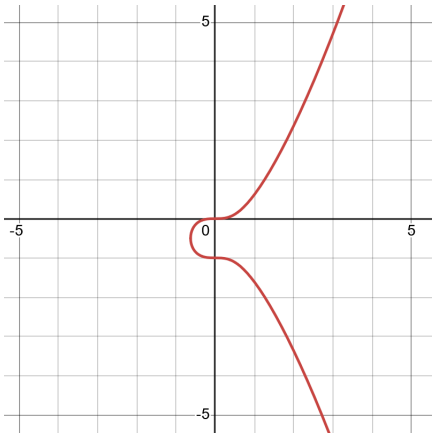


FIGURE 31. The elliptic curve $E : y^2 + y = x^3$.

5.8. **Torsion, Rank and Reduction on Elliptic Curves.** In this last section (diverging from the textbook), we will explore some additional topics on elliptic curves. Here are the main goals.

- Discuss the *torsion subgroup* of an elliptic curve over $\mathbb{Q}$, the *Nagell-Lutz Theorem* and *Mazur's Theorem.*
- Discuss the *rank* of an elliptic curve over $\mathbb{Q}$.
- Explore elliptic curves modulo p, as well as the *reduction map* and the *Hasse-Weil bound.*
- Briefly discuss Fermat's Last Theorem.

The theorems we share in this section will be without proof; however, we will give several examples of their applications.

In the previous section, we reviewed the Mordell-Weil Theorem, which states that for an elliptic curve $E/\mathbb{Q}$ its Mordell-Weil group $E(\mathbb{Q})$ is a finitely generated abelian group. This is impressive on its own, but due to results in abstract algebra, this tells us more about $E(\mathbb{Q})$. The key phrase for those in the know is "the fundamental theorem of finitely generated abelian groups."

Corollary 5.8.1 (Structure Theorem for the Mordell-Weil Group)**.** *For an elliptic curve $E/\mathbb{Q}$, one has*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \times E(\mathbb{Q})[\text{tors}]$$

for some integer $r := r(E, \mathbb{Q}) \geq 0$ and some finite abelian group $E(\mathbb{Q})[\text{tors}]$.

The integer r is called the **rank of E over $\mathbb{Q}$,** which is the maximal number of $\mathbb{Z}$-linearly independent rational points on the curve; and $E(\mathbb{Q})[\text{tors}]$ is the **torsion subgroup of E over $\mathbb{Q}$,** which is the subgroup of rational points with finite order.

Remark 5.8.1. This structure theorem implies that the subgroup of $E(\mathbb{Q})$ of points with finite order is a *finite* subgroup of $E(\mathbb{Q})$ – something which is not obvious. Additionally, the rank r can be zero, in which case $E(\mathbb{Q}) = E(\mathbb{Q})[\text{tors}]$, so that the Mordell-Weil group is finite.

Example 5.8.1. Consider the elliptic curve $E/\mathbb{Q} : y^2 = x^3+5x$. We see that $P := (0, 0)$ lies on E; in fact, it is an order 2 point by Exercise 5.7.5. As it turns out, we have $E(\mathbb{Q})[\text{tors}] = \langle P \rangle = \{O, P\}$, i.e., the only nontrivial rational torsion point on E is P. Additionally, this curve has rank $r = 1$ over $\mathbb{Q}$, with the point $Q := (20, 90) \in E(\mathbb{Q})$ having infinite order. As a consequence of these two facts (which we take for granted), we can conclude that

$$E(\mathbb{Q}) = \langle Q \rangle \oplus \langle P \rangle \cong \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

Thus, any point $R \in E(\mathbb{Q})$ has the form

$$P = a(20, 90) \oplus b(0, 0)$$

for some (unique) $a \in \mathbb{Z}$ and $b \in \mathbb{Z}/2\mathbb{Z}$. For example, we have the point $\left(\frac{1}{4}, -\frac{9}{8}\right) \in E(\mathbb{Q})$, and we can check that

$$\left(\frac{1}{4}, -\frac{9}{8}\right) = (20, 90) - (0, 0).$$

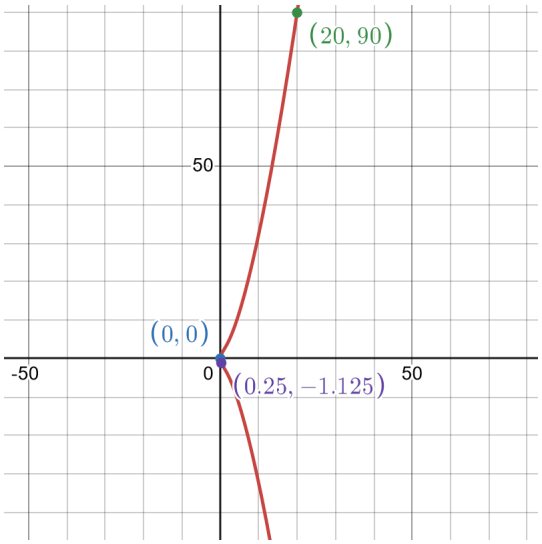


FIGURE 32. The elliptic curve $E : y^2 = x^3 + 5x$ in $\mathbb{R}^2$.

The rank and torsion subgroup of an elliptic curve are central objects of research in modern number theory. Let us highlight what is known about these two invariants; we start with the *torsion subgroup.*

The torsion subgroup of an elliptic curve. Recall that for an elliptic curve $E/\mathbb{Q}$, a point $P \in E(\mathbb{C})$ is a **torsion point** if P has finite order in $E(\mathbb{C})$, i.e., if there exists $n \in \mathbb{Z}^+$ with $nP = O$.

Definition 5.8.1. Given an elliptic curve $E/\mathbb{Q}$, the **torsion subgroup of E over $\mathbb{Q}$** is

$$E(\mathbb{Q})[\text{tors}] = \{P \in E(\mathbb{Q}) : nP = O \text{ for some } n \in \mathbb{Z}^+\}.$$

We also let $E[\text{tors}] := E(\mathbb{C})[\text{tors}]$ the subgroup of $E(\mathbb{C})$ of *all* torsion points.

Remark 5.8.2. It is worth taking a moment to convince yourself that $E(\mathbb{Q})[\text{tors}]$ is indeed a group: that is, if $P, Q \in E(\mathbb{Q})$ have finite order, then so do $P \oplus Q$ and $-P$.

For an elliptic curve $E/\mathbb{Q}$, we have

$$E(\mathbb{Q})[\text{tors}] \subseteq E[\text{tors}].$$

As it turns out, one always has $\#E(\mathbb{Q})[\text{tors}] \leq 16$, and $E(\mathbb{C})[\text{tors}] \cong \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}$ – these facts are not obvious! What is also not obvious is that the x and y-coordinates of torsion points are *algebraic numbers,* i.e., roots of polynomials with integer coefficients. Examples of algebraic numbers are i, $\sqrt{7}$ and $\sqrt[4]{5}+2$; nonexamples are called *transcendental numbers,* and include e^1, π and $\log(2)$. Algebraic numbers are the central objects of study in algebraic number theory.

Definition 5.8.2. For an elliptic curve $E/\mathbb{Q}$, a torsion point $P \in E$ is called an **n-torsion point** if P has order dividing n, i.e $nP = O$. The **n-torsion subgroup of E** is

$$E[n] := \{P \in E(\mathbb{C}) : nP = O\}.$$

We also let $E(\mathbb{Q})[n]$ be the **n-torsion subgroup of E over $\mathbb{Q}$.**

Remark 5.8.3. A general result is that for any elliptic curve $E/\mathbb{Q}$, for each $n \in \mathbb{Z}^+$ one has $E[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. This is not something that is obvious, but knowing that $\#E[n] = n^2$ can prove useful when trying to determine whether you have found all n-torsion points.

Example 5.8.2.

1. Every elliptic curve $E/\mathbb{Q}$ has a fixed rational point $O \in E(\mathbb{Q})$ by definition, chosen as the identity element. This is a torsion point of order one.
2. In Example 5.8.1, we saw that the elliptic curve $E : y^2 = x^3 + 5x$ has an order two torsion point, namely $(0, 0) \in E(\mathbb{Q})[2]$. We also claimed that

$$E(\mathbb{Q})[2] = \{O, (0, 0)\}.$$

   However, if we extend to $\mathbb{C}$, then we have two more 2-torsion points:

$$E[2] = \{O, (0, 0), (\pm\sqrt{-5}, 0)\}.$$

   These are all the 2-torsion points by Remark 5.8.3. (This will also be explored in the next HW.)
3. We have analyzed the elliptic curve $E/\mathbb{Q} : y^2 = x^3 + 17$ several times. As it turns out, we have $E(\mathbb{Q})[\text{tors}] = \{O\}$; we will prove this later in the section. However, this curve's Mordell-Weil group $E(\mathbb{Q})$ turns out to be infinite, with rank $r = 2$. Its *LMFDB* page is here.

   We saw previously that $(-1, 4) \in E(\mathbb{Q})$. Noting that $E(\mathbb{Q})[\text{tors}] = \{O\}$, we immediately conclude that the order of $(-1, 4)$ is *infinite* – in particular, you can add $(-1, 4)$ to itself an infinite amount of times to create an infinite amount of distinct rational points on E!
4. In contrast to the previous examples, the elliptic curve $E/\mathbb{Q} : y^2 = x^3 + 7$ has *no* nontrivial rational points, i.e., its Mordell-Weil group $E(\mathbb{Q})$ is trivial. Proving this is beyond the scope of our course, but in Exercise 5.7.6 you will prove this curve has no integral points using elementary techniques.

Example 5.8.3. Here is a pictorial example of a torsion point $P := (23, -120)$ and its multiples on the elliptic curve $E : y^2 = x^3 + 93x + 94$. Try to convince yourself why this is a point of order six.

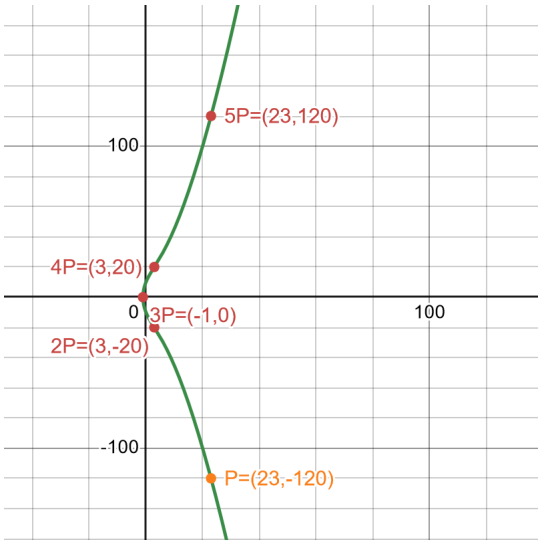


FIGURE 33. The rational elliptic curve $E : y^2 = x^3 + 93x + 94$, pictured in $\mathbb{R}^2$ with nP for $P := (23, -120)$ and $1 \leq n \leq 5$.

Many examples of torsion points you come across will have integer coordinates. As it turns out, this is not a complete coincidence: rational torsion points of order ≥ 3 on elliptic curves in short Weierstrass form are *integral*. The following theorem describes this; it also gives a way to check for rational torsion points.

Theorem 5.8.2 (The Nagell-Lutz Theorem)**.** *Consider an elliptic curve*

$$E/\mathbb{Q} : y^2 = x^3 + Ax + B.$$

If $P \in E(\mathbb{Q})[\text{tors}]$ *is a nontrivial torsion point, then* P *is integral. Furthermore, if* $2P \neq O$*, then writing* $P = (x, y)$ *one has* $y \mid 4A^3 + 27B^2$*.*

Combined with the description of 2-torsion points from the next HW, the Nagell-Lutz Theorem gives one way to completely determine torsion subgroups of rational elliptic curves.

Example 5.8.4. Let us see how the Nagell-Lutz Theorem can be applied when computing torsion subgroups. Consider the elliptic curve

$$E/\mathbb{Q} : y^2 = x^3 - 1.$$

We spot the point $(1, 0) \in E(\mathbb{Q})$.

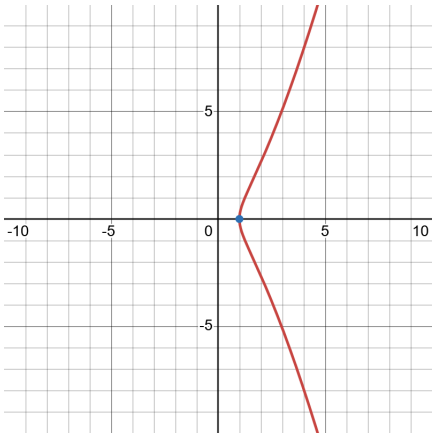


FIGURE 34. The elliptic curve $E : y^2 = x^3 - 1$ in $\mathbb{R}^2$.

We have $A = 0$ and $B = -1$, so that $4A^3 + 27B^2 = -27 = -3^3$, which has eight divisors: $\pm 1, \pm 3, \pm 9$ and ± 27. Thus, if $P = (x, y) \in E(\mathbb{Q})$ is a torsion point of order $\neq 2$, then by the Nagell-Lutz Theorem $|y| = 1, 3, 9$ or 27.

From $y^2 = x^3 - 1$, one has $x^3 = y^2 + 1$. Thus, we can plug in the eight possibilities for y and check directly whether $x = \sqrt[3]{y^2 + 1}$ is an integer to determine whether we have a torsion point of order > 2. For example, if $y = \pm 1$ then $x^3 = (\pm 1)^2 + 1 = 2$, which has no integral solutions. If $y = \pm 3$ then $(\pm 3)^2 + 1 = 10$, and $\sqrt[3]{10} \notin \mathbb{Z}$. Similarly $(\pm 9)^2 + 1 = 82$ and $(\pm 27)^2 + 1 = 272$, neither of which are cubes of integers. We conclude that any torsion point on E has order at most two. Exercise 5.7.5 then lets us deduce that $E(\mathbb{Q})[2] = \{O, (1, 0)\}$, and thus conclude $E(\mathbb{Q})[\text{tors}] = \{O, (1, 0)\}$. This torsion subgroup can be verified here.

Example 5.8.5. It can be the case that for an elliptic curve $E/\mathbb{Q} : y^2 = x^3 + Ax + B$, the number $4A^3 + 27B^2$ has many divisors, making Nagell-Lutz impractical to use by hand. Later in this section, we will have another technique to compute torsion subgroups: via *reduction.*

As noted previously, torsion subgroups of elliptic curves are finite abelian groups. If you are familiar with abstract algebra, then you may know there are only finitely many groups of a fixed order. An interesting question to ask, then, is what possibilities there are for torsion subgroups of elliptic curves. This question has been completely settled by Barry Mazur (my PhD advisor's PhD advisor).

Theorem (Mazur's Theorem)**.** *Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})[\text{tors}]$ is one of the following 15 finite abelian groups, up to isomorphism:*

$$E(\mathbb{Q})[\text{tors}] \cong \begin{cases} \mathbb{Z}/N\mathbb{Z}, & \text{for some } N = 1, 2, \ldots, 10, 12; \\ \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z}, & \text{for some } N = 1, 2, 3, 4. \end{cases}$$

Moreover, each group above is the torsion subgroup of some elliptic curve $E/\mathbb{Q}$.

It is an active area of research to determine which finite abelian groups a torsion subgroup $E(F)[\text{tors}]$ can be over varying fields F, such as number fields. For example,

following Mazur's Theorem there are partial classification results for what $E(F)[\text{tors}]$ can be when the field extension degree $[F : \mathbb{Q}]$ is bounded. However, full results are only known when $[F : \mathbb{Q}] \leq 3$, i.e., when $F = \mathbb{Q}$ or when F is a quadratic or cubic number field.

The rank of an elliptic curve. As seen in Mazur's Theorem above, the possible torsion subgroups for an elliptic curve over $\mathbb{Q}$ is completely known, with sizes bounded by 16. In contrast, **it is unknown whether ranks of elliptic curves over $\mathbb{Q}$ are bounded.**

Every few years, a new elliptic curve gets discovered with a higher guaranteed rank than the last known lower bound. Here is a history of known lower bounds on the largest rank, with some dates omitted:

- rank ≥ 3, 1939 (Billing).
- rank ≥ 4, 1945 (Wiman).
- rank ≥ 6, 1975 (Penney, Pomerance).
- rank ≥ 12, 1982 (Mestre).
- rank ≥ 21, 1994 (Nagao, Kouya).
- rank ≥ 24, 2000 (Martin, McMillen).
- rank ≥ 28, 2006 (Elkies).
- rank ≥ 29, 2024 (Elkies-Klagsbrun).

The largest known rank *equality* is $r = 20$, due to Elkies-Klagsbrun (2020). For more information on ranks, see here, or search "history of elliptic curve rank records." This site also lists for each rank an example of an elliptic curve with at least that many linearly independent points. For example, the 2024 Elkies-Klagsbrun result finds 29 linearly independent points on the elliptic curve

$$E : y^2 + xy = x^3 - 270061832416309222184346521452974537847680546218363579547373 85x + 55258058551342376475736699591118191821521067032535079608372404779149413277716173425636721497$$

(65 and 92 digits, respectively). The points they provided were equally complicated!

Elliptic curves mod p, and reduction. A notion of elliptic curves we have yet to touch on is that of elliptic curves over *finite fields,* i.e., fields with finite cardinality. The most important class of finite fields are those of the form $\mathbb{Z}/p\mathbb{Z}$, where p is prime (recall Theorem 2.11.6). Up to this point, the elliptic curves we considered were defined over fields which contain $\mathbb{Q}$; this is the "characteristic zero" case. Now we consider the finite "positive characteristic" case. From here on out, for each prime p we write $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. To simplify matters (and avoid defining *algebraic closures*), we will only consider elliptic curves given in Weierstrass form.

To not belabor the point, we first note that the notion of curves, the projective plane, and homogenization, singular points and irreducibility all have definitions that extend to equations defined by congruences modulo p. With this in mind:

Definition 5.8.3. For a prime p, the set of solutions to a nonsingular congruence

$$y^2 + a_1xy + a_3y \equiv x^3 + a_2x^2 + a_4x + a_6 \pmod{p}$$

where $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}_p$ is called an **elliptic curve over $\mathbb{F}_p$,** written as $E/\mathbb{F}_p$.

One major subtlety arises when working modulo 2: the equation $y^2 = x^3 + Ax + B$ is *always* singular modulo 2, with a singular point guaranteed over the algebraic closure of $\mathbb{F}_2$. To avoid this, we introduce a useful invariant of an elliptic curve in Weierstrass form.

Definition 5.8.4. For a field F and a cubic curve

$$C/F : y^2 = x^3 + Ax + B,$$

the **discriminant** of C is $\Delta := \Delta(C, F) := -16(4A^3 + 27B^2)$. If C is given in general Weierstrass form

$$C/F : y^2 + z_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

then the discriminant is

$$\Delta := -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$

where

$$\begin{aligned} b_2 &= a_1^2 + 4a_4, \\ b_4 &= a_1a_3 + 2a_4, \\ b_6 &= a_3^2 + 4a_6, \\ b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2. \end{aligned}$$

Proposition 5.8.3. *For a field F and a curve C/F in Weierstrass form, one has that C is nonsingular (and is thus an elliptic curve) if and only if $\Delta \neq 0$ in F.*

We will often calculate discriminants by hand only when the curve is in short Weierstrass form.

Example 5.8.6.

- For any $A, B \in \mathbb{F}_2$, the congruence

  $$C : y^2 \equiv x^3 + Ax + B \pmod 2$$

  has discriminant $\Delta \equiv 0 \pmod 2$, so C is automatically singular.
- For a more direct example, consider the curve

  $$D/\mathbb{F}_2 : y^2 = x^3 + 1.$$

  Then D has the $\mathbb{F}_2$-rational singular point $(0, 1) \in D(\mathbb{F}_2)$, as can be checked.
- Consider for a prime p the curve

  $$E/\mathbb{F}_p : y^2 = x^3 + 2x + 1.$$

  Then $\Delta = -16(4 \cdot 2^3 + 27 \cdot 1^2) = -16 \cdot 59$. Thus, when $p \neq 2, 59$ we find that $E/\mathbb{F}_p$ is an elliptic curve.

Elliptic curves over finite fields do not have the smooth shapes we have seen in the real plane. For example, here is an elliptic curve pictured in the "finite plane" $\mathbb{F}_{103}^2$:

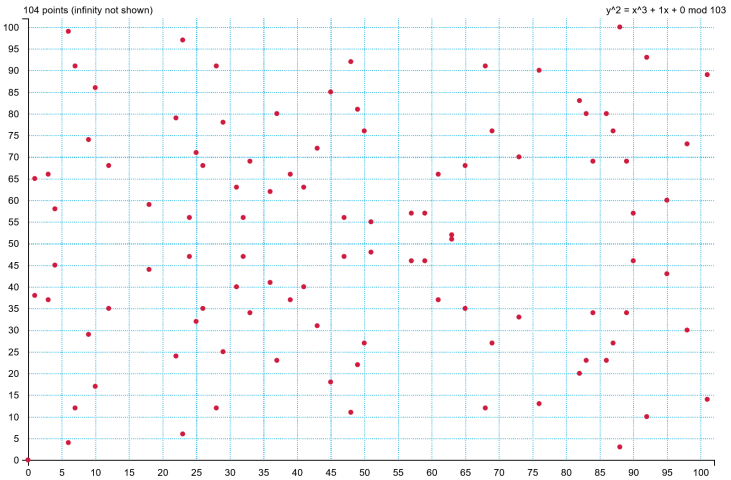


FIGURE 35. The elliptic curve $E : y^2 = x^3 + x$, pictured in $\mathbb{F}_{103}^2$. Here is the site used to graph it.

(Do you spot any symmetries in the picture above?)

Even through the picture over finite fields is radically different, for an elliptic curve $E/\mathbb{F}_p$ the group $E(\mathbb{F}_p)$ is *still* an abelian group under the usual chord and tangent method, which holds algebraically.

Theorem 5.8.4. *For an elliptic curve $E/\mathbb{F}_p$, the chord and tangent method makes $E(\mathbb{F}_p)$ a* finite *abelian group.*

Remark 5.8.4. The formulas from Exercises 5.7.1 and 5.7.2 for the group law of an elliptic curve over $\mathbb{Q}$ in Weierstrass form still hold over finite fields – just keep in mind that undefined values may suggest vertical tangents, or a necessity to work in the projective plane.

We have previously seen that for Diophantine equations, local information guides global solutions: for example, if a Diophantine equation does not have a solution modulo some prime p, then it cannot have an integral solution. Elliptic curves are no exception – however, one can leverage their additional group structure to say more about their rational points. This leads us to the notion of *reduction.*

Definition 5.8.5. Consider an elliptic curve

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$. Its discriminant $\Delta := \Delta(E, \mathbb{Q})$ is an integer. For a prime p, we can reduce this equation to get a cubic curve over $\mathbb{F}_p$, which we denote by $\widetilde{E}$:

$$\widetilde{E}/\mathbb{F}_p : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

The discriminant of this cubic curve over $\mathbb{F}_p$ satisfies

$$\Delta(E, \mathbb{F}_p) \equiv \Delta \pmod{p}.$$

Thus $\Delta \not\equiv 0 \pmod{p}$ if and only if $\widetilde{E}$ is an elliptic curve over $\mathbb{F}_p$; in this case, we say that E has **good reduction at p.** Otherwise, we say that E has **bad reduction at p.**

Definition 5.8.6. Given an elliptic curve $E/\mathbb{Q}$ defined over $\mathbb{Z}$, for a prime p let us define the **mod-p reduction map**

$$\text{red}\colon E(\mathbb{Q}) \to \widetilde{E}(\mathbb{F}_p)$$

as follows. For a point $P = \left(\frac{a}{b}, \frac{c}{d}\right) \in E(\mathbb{Q})$, we can write P in projective coordinates as

$$P = \left[\frac{a}{b} : \frac{c}{d} : 1\right] = [ad : bc : bd] = [a' : b' : c'],$$

where $a', b', c' \in \mathbb{Z}$ are such that at least one of a', b', c' is coprime to p. Then we define

$$\text{red}\left(\frac{a}{b}, \frac{c}{d}\right) := [a' \mod p : b' \mod p : c' \mod p].$$

Example 5.8.7. Let us revisit Example 5.6.5, which was our first chord and tangent method example. We have the elliptic curve

$$E/\mathbb{Q} : y^2 = x^3 + 17.$$

Its discriminant is $\Delta = -16(4A^3 + 27B^2) = -16 \cdot 27 \cdot 17^2$. Thus E has good reduction at all primes $p \neq 2, 3, 17$. Consider $p = 5$: then we have the map

$$\text{red}\colon E(\mathbb{Q}) \to \widetilde{E}(\mathbb{F}_5).$$

We saw that $\left(-\frac{8}{9}, -\frac{109}{27}\right) \in E(\mathbb{Q})$. Let us reduce this point modulo 5:

$$\begin{aligned}
\text{red}\left(-\frac{8}{9}, -\frac{109}{27}\right) &= \text{red}\left[-\frac{8}{9} : -\frac{109}{27} : 1\right] \\
&= \text{red}[-3 \cdot 8 : -109 : 27] \\
&= \text{red}[-24 : -109 : 27] \\
&:= [-24 \mod 5 : -109 \mod 5 : 27 \mod 5] \\
&= [-4 \mod 5 : -4 \mod 5 : 2 \mod 5] \\
&= [1 : 1 : 2].
\end{aligned}$$

Let us double-check that $[1 : 1 : 2]$ is in $\widetilde{E}(\mathbb{F}_5)$. Homogenizing this mod-$p$ equation gives

$$\widetilde{E}_H : Y^2Z \equiv X^3 + 17Z^3 \pmod{5},$$

and we check that $1^2 \cdot 2 \equiv 1^3 + 17 \cdot 2^3 \pmod{5}$.

There is a particularly nice connection between the torsion subgroup of an elliptic curve over $\mathbb{Q}$ and its reductions, described in the following theorem. Similar to the Nagell-Lutz Theorem, it can help us calculate torsion subgroups.

Theorem 5.8.5. *For an elliptic curve $E/\mathbb{Q}$ defined over $\mathbb{Z}$, if $p \geq 3$ is a prime of good reduction for E, then the reduction map*

$$\text{red}\colon E(\mathbb{Q})[\text{tors}] \to \widetilde{E}(\mathbb{F}_p)$$

is an injective group homomorphism.

Example 5.8.8. Let us consider the familiar elliptic curve $E : y^2 = x^3 + 17$. We previously checked that E has good reduction at all primes $p \neq 2, 3, 17$. We claimed in Example 5.8.2 that $E(\mathbb{Q})[\text{tors}] = \{O\}$, which we now prove. One checks that $\#\widetilde{E}(\mathbb{F}_5) = 6$ and $\#\widetilde{E}(\mathbb{F}_7) = 13$. By Theorem 5.8.5, we know that $E(\mathbb{Q})[\text{tors}]$ injects as a subgroup into these two groups. Therefore, by our version of Lagrange's Theorem, any element $P \in E(\mathbb{Q})[\text{tors}]$ has order dividing both 6 and 13, which forces $|P| = 1$, i.e. $P = O$. We conclude that $E(\mathbb{Q})[\text{tors}]$ is trivial.

As we have noted, for an elliptic curve $E/\mathbb{F}_p$ the Mordell-Weil group $E(\mathbb{F}_p)$ is a finite abelian group. One natural question is *what is the size of* $E(\mathbb{F}_p)$*?* Since $E(\mathbb{F}_p) \subseteq \mathbb{F}_p^2 \cup \{O\}$, one has the naïve bound

$$\#E(\mathbb{F}_p) \leq p^2 + 1.$$

However, there is a classic result which gives a sharper bound.

Theorem 5.8.6 (Hasse-Weil bound)**.** *For an elliptic curve $E/\mathbb{F}_p$, one has*

$$|p + 1 - \#E(\mathbb{F}_p)| \leq 2\sqrt{p}.$$

Given an elliptic curve $E/\mathbb{Q}$ defined over $\mathbb{Z}$, for each prime p of good reduction the Hasse-Weil bound gives a bound on the discrepancy between $\#\widetilde{E}(\mathbb{F}_p)$ and $p + 1$. This discrepancy, written as

$$a_p := a_p(E) := p + 1 - \#E(\mathbb{F}_p),$$

is called the **trace of Frobenius of E modulo p.** There are several outstanding conjectures on the values and distribution of a_p for varying $E/\mathbb{Q}$ and for varying p; see for example the *Sato-Tate conjecture* and *Lang-Trotter conjecture.* In Exercise 5.7.9, one shows that for the elliptic curve $E/\mathbb{Q} : y^2 = x^3 + x$, one has for each prime $p \equiv 3 \pmod 4$ that $a_p = 0$, i.e. $\#E(\mathbb{F}_p) = p + 1$.

One more note regarding elliptic curves over finite fields: they are a crucial component of *elliptic curve cryptography.* Elliptic curves are also a centerpiece of post-quantum cryptography, particularly in *isogeny-based cryptography.* We will not touch on this in our course.

Fermat's Last Theorem. To conclude these notes, let us discuss one of the most important results in modern mathematics.

Theorem 5.8.7 (Fermat's Last Theorem)**.** *There are no positive integral solutions to the equation*

$$x^n + y^n = z^n$$

when $n \geq 3$.

We discussed this theorem briefly in §5.3; as noted previously, when $n \leq 2$ there are infinitely many integral solutions, and they are describable/parametrizable. In 1637, Pierre de Fermat claimed he had a proof of this result for $n \geq 3$ "too large to fit in the margin" of his copy of *Arithmetica* (a book written by Diophantus). It was likely that the proof he had in mind was incorrect.

Fermat's Last Theorem was fully proven 300+ years after Fermat's claim, in 1994/1995. The proof uses high level arithmetic geometry, which would take years to understand. A

key idea in this proof is to show that a positive integral solution (a, b, c) to $x^n + y^n = z^n$ implies the existence of a certain elliptic curve $E_{a,b,c}$ which has self-contradicting properties.

Let us explain this a bit more, beyond what was covered in §5.3. Suppose for contradiction that there exists a positive integral solution:

$$a^n + b^n = c^n,$$

with $a, b, c \in \mathbb{Z}^+$. From this, we define the **Frey (elliptic) curve**

$$E_{a,b} : y^2 = x(x - a^n)(x - b^n).$$

In 1986, Ken Ribet proved that a Frey curve is not *modular*. However, Wiles' proof showed that an elliptic curve of the form $E : y^2 = x(x - A)(x - B)$ must always be modular when $A, B \in \mathbb{Q}$. Contradiction!

What does "modular" mean? This is something to learn in an advanced number theory course! *Fin.*

Exercises.

Bonus Exercise 5.8.1. This exercise deals with the ":)-theorem."

In the following, let us define the **radical** function: for $\in \mathbb{Z}^+$, we set

$$\text{rad}\left(\right) := \prod_{\text{prime} \mid } .$$

Then the :)-theorem is as follows.

Theorem (:)-theorem)**.** *For each* > 0*, there are finitely many* $\in \mathbb{Z}^+$ *with* $\gcd\left(,,\right) = 1$ *and* $+ = $*, such that*

$$> \text{rad}\left(\cdot \cdot \right)^{1+} .$$

Prove the :)-theorem. (*Hint:* good luck!)

References

[LMFDB] The LMFDB Collaboration, The L-functions and modular forms database, https://www.lmfdb.org, 2024, [Online; accessed April 14 2026].

[NZM91] I. Niven, H.S. Zuckerman and H.L. Montgomery, *An introduction to the theory of numbers*, 5th Ed., John Wiley & Sons, Inc., New York (1991).

[Rou91] G. Rousseau, *On the Quadratic Reciprocity law*, J. Austral. Math. Soc. Ser. A (1991), no. 3, 423–425.

[ST15] J. Silverman and J. Tate, *Rational points on elliptic curves*, 2nd Ed., Undergraduate Texts in Mathematics, Springer, Cham (2015).