CLASS NOTES FOR MATH 4573: ELEMENTARY NUMBER THEORY

TYLER GENAO

Disclaimer: these are my class notes for an elementary number theory course (MATH 4573) that I taught at OSU during Spring 2024. I primarily followed [NZM91] for these notes. These notes include those that my lecture would follow in class, along with homework problems at the end of each section. Because of the former point, some parts might be considered terse, as it was "optimized" for my lecture. If you find these notes helpful, or have comments or suggestions, feel free to reach out to me at genao.5@osu.edu – I would love to hear it!

(In these notes, "Exercises" were assigned as homework for the class, and "Bonus Exercises" were optional homework problems.)

Contents

1. Chapter 1: Divisibility

1.1. **Introduction.** In a general sense, number theory is the subject which studies the arithmetic (additive and multiplicative properties) of the integers $0, \pm 1, \pm 2, \ldots$. We'll use $\mathbb{Z}$ to denote the set of integers, $\mathbb{Z}^+$ the set of natural numbers (excluding zero!) and $\mathbb{Q}$ the set of rational numbers (fractions of integers).

Here are some examples of number-theoretic results:

1. An integer n is divisible by 3 if and only if the sum of its digits is divisible by 3;
2. the equation $x^2 + y^2 = z^2$ has infinitely many **integral** solutions ($x, y, z \in \mathbb{Z}$);
3. for integers $n \geq 3$, the equation $x^n + y^n = z^n$ has **no** nontrivial solutions in $\mathbb{Z}^3$ ($(x, y, z) \neq (0, 0, 0)$) (Fermat's Last Theorem, proven in 1995 by Andrew Wiles);
4. There are infinitely many **prime numbers** (positive integer divisible only by 1 and itself) (Euclid).

Number theory is often motivated by **empirical data.** Some results include:

a) Every natural number is a sum of 4 squares: $n = a^2 + b^2 + c^2 + d^2$. This is TRUE, proven by Lagrange in the 1700's.

b) No n'th power is a sum of fewer than n n'th powers. This is FALSE: for example, in 1987 Noam Elkies used elliptic curves to show that

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

c) *Goldbach Conjecture:* every integer $n > 2$ is the sum of two primes. For example, $4 = 2 + 2$, $6 = 3 + 3$, $20 = 7 + 13$, $50 = 3 + 47$, $100 = 29 + 71$. This; is OPEN: it's true for $n < 40,000,000$, but is it true for all $n \in \mathbb{Z}^+$?

Note: For a new theorem, it's always good to practice some small examples to get a feeling for why it might be true.

Number theory is a *very* wide subject. Much of the research in modern number theory falls into one of the following subfields (which are not mutually exclusive):

1. Algebraic number theory: generalizations of $\mathbb{Z}$ and $\mathbb{Q}$ to other algebraic structures whose elements mimic the arithmetic behavior of integers and fractions.
2. Analytic number theory: estimates on proportions and distributions of positive integers, using analytic techniques.
3. Arithmetic geometry: studying rational and integral solutions to algebraic curves (and varieties).

This is not a comprehensive list!

Two more general principles are worth noting, since they show up in many proofs in number theory (they are in fact equivalent):

1. **Principle of Well-Ordering:** any nonempty subset $X \subseteq \mathbb{Z}^+$ has a *least positive element* (i.e., a smallest number).

2. **Principle of Induction:** If a statement $P(n)$ is true for $n = 1$, and is true for $P(n)$ whenever it is true for $P(n-1)$, then $P(n)$ is true for all $n \in \mathbb{Z}^+$.

1.2. **Divisibility.** Here is the first definition for our class.

Definition 1.2.1. An integer a is divisible by nonzero integer b if there exists $x \in \mathbb{Z}$ with $bx = a$. In this case, we will write $b \mid a$.

Example 1.2.1. $1 \mid 12$, $-3 \mid 15$, $20 \mid 100$, $7 \nmid 5$.

If $b \mid a$ and $b \neq a$, we say that b is a *proper divisor of* a.

Theorem 1.2.1 (Theorem A)**.** *The following statements are true for integers $a, b \in \mathbb{Z}$.*

(1) If $b \mid a$, then $b \mid ac$ for all $c \in \mathbb{Z}$.
(2) If $b \mid a$ and $c \mid b$, then $c \mid a$.
(3) If $b \mid a$ and $b \mid c$, then for all integers $x, y \in \mathbb{Z}$ one has $b \mid (ax + cy)$.
(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 $c \neq 0$, then $mb \mid ma$.

Here is a proof of one of these items (3).

Proof. Assume that $b \mid a$ and $c \mid a$; then we can write $bm = a$ and $bn = c$ for some $m, n \in \mathbb{Z}$. Thus, for any $x, y \in \mathbb{Z}$, we get

$$ax + cy = (bm)x + (bn)y = b(mx + ny),$$

and thus $b \mid (ax + cy)$. □

The next theorem concerns division of an integer. It isn't quite an algorithm – but it can be made into one (which we'll see after this).

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

Proof. Consider the set of integers

$$A := \{\ldots a - 3b, a - 2b, a - b, a, a + b, a + 2b, a + 3b, \ldots\},$$

i.e., $A := \{a + kb : k \in \mathbb{Z}\}$. Then by the **well-ordering principle,** A has a least non-negative element; call it r. Then r has the form $r = a - qb$ for some $q \in \mathbb{Z}$, and so $a = qb + r$. Observe that $r < b$: otherwise, we'd have that $a - (q-1)b < r$, contradicting minimality of r.

We need to show that q and r are unique. For contradiction, suppose there's another element $r_1 = a - q_1 b$ of A with $0 \leq r_1 < a$. By minimality of r, we know $r_1 - r > 0$. Now, $r_1 - r = b(q_1 - q)$, so that $b \mid (r_1 - r)$; and since $b, r_1 - r > 0$, we must have by Theorem 1.2.1 that $b \leq r_1 - r$, which is impossible since $0 \leq r_1 < b$. □

Note that the assumption that $b > 0$ is not necessary: the conclusion would instead be stated as $a = bq + r$, $0 \leq r < |b|$.

We'll often want to compare divisors of two integers. To this end, we will define the following.

Definition 1.2.2. Given two integers a and b, the *greatest common divisor of a and b,* (GCD) written $\gcd(a, b)$, is the largest integer $c \in \mathbb{Z}^+$ with $c \mid a$ and $c \mid b$.

Example 1.2.2. $\gcd(4, 8) = 4$; $\gcd(6, 15) = 3$; $\gcd(7, 20) = 1$.

It turns out that the GCD of any two integers is a linear combination of the two.

Theorem 1.2.3. *If g is the GCD of a and b, then there exist x and y with $g = ax + by$.*

Proof. Consider the set of linear combinations of a and b,

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

By the well-ordering principle, there exists a least positive $h \in B$; write $h = ax + by$.

We claim that h is the GCD of a and b. First, we claim that $h \mid a$ and $h \mid b$. By the division algorithm, we get $a = qh + r$, where $0 \leq r < h$. For contradiction, suppose that $h \nmid a$; then $r \neq 0$. But since

$$r = a - qh = a - q(ax + by) = a(1 - qx) + b(-qy),$$

we have $r \in B$, contradicting the minimality of h. Thus, we have $h \mid a$, and a similar argument shows $h \mid b$.

By definition of the GCD, we have $h \leq g$. But since $h = ax + by$ and $g \mid a$ and $g \mid b$, we have that $g \mid h$ by (3) from Theorem A, so that $g = h$. □

Thus, the GCD has two definitions:

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

Here are some more GCD properties.

Theorem 1.2.4 (Theorem B)**.** *The following properties hold for the GCD of a and b.*

- *For any integer m, one has*
$$\gcd(ma, mb) = m \gcd(a, b).$$
- *If $d \mid a$ and $d \mid b$ and $d > 0$, then*
$$\gcd(a/d, b/d) = \gcd(a, b)/d;$$
in particular, if $g := \gcd(a, b)$ then
$$\gcd(a/g, b/g) = 1.$$
- *If $\gcd(a, m) = \gcd(b, m) = 1$, then*
$$\gcd(ab, m) = 1.$$
- $\gcd(a, b) = \gcd(b, a) = \gcd(a, -b) = \gcd(a, b + am)$.

Proof. Here's a proof of the fourth item: in particular, that $\gcd(a, b + am) = \gcd(a, b)$. Write $g := \gcd(a, b)$ and $h := \gcd(a, b + am)$. Want to show that $g = h$.

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,

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

and so $g \leq h$, by the alternative definition of the GCD. On the other hand,

$$g = aw + bx = aw + bx + amx - amx = a(w - mx) + (b + am)x =: aX + (b + am)Y.$$

Thus, g is a $\mathbb{Z}$-linear combination of a and $b + am$, and so $h \leq g$. Thus, $g = h$. □

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

You'll prove this one:

Theorem 1.2.5. *If $c \mid ab$ and $\gcd(b, c) = 1$, then $c \mid a$.*

Proof. Start with: $\gcd(ab, ac) = a \gcd(b, c) = a$. Now, $c \mid ab$ and $c \mid ac$, thus $c \mid \gcd(ab, ac) = a$. □

Here's a practical way to compute GCD's.

Theorem 1.2.6 (The Euclidean algorithm)**.** *Given integers $a, b \in \mathbb{Z}$ with $b > 0$, one can repeatedly apply the Division Algorithm to compute $\gcd(a, b)$.*

$$\begin{aligned}
a &= bq_1 + r_1,\ 0 < r_1 < b;\\
b &= r_1q_2 + r_2,\ 0 < r_2 < r_1;\\
r_1 &= r_2q_3 + r_3,\ 0 < r_3 < r_2;\\
&\ldots\\
r_{j-2} &= r_{j-1}q_j + r_j,\ 0 < r_j < r_{j-1};\\
r_{j-1} &= r_jq_{j+1}.
\end{aligned}$$

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

Proof. To see that r_j is the GCD, observe that

$$\begin{aligned}
\gcd(a, b) &= \gcd(bq_1 + r_1, b) \text{First equation}\\
&= \gcd(bq_1 + r_1 - bq_1, b) \text{Theorem B}\\
&= \gcd(r_1, b)\\
&= \gcd(r_1, r_1q_2 + r_2) \text{Second equation}\\
&\gcd(r_1, r_2)\\
&\ldots\\
&= \gcd(r_{j-1}, r_j)\\
&= \gcd(r_jq_{j+1}, r_j)\\
&= r_j.
\end{aligned}$$

Next, we will show that $\gcd(a, b)$ is a linear combination of a and b via induction. Clearly, $r_1 = a - bq_1$ is a linear combination of a and b. Suppose then that we can write r_i as a linear combination of a and b for $i < j$. Then we have $r_j = r_{j-2} - r_{j-1}q_j$, and since each r_{j-2} and r_{j-1} is a linear combination of a and b, so is r_j. □

Example 1.2.3. Compute the GCD of 42823 and 6409, and express it as a $\mathbb{Z}$-linear combination of the two.

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

Thus, $\gcd(42823, 6409) = 17$.

As for the combination: the goal is to compute each auxiliary remainder as a combination using the equations above; each line of equation involves the remainder r_i (i'th step), and combines the linear combinations of r_{i-1} and r_{i-2} (sans including the first and second step).

From these calculations, we have $a = 42823, b = 6409, r_1 = 4369, r_2 = 2040, r_3 = 289$ and $r_4 = 17 = \gcd(a, b)$. First,

$$r_1 = a - 6b.$$

Then,

$$r_2 = b - r_1 = b - (a - 6b) = -a + 7b.$$

Then,

$$r_3 = r_1 - 2r_2 = (a - 6b) - 2(-a + 7b) = 3a - 20b.$$

Finally,

$$r_4 = \gcd(a, b) = r_2 - 7r_3 = (-a + 7b) - 7(3a - 20b) = -22a + 147b.$$

Therefore, we have

$$\gcd(42823, 6409) = -22 \cdot 42823 + 147 \cdot 6409.$$

A nice way to compute the GCD and $\mathbb{Z}$-linear combination is with **Blankinship's algorithm.** The goal is to compute $\gcd(a, b)$ via computing each r_i one step at a time, using elementary row operations.

1. Given $a, b \in \mathbb{Z}$, 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, the second Equation B.
2. Since $r_1 = a - bq_1$, subtract $B \cdot q_1$ from A to get
$$\left[\begin{array}{c|cc} a & 1 & 0 \\ b & 0 & 1 \end{array}\right] \xrightarrow{A \mapsto A - q_1 B} \left[\begin{array}{c|cc} a - bq_1 = r_1 & 1 & -q_1 \\ b & 0 & 1 \end{array}\right].$$
3. Since $r_2 = b - r_1 q_2$, subtract $A \cdot q_2$ from B to get
$$\left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ b & 0 & 1 \end{array}\right] \xrightarrow{B \mapsto B - A \cdot q_2} \left[\begin{array}{c|cc} r_1 & 1 & -q_1 \\ b - r_1 q_2 = r_2 & -q_2 & q_1 q_2 \end{array}\right].$$
4. Repeat this until you obtain 0 on the left column somewhere; then the remaining nonzero number is $\gcd(a, b)$, and its right entries give the $\mathbb{Z}$-linear combination.

Example 1.2.4. Write the GCD of 267 and 112 as a $\mathbb{Z}$-linear combination of the two.

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

Therefore, $\gcd(267, 112) = 1$ and $1 = 267 \cdot (-13) + 112 \cdot (31)$.

Antithetic to the GCD, let's define the *least common multiple.*

Definition 1.2.4. Given $a, b \in \mathbb{Z}$, the least common multiple of a and b is the smallest positive common multiple c (meaning $a \mid c$ and $b \mid c$). Written $\operatorname{lcm}(a, b)$.

Theorem 1.2.7. *Any common multiple c of a and b satisfies* $\operatorname{lcm}(a, b) \mid c$.

Proof. Suppose $a \mid c$ and $b \mid c$. Since $\operatorname{lcm}(a, b) \leq c$, by the division algorithm, let us write $c = \operatorname{lcm}(a, b)q + r$, $0 \leq r < \operatorname{lcm}(a, b)$. Thus, $r = c - \operatorname{lcm}(a, b)q$, so that $a \mid r$ and $b \mid r$, so r is a common multiple with $0 \leq r < \operatorname{lcm}(a, b)$. By minimality of $\operatorname{lcm}(a, b)$, this forces $r = 0$, so that $\operatorname{lcm}(a, b) \mid c$. □

Exercise 1.2.1 (Playing with GCD's)**.** This exercise will get you acquainted with GCD calculations by hand.

a) Use the Euclidean algorithm to compute the GCD of $a = 1876$ and $b = 365$.
b) Next, compute the GCD of $a = 1876$ and $b = 365$ using Blankinship's algorithm with a matrix. Using this work, also write your GCD as a $\mathbb{Z}$-linear combination of a and b.
c) Use Blankinship's algorithm to compute the GCD of $a = 4999$ and $b = 1109$, and write this GCD as a $\mathbb{Z}$-linear combination of a and b.

Exercise 1.2.2 (Trial run: induction)**.** Prove the following result, using the proof technique of **induction:**

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

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

We'll break this proof down into steps:

b) *Check the base case:* write down what you got for the first case $n = 1$, and confirm that it's a multiple of 5.

c) *Induction hypothesis:* assume that the proposition is true for all integers $1 \leq k < n$. Using this assumption, prove the proposition for $k = n$.

(*Hint:* for c), observe that $9^n - 4^n = (5+4) \cdot 9^{n-1} - 4 \cdot 4^{n-1}$.)

Exercise 1.2.3 (Trial run: contradiction)**.** Prove the following result, using the proof technique of **contradiction:**

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

Exercise 1.2.4 (Trial run: contrapositive)**.** We will prove the following result by proving its equivalent **contrapositive:**

Proposition. *Let a be a positive integer. If $a > 1$, 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 $2^a + 1$ for $a \geq 1$; prove it if you can.

Exercise 1.2.5 (How do the GCD and LCM equate?)**.** Show that if $a, b \in \mathbb{Z}^+$ satisfy $\gcd(a, b) = \operatorname{lcm}(a, b)$, then $a = b$.

Exercise 1.2.6 (Sharpening divisibilities)**.** Show that $a \mid bc$ if and only if $\frac{a}{\gcd(a,b)} \mid c$.

1.3. **Primes.**

Definition 1.3.1. An integer $p > 1$ is **prime** if it has no proper divisors, i.e., no $1 < d < p$ with $d \mid p$.

An integer n which is not prime is called **composite.** Divisors of n are called **factors.** (We usually focus only on positive factors.)

For example: 2 is prime, 3 is prime, 4 is *not* prime (2 is a proper divisor), 17 is prime, 20 is *not* prime ($2 \mid 20$), etc.

Theorem 1.3.1. *Every integer $n > 1$ is a product of primes.*

Proof. Fix $n > 1$.

1. If n is already prime, done.
2. If n is composite, factor it as $n = ab$, with $1 < a, b < n$. If a and b are both prime, we're done.
3. If e.g. a is composite, factor it, etc. This process eventually ends since each proper divisor of the previous number is strictly smaller.
4. Thus n is a product of primes.

□

In general, each integer $n > 1$ has a factorization into prime powers

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

where p_i's are primes. As we will show soon, this factorization is always *unique* for an integer n, up to reordering the prime powers. ("The Fundamental Theorem of Arithmetic," or FTA.)

$\mathbb{Z}$ is unique in that not all rings satisfy a FTA! (For example, the "algebraic integer ring" $\mathbb{Z}[\sqrt{-6}] := \{a + b\sqrt{-6} \ : \ a, b \in \mathbb{Z}\}$ fails this.)

Towards proving FTA for $\mathbb{Z}$, we will need the following lemma.

Lemma 1.3.2. [NZM91, Theorem 1.15] *If p is prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. More generally, if $p \mid a_1 a_2 \cdots a_r$, then for some i $p \mid a_i$.*

Proof. If $p \mid ab$ and $p \nmid a$, then $\gcd(p, a) = 1$, so by a theorem from last class [Theorem 1.10] we have $p \mid b$.

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

Theorem 1.3.3 (The Fundamental Theorem of Arithmetic)**.** *The factorization of an integer $n > 1$ is unique up to ordering the prime powers.*

Proof. Suppose we have two factorizations of n into products of primes,

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

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

For each $1 \leq i \leq r$, we have $p_i \mid q_1^{f_1} q_2^{f_2} \cdots q_t^{f_t}$, so by the lemma, for some j we have $p_i \mid q_j^{f_j} = q_j q_j \cdots q_j$; applying the lemma again gives $p_i \mid q_j$, and thus $p_i = q_j$ since q_j is prime. This applies to *every* prime p_i with $1 \leq i \leq r$ (since $p_i \neq p_k$ when $i \neq k$); thus, $r \leq t$. Applying the same argument to the q_j's shows that $t \leq r$, so $r = t$.

Relabel all primes so that $p_i = q_i$:

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

We claim each $e_i = f_i$. If $e_i \neq f_i$ for some i, say (without loss of generality) $e_i < f_i$, then canceling $p_i^{e_i}$ from both sides implies $p_i \mid p_1^{e_1} p_2^{e_2} \cdots \widehat{p_i^{e_i}} \cdots p_r^{e_r}$, which by the lemma implies $p_i \mid p_j$ for some $j \neq i$, impossible. □

Example 1.3.1. Factorize the following:

- $10 = 2 \cdot 5$.
- $80 = 2^4 \cdot 5$.
- $81 = 3^4$.
- $31 = 31$.

Every integer has a unique factorization into prime powers. Given that there infinitely many integers, how many primes are there?

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

Proof. Fix any finite amount of primes: call these $p_1, p_2, \ldots, p_r$. Then 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 $p_i \mid 1$. And since $n > 1$, n must have a prime factor. Therefore, there is a prime which divides n not equal to $p_1, p_2, \ldots, p_r$. □

Sometimes, we will write a factorization as an "infinite" product

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

where $v_p(n)$ is the power of p which divides n ($v_p(n)$ is called the **p-adic valuation of n**). This product is *finite* since almost all $v_p(n)$'s are equal to zero.

Unique factorization makes GCD's and LCM's easy to compute:

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

Similarly,

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

Example 1.3.2. Let $a = 108$ and $b = 225$. Then $a = 2^2 \cdot 3^3$ and $b = 3^2 \cdot 5^2$, and so

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

and

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

Exercise 1.3.1 (Proof or Counterexample?)**.** Determine whether the following are true or false. If a statement is true, then prove it; if it is false, then provide a counterexample.

a) For prime $p \in \mathbb{Z}^+$, if $p \mid a^3$ then $p \mid a$.
b) If $\gcd(a,b) = \gcd(a,c)$ then $\operatorname{lcm}(a,b) = \operatorname{lcm}(a,c)$.
c) For $n \in \mathbb{Z}^+$, if $n \mid a^3$ then $n \mid a$.
d) If $n \mid a^2 - 1$ then $n \mid a^4 - 1$.

Exercise 1.3.2 (Sum odd squares aren't perfect)**.** Recall that a *perfect square* $a \in \mathbb{Z}$ 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 $x^2 + y^2$ isn't a perfect square.

Exercise 1.3.3 (Square minus one)**.** Show that if $n \in \mathbb{Z}$ 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 (Bounds on prime factors)**.**

a) Show that if $n \in \mathbb{Z}^+$ 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 (A difference between primes and composites)**.**

a) Show that $p \nmid (p-1)!$ for all primes $p \in \mathbb{Z}^+$.
b) Show that $n \mid (n-1)!$ for all composite $n > 4$. (*Hint:* Write $n = ab$, where $1 < a, b < n$. Then split the work up into two cases depending on a and b.)

Exercise 1.3.6 (Primes of the form $3 + 4k$)**.** 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 "congruent to 3 modulo 4" (which is the case where $a = 3$ and $b = 4$).

a) Show that an integer $n \in \mathbb{Z}^+$ of the form $3 + 4k$ has at least one prime factor of the same form.
b) Mimicking Euclid's proof on the infinitude of primes (see [NZM91, Theorem 1.17]), use part a) to prove the following:

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

(*Hint:* Construct an n of the form $3 + 4k$.)

Bonus Exercise 1.3.7 (Primes of the form $1 + 4k$)**.** Prove the following result:

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 $n \in \mathbb{Z}$ and any odd prime $p \in \mathbb{Z}^+$, 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 1.3.8 (Furstenberg's proof of the infinitude of primes)**.** This exercise will give a topological proof that there are infinitely many primes.

Let us define a topology on $\mathbb{Z}$ as follows. Say that a subset $U \subseteq \mathbb{Z}$ is open iff it is a union of nonconstant 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

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

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

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

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

The following two exercises will use *computational mathematics* to count the distribution of primes in the natural numbers $\mathbb{Z}^+$.

Bonus Exercise 1.3.9 (The prime number theorem, experimentally)**.** This exercise will attempt to convince ourselves that the *prime number theorem* is true. Let us define the "prime counting function" $\pi\colon \mathbb{R} \to \mathbb{Z}^+$, where for each real number x, the integer $\pi(x)$ is the number of (positive) primes less than or equal to x.

Theorem (The prime number theorem)**.** *One has*

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

a) Create code that calculates $\pi(x)$ for any real number x.
b) Compare the values of $\pi(x)$ and $\frac{x}{\log(x)}$ for $x = 10, 10^2, \ldots, 10^{10}$; analyze what is happening to these two values as x gets increasingly large.
∗c) Find and understand an elementary proof of the prime number theorem.

Bonus Exercise 1.3.10 (Primes in arithmetic progressions, experimentally)**.** Given a pair $(a, b) \in \mathbb{Z}^+ \times \mathbb{Z}^+$, 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) Using a computer, calculate $\pi_{3,4}(x)$ 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 1.3.11 (Conjecture on primes in polynomial progressions)**.** 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 $f(x) \in \mathbb{Z}[x]$ 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) $\gcd(f(1), f(2), f(3), \ldots) = 1$.

Then there are infinitely many primes of the form $f(n)$ where n ranges over $\mathbb{Z}^+$.

Compare the conclusion of the Bunyakovky conjecture to Exercise 2.1.15.

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 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?

Bonus Exercise 1.3.12 (Failure of unique factorization in a number ring)**.** The Fundamental Theorem of Arithmetic is a special result which applies to elements $n \in \mathbb{Z}$ not equal to 0 or ± 1; however, 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 $N \colon \mathbb{Z}[\sqrt{-6}] \to \mathbb{Z}_{\geq 0}$ 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 iff $N(\alpha) = 1$.

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, "irreducible" in R is an analogous notion to "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, this analogy breaks down when considering *uniqueness* of this 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 $\mathbb{Z}[\sqrt{-6}]$.

Therefore, $\mathbb{Z}[\sqrt{-6}]$ does not have a "unique factorization theorem" for its elements. 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 "Dedekind domain.")

1.4. **The Binomial Theorem.** The binomial theorem describes the coefficients in the expansion of a binomial $(x+y)^n$ with $n \in \mathbb{Z}$, which is important for many reasons, including algebraic calculations with integers.

First, an algebraic definition.

Definition 1.4.1. Let $n, k \in \mathbb{Z}$ with $k \geq 0$. Then the **binomial coefficient** $\binom{n}{k}$, read "n choose k", is

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

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

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

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

Lemma 1.4.1 (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)**.** *For any integer $n \geq 1$ and any real numbers x, y, one has*

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

Proof. First, for a set of numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$, 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 of the form

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

where I is any subset of the index set $\{1, 2, \ldots, n\}$.

Consider the monomials corresponding to I with k elements, $0 \leq k \leq n$. In our context, all $x_i = x$ and $y_i = y$, and such a monomial has the form $x^k y^{n-k}$. By the previous lemma, there are $\binom{n}{k}$ such monomials, and thus such indexing sets I contribute the term $\binom{n}{k} x^k y^{n-k}$ to the expansion of $(x+y)^n$. □

It is worth noting that this proof 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 (although we'd need to generalize the binomial coefficient definition a bit).

In Pascal's Triangle, row n gives the binomial coefficients for $\binom{n-1}{k}$.

Exercise 1.4.1 (Sum of all coefficients)**.** Use the binomial theorem to show that for each $n \geq 0$,

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

Exercise 1.4.2 (n'th derivative product formula)**.** Let $f(x)$ and $g(x)$ be n-times differentiable functions. Prove the n'th derivative formula

$$(f(x)g(x))^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(x)g^{(n-k)}(x).$$

2. Chapter 2: Congruences

2.1. **Congruences.** Congruences are another way to write divisibilities. However, this way of thinking lends itself to several important results.

Definition 2.1.1. For $a, b, m \in \mathbb{Z}$, 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

$$a = b + mk.$$

This is equivalent to

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

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

In the situation above, we call m the **modulus**. This is the namesake of "modular arithmetic." We always assume $m > 0$.

Example 2.1.1. To illustrate notation, we have $13 \equiv 1 \pmod{12}$, $4 \equiv -3 \pmod{7}$, and for any odd number $n \in \mathbb{Z}$, $n \equiv 1 \pmod{2}$.

Here are some basic properties of congruences.

Theorem 2.1.1. [NZM91, Theorem 2.1] *Let $a, b, c, d \in \mathbb{Z}$. Then:*

(1) $a \equiv b \pmod{m}$, $b \equiv a \pmod{m}$ and $a - b \equiv 0 \pmod{m}$ are equivalent statements.
(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 $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$.
(4) If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$.
(5) If $a \equiv b \pmod{m}$ and $d \mid m$ with $d > 0$, then $a \equiv b \pmod{d}$.
(6) If $a \equiv b \pmod{m}$, then for all $c > 0$, $ac \equiv bc \pmod{mc}$.

We'll just prove (3) and (4).

Proof. For both parts, we assume that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$. Then we can write $a = b + mk$ and $c = d + m\ell$ for some $k, \ell \in \mathbb{Z}$.

(3) We check that

$$a + c = b + mk + d + m\ell = b + d + m(k + \ell),$$

so that $a + c \equiv b + d \pmod{m}$.

(4) We check that

$$ac = (b + mk)(d + m\ell) = bd + m(kd + b\ell + k\ell),$$

so that $ac \equiv bd \pmod{m}$. □

It turns out that evaluating polynomials at congruent numbers returns congruent numbers:

Theorem 2.1.2. [NZM91, Theorem 2.2] *Let $f(x) \in \mathbb{Z}[x]$ (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 $c_i \in \mathbb{Z}$. By Thm 2.1.(4), since $a \equiv \pmod{m}$, we find that for all $i \geq 0$, $a^i \equiv b^i \pmod{m}$, and thus $c_i a^i \equiv c_i b^i \pmod{m}$. Then by Thm 2.1.(2), adding the terms 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., $f(a) \equiv f(b) \pmod{m}$. □

The next theorem describes how canceling terms over a modulus works.

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) in particular, if $\gcd(a, m) = 1$, then $ax \equiv ay \pmod{m}$ iff $x \equiv y \pmod{m}$.

(3) Given $m_1, m_2 \ldots, m_r$, one has

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

for all $1 \leq i \leq r$ iff

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

Proof. For (1): suppose that $ax \equiv ay \pmod{m}$. Then $ax = ay + mk$ for some $k \in \mathbb{Z}$. Thus, $a(x - y) = mk$, and so

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

and thus (flipping it around)

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

However, $\frac{m}{\gcd(a,m)}$ and $\frac{a}{\gcd(a,m)}$ are coprime since $\gcd\left(\frac{m}{\gcd(a,m)}, \frac{a}{\gcd(a,m)}\right) = \frac{1}{\gcd(a,m)} \cdot \gcd(m, a) = 1$, and thus

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

Thus, $x \equiv y \pmod{\frac{m}{\gcd(a,m)}}$. I'll let you prove the converse direction.

For (2): This is a consequence of (1).

For (3): Assume that $x \equiv y \pmod{m_i}$ for each $1 \leq i \leq r$, one has $m_i \mid (x-y)$. Thus, $x-y$ is a common multiple of $m_1, m_2, \ldots, m_r$, and so $\operatorname{lcm}(m_1, m_2, \ldots, m_r) \mid (x-y)$, and thus $x \equiv y \pmod{\operatorname{lcm}(m_1, m_2, \ldots, m_r)}$. I'll let you prove the converse direction. □

In general, given a modulus m, for any integer $x \in \mathbb{Z}$, by the division algorithm we can write $x = mq + r$ for some $0 \leq r < m$. Thus, $x \equiv r \pmod{m}$. Therefore, any integer is congruent to exactly one of $0, 1, 2, \ldots, m - 1$ modulo m. This motivates the following definitions.

Definition 2.1.2. Given $x \in \mathbb{Z}$, if $x \equiv y \pmod{m}$, then y is a **residue** of x modulo m. A set $\{x_1, x_2, \ldots, x_m\}$ is called a **complete residue system modulo** m (CRS mod m) if for any integer $x \in \mathbb{Z}$, there exists a unique x_i in the set with $x \equiv x_i \pmod{m}$.

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 (prove it!).

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

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

is an arithmetic progression, called the **congruence class,** or **residue class, of** a **modulo** m**.**

When talking about congruence classes modulo m, we'll often refer to them by their representatives (e.g. the odd numbers, $S(1, 2)$, is just 1 modulo 2).

Each number in $S(a, m)$ is congruent to a modulo m. There are m distinct residue classes modulo m, defined by e.g. $S(0, m), S(1, m), S(2, m) \ldots, S(m-1, m)$: thus we 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 (RRS mod m) is a set of non-congruent integers $\{r_1, r_2, \ldots, r_n\}$ with that $\gcd(r_i, m) = 1$, such that any integer $x \in \mathbb{Z}$ coprime to m is congruent to exactly one such r_i.

The following theorem helps us construct reduced residue systems.

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

$$\gcd(a, m) = \gcd(b + mk, m) = \gcd(b + mk - (mk), m) = \gcd(b, m). \qquad \square$$

By this theorem, a reduced residue system can be created from a complete residue system by removing the representatives that are not coprime to m. As we'll see later, a RRS is important for studying multiplicative arithmetic modulo m.

Any two RRS's mod m will have an equal number of elements.

Definition 2.1.5. The number of elements in a RRS mod m is denoted $\phi(m)$, and is called the **Euler phi function,** or **Euler's totient function.**

Here's an alternative definition for $\phi(m)$.

Theorem 2.1.5. [NZM91, Theorem 2.5] *For each* $m > 0$*,* $\phi(m)$ *is the number of positive integers* $\leq m$ *that are coprime to* m*.*

Proof. This follows from constructing a RRS mod m with integers $0 \leq k < m$. $\square$

Euler's phi function appears in many contexts, and is interesting to study in its own right. We'll see more of it soon, and in future chapters.

Example 2.1.2.

1. A RRS mod 7 is $1, 2, 3, 4, 5, 6$; thus $\phi(7) = 6$.

2. A RRS mod 8 is $1, 3, 5, 7$; thus $\phi(8) = 4$.
3. $\phi(6) = 2$, since a RRS mod 6 is $1, 5$.
4. $\phi(12) = 4$, since a RRS mod 12 is $1, 5, 7, 11$.

You can modify a residue system simply by multiplying by a number coprime to m:

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

Example 2.1.3. A RRS mod 8 is $1, 3, 5, 7$; by this theorem, so is $3, 9, 15, 21$.

Proof. First, we note that $ar_i \equiv ar_j \pmod{m}$ implies $a_i \equiv a_j$, due to [NZM91, Theorem 2.3]. Therefore, the set $\{ar_1, ar_2, \ldots, ar_n\}$ has the same size as the complete/reduced residue system $\{r_1, r_2, \ldots, r_n\}$ and no two elements in it are congruent; therefore, it is also a complete/reduced residue system modulo m. To see this: the multiplication-by-a map $\{r_1, r_2, \ldots, r_n\} \to \{ar_1, ar_2, \ldots, ar_n\}$ is injective, by [NZM91, Theorem 2.3]. Then it is also surjective since both sets have the same size, and thus each r_i is congruent to ar_j for a unique $1 \leq j \leq n$. Thus, any integer $x \in \mathbb{Z}$ is congruent to r_i for some i, and thus to ar_j for some j. □

In practice, our residue systems will usually consist only of nonnegative integers which are at most m.

Fermat and Euler's theorems. The following theorem is a classical result due to Euler.

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

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

Proof. Let $r_1, r_2, \ldots, r_n$ be a RRS mod m. By the previous [NZM91, Theorem 2.6], so is $ar_1, ar_2, \ldots, ar_n$, and the proof showed that the sets $\{r_1, r_2, \ldots, r_n\}$ and $\{ar_1, ar_2, \ldots, ar_n\}$ are the same modulo m (possibly in a different order). Therefore, noting that the size of the two RRS'w mod m is $\phi(m)$, we have

$$\prod_{i=1}^{n} r_i \equiv \prod_{i=1}^{n} ar_i \pmod{m},$$

so that

$$\prod_{i=1}^{n} r_i \equiv a^{\phi(m)} \prod_{i=1}^{n} r_i \pmod{m}.$$

Since $\gcd\left(\prod_{i=1}^{n} r_i, m\right) = 1$, by [NZM91, Theorem 2.3.(2)] we can cancel out $\prod_{i=1}^{n} r_i$ from both sides to conclude that

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

□

Here's a special case of Euler's theorem, which came first. This takes $m = p$ to be prime.

Theorem 2.1.8 (Fermat's little theorem)**.** *For prime* $p \in \mathbb{Z}$, *if* $p \nmid a$ *then one has*

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

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

Example 2.1.4. By Euler's theorem:

1. Taking $m = 6$ and $a = 5$, $\phi(6) = 2$, and thus $5^2 \equiv 1 \pmod 6$ is true. Check it: $6 \mid (25 - 1) = 24$.
2. Taking $m = 12$ and $a = 7$, $\phi(12) = 4$, and $7^4 \equiv 1 \pmod{12}$ is true. Check it: $7^4 - 1 = 2400 = 12 \cdot 200$.
3. Taking $m = 11$ and $a = 2$, $\phi(11) = 10$, and $2^{10} \equiv 1 \pmod{11}$ is true. Check it: $2^{10} - 1 = 1023 = 11 \cdot 93$.

The multiplicative inverse of an integer isn't usually an integer. However, we can invert certain integers modulo m, and procure a "multiplicative inverse" mod m.

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. Since $\gcd(a, m) = 1$ 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 mod m gives $ax \equiv 1 \pmod m$.

For uniqueness: if $y \in \mathbb{Z}$ is such that $ay \equiv 1 \pmod m$, then multiplying this by x gives $(ax)y \equiv x \pmod m$, i.e., $y \equiv x \pmod m$. □

Definition 2.1.6. Given $a \in \mathbb{Z}$ 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 as $x := a^{-1}$ (where we abuse notation and allow a to represent both itself and its residue class mod m).

There are at least three ways to get the multiplicative inverse of a mod m:

1. Use the Euclidean/Blankinship's algorithm to express $1 = ax + my$; it follows that $a^{-1} \equiv x \pmod m$.
2. Deduce that $a^{-1} \equiv a^{\phi(m)-1} \pmod m$ from Euler's theorem.
3. If m is small, we can multiply a by each integer $1 \leq k < m$ until we have $ak \equiv 1 \pmod m$.

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

1. $2 \mod 3$: 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 $2^{-1} \equiv 1 \pmod 3$.
2. $14 \mod 5$: $14 \equiv -1 \pmod 5$, so it's clear that $(-1)^{-1} \equiv -1 \pmod 5$.
3. $2 \mod 37$: since 37 is odd, by Fermat's little theorem, we know that $2^{\phi(37)} \equiv 1 \pmod{37}$, so the inverse of $2 \mod 37$ is represented by the integer $2^{\phi(37)-1} = 2^{35}$.

The following theorem is a classical one which characterizes prime numbers.

Theorem 2.1.10 (Wilson's Theorem)**.** *For any integer* $p > 1$*,* p *is prime if and only if* $p \mid ((p-1)! + 1)$*, i.e.,* $(p-1)! \equiv -1 \pmod p$*.*

Before we prove this, we 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 $x \equiv \pm 1 \pmod{p}$.*

Equivalently, the lemma says that $x^2 - 1$ has roots $1, -1$ modulo p. We will study solutions to polynomials modulo prime p closely in this class.

Proof. The backward direction is clear. For the forward direction, if $x^2 \equiv 1 \pmod{p}$, then $p \mid (x^2-1) = (x+1)(x-1)$. By primality of p, 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, the backward direction: suppose we can write $pk = (p-1)! + 1$ for some $k \in \mathbb{Z}$. If p is composite, then $p = ab$ with $1 < a, b < p$. Thus, $a \leq p - 1$, and so $a \mid (p-1)!$. Since $a \mid p$, this forces $a \mid (pk - (p-1)!) = 1$, which is impossible since $a > 1$. We conclude that p is prime.

Forward direction: suppose that p is prime. Observe that $p \mid ((p-1)! + 1)$ is equivalent to $(p-1)! \equiv -1 \pmod{p}$, which we will show. 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 $kx_k \equiv 1 \pmod{p}$ (multiplicative inverse).

If $k = x_k$, then we have $k^2 \equiv 1 \pmod{p}$, so by the lemma $k \equiv \pm 1 \pmod{p}$. Thus, for $k \neq 1, p-1$, we have $k \neq x_k$. Therefore, each number $1 < k < p-1$ has a *distinct, unique* inverse number $1 < x_k < p-1$; we can assume that $k \leq \frac{p-1}{2}$ and $\frac{p-1}{2} < x_k$. With all of this in mind, we make the following calculations:

$$\begin{aligned}
(p-1)! &= \prod_{k=1}^{p-1} k \\
&= 1 \cdot (-1) \cdot \prod_{k=2}^{p-2} k \\
&= -1 \cdot \prod_{k=2}^{\frac{p-1}{2}} kx_k \\
&\equiv -1 \cdot \prod_{k=1}^{\frac{p-1}{2}} 1 \\
&= -1 \pmod{p}.
\end{aligned}$$

□

One application of Wilson's theorem is to study solutions to $x^2 + 1$ modulo p. (Note that over $\mathbb{C}$, the roots of $x^2 + 1$ are $\pm i$.)

Theorem 2.1.12. [NZM91, Theorem 2.12] *For prime p, the congruence $x^2 \equiv -1 \pmod{p}$ has solutions if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.*

Proof. For $p = 2$, 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.,

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

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

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

Pairing off k with $p-k$ above implies that

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

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

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

where $a := \prod_{k=1}^{\frac{p-1}{2}} k$. Thus, 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.

Conversely, suppose that $a^2 \equiv -1 \pmod p$ for some $a \in \mathbb{Z}$. 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 $a^{p-1} \equiv 1 \pmod p$, so that

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

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 $k \in \mathbb{Z}$, so that $p = 1+4k$, i.e., $p \equiv 1 \pmod 4$. □

Example 2.1.6.

1. $5 \equiv 1 \pmod 4$, and we can check that x^2+1 modulo 5 has solutions $x = \pm 2 = 2, 3$, i.e., $2^2, 3^2 \equiv -1 \pmod 5$.
2. $7 \equiv 3 \pmod 4$, so x^2+1 is *irreducible* modulo 7. We can directly check that the squares mod 7 are $1^2 = 1, 2^2 = 4, 3^2 = 9 \equiv 2, 4^2 \equiv (-3)^2 = 3^2 = 9$, etc., none of which are -1 mod 7.
3. $13 \equiv 1 \pmod 4$, so x^2+1 has a root modulo 13. What is it?

The following theorem connects solutions for x^2+1 modulo p, to representing p as a sum of two squares.

Theorem 2.1.13. [NZM91, Lemma 2.13] *For an odd prime p, $p \equiv 1 \pmod 4$ 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 backward direction: suppose $p = a^2 + b^2$. Since any square satisfies $x^2 \equiv 0, 1 \pmod 4$, we find that $p \equiv 0, 1, 2 \pmod 4$. Since p is odd, $p \not\equiv 0, 2 \pmod 4$, which forces $p \equiv 1 \pmod 4$.

For the forward direction, assume that $p \equiv 1 \pmod 4$. Then by [NZM91, Theorem 2.12], there exists $c \in \mathbb{Z}$ with $c^2 \equiv -1 \pmod p$.

Since p is prime, $\sqrt{p}$ is not an integer. Let us set $K := \lfloor \sqrt{p} \rfloor$, the largest integer less than $\sqrt{p}$. Then we have $K < \sqrt{p} < K + 1$. Consider pairs (u, v) where $0 \leq u, v \leq K$. u and v each take on $K + 1$ possible values, thus there are $(K + 1)^2$ pairs. Since $K + 1 > \sqrt{p}$, there are $> p$ such pairs.

Define a function $f \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by $f(u, v) := u + cv$. Modulo p, $f(u, v)$ can only take p distinct values; however, there are (u, v) pairs above, so by the pigeonhole principle, two distinct pairs (u_1, v_1) and (u_2, v_2) have the same value: $f(u_1, v_1) \equiv f(u_2, v_2) \pmod p$, i.e., $u_1 + cv_1 \equiv u_2 + cv_2 \pmod p$. Thus,

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

Squaring both sides gives

$$(u_1 - u_2)^2 \equiv c^2(v_1 - v_2)^2 \equiv -(v_1 - v_2)^2 \pmod p.$$

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 $p \mid (a^2 + b^2)$.

We claim this is an equality. First, note that $a^2 + b^2 > 0$ since $a \neq 0$ or $b \neq 0$ (the pairs were distinct). Next, recall that $0 \leq u_i, v_i \leq K$, so that $|u_1 - u_2| = |a|, |v_1 - v_2| = |b| \leq K$. Since $K < \sqrt{p}$, we square both sides and get $a^2, b^2 < p$, and thus $0 < a^2 + b^2 < 2p$. However, $p \mid (a^2 + b^2)$, and the only multiple of p strictly between 0 and $2p$ is p. We conclude that $p = a^2 + b^2$. □

Therefore, we have shown that for any odd prime p,

$$p \equiv 1 \pmod 4 \Leftrightarrow p = a^2 + b^2 \Leftrightarrow x^2 \equiv -1 \pmod p \text{ has a solution.}$$

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, since 7, 11, 19 and 23 are 3 modulo 4, these five numbers are not sums of two squares – this can be checked easily by hand.

Exercise 2.1.1 (Residual practice)**.** In this exercise, you will review some modular arithmetic with hands-on examples.

a) List all integers $1 \leq n \leq 100$ which are congruent to 1 mod 18.
b) Give a complete residue system modulo 13 comprised of multiples of 4.
c) Give a reduced residue system modulo 16. What is $\phi(16)$?
d) Give a reduced residue system modulo 11. For each representative of this residue system, write down its multiplicative inverse modulo 11.

Exercise 2.1.2 (Digit arithmetic)**.** 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 a_0 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 $a_0 = 0$ or 5.

(*Hint:* for each of these parts, work with the base 10 expansion of k: i.e., 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 2.1.3 (Polynomials under modular arithmetic)**.**

a) Let $f(x) \in \mathbb{Z}[x]$ be a polynomial with integer coefficients. Show that if $f(a) \equiv n \pmod{m}$, then for all $k \in \mathbb{Z}$ one has $f(a + km) \equiv n \pmod{m}$.
b) Let $f(x), g(x) \in \mathbb{Z}[x]$ be polynomials of degree $n \geq 0$. Writing

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

and

$$g(x) = \sum_{i=0}^{n} b_i x^i,$$

show that if $a_i \equiv b_i \pmod{m}$ for each $0 \leq i \leq m$, then for all $c \in \mathbb{Z}$ one has $f(c) \equiv g(c) \pmod{m}$.

Exercise 2.1.4 (Squares can't span)**.** Show that if $m \geq 3$, then $\{0^2, 1^2, \ldots, (m-1)^2\}$ is not a complete residue system modulo m.

Exercise 2.1.5 (Roots of $x^{\phi(m)} - 1$ modulo m)**.**

a) Show that for any modulus $m > 0$, one has that $a \in \mathbb{Z}$ is a root of $x^{\phi(m)} - 1$ modulo m if and only if $\gcd(a, m) = 1$. Thus, any reduced residue system modulo m is the set of all roots of $x^{\phi(m)} - 1$ modulo m.
b) Show that for prime $p \in \mathbb{Z}$, one has for all $a \in \mathbb{Z}$ that

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

(This strengthens Exercise 2.1.11.)

Exercise 2.1.6 (Power digits)**.**

a) Prove that the square of an integer has $0, 1, 4, 5, 6$ or 9 for its unit digit.
b) Prove that the fourth power of an integer has $0, 1, 5$ or 6 for its unit digit.
c) Without using a calculator, prove that $(123456789)^5$ has unit digit 9.

Exercise 2.1.7 (Hidden powers)**.**

a) Show that for all integers $n, k \in \mathbb{Z}$, if $7 \nmid n$ then $7 \mid (n^{6k} - 1)$.
b) Show that for any integer $n \in \mathbb{Z}$, one has $42 \mid (n^7 - n)$.

Exercise 2.1.8 (Inverting reduced residue systems)**.** Show that if $\{x_1, x_2, \ldots, x_r\}$ is a reduced residue system modulo m, then so is $\{x_1^{-1}, x_2^{-1}, \ldots, x_r^{-1}\}$.

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

a) Show that if g is a primitive root modulo m, then for all integers n coprime to m, there exists a unique integer $0 \leq e \leq \phi(m) - 1$ such that $g^e \equiv n \pmod{m}$. In particular, a primitive root modulo m, if it exists, "generates" all residue classes modulo m which are coprime to m.
b) Determine with proof whether a primitive root exists modulo the following integers.
   i) Modulo 6;
   ii) Modulo 8;
   iii) Modulo 9.
c) Use Exercise 2.1.8 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.

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

Exercise 2.1.11 (Breaking a golden rule)**.** As you already know, for any real numbers x and y and for any integer $n > 0$, one usually has $(x + y)^n \neq x^n + y^n$. However, this expectation changes when considering integers modulo a prime p.

Show that for any integers $a, b \in \mathbb{Z}$, one has $(a + b)^p \equiv a^p + b^p \pmod{p}$.

Exercise 2.1.12 (Totients for prime powers)**.** Show that for prime powers p^e, one has $\phi(p^e) = p^e - p^{e-1}$.

Exercise 2.1.13 (Modular arithmetic in the exponents)**.**

a) Show that for $e, f, m \in \mathbb{Z}$ with $m > 0$, if $e \equiv f \pmod{\phi(m)}$, then for all integers a coprime to m, one has $a^e \equiv a^f \pmod{m}$.
b) Show that modulo 13, one has for all integers a with $13 \nmid a$ that

$$a^{16} + 42a^{12} + 11a^4 + 1 \equiv 4 - a^4 \pmod{13}.$$

c) Show that for any polynomial $f(x) \in \mathbb{Z}[x]$, there exists a polynomial $g(x) \in \mathbb{Z}[x]$ of degree $< p$ such that for all $a \in \mathbb{Z}$ with $p \nmid a$, one has

$$f(a) \equiv g(a) \pmod{p}.$$

Bonus Exercise 2.1.14 (Primes and the number $n!+1$)**.** This problem explores primes and their connection to numbers of the form $n! + 1$ for $n \in \mathbb{Z}^+$. 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 $n \in \mathbb{Z}^+$ 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: https://oeis.org/A002981.

Bonus Exercise 2.1.15 (Polynomials and prime values)**.** Prove that no polynomial $f(x) \in \mathbb{Z}[x]$ of degree > 1 has the property that $f(n)$ is prime for all $n \in \mathbb{Z}^+$. See also the Bunyakovsky conjecture (Bonus Exercise 1.3.11).

Bonus Exercise 2.1.16 (Splitting behavior in the Gaussian integer ring)**.** This is a continuation of Bonus Exercise 1.3.12. In [NZM91, Theorem 2.12] and [NZM91, Lemma 2.13], it was shown that for prime $p > 2$,

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

In this exercise, we will study how prime numbers $p \in \mathbb{Z}$ 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.

Theorem. *A prime $p \in \mathbb{Z}^+$ 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 $p > 2$ 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 a prime $p > 2$ is irreducible in $\mathbb{Z}[i]$ if and only if $p \equiv 3 \pmod{4}$, if and only if $x^2 + 1$ is irreducible modulo p.
d) How does $p = 2$ factorize in $\mathbb{Z}[i]$? How does $x^2 + 1$ factor modulo 2?

This exercise shows that for any prime $p \in \mathbb{Z}^+$, its behavior in $\mathbb{Z}[i]$ is determined by the factorization of $x^2 + 1$ modulo p. In a more general setting, this is a consequence of a theorem of Dedekind and Kummer.

2.2. **Solutions of Congruences.** This section will elucidate some of the technicalities in looking for solutions to polynomials modulo m.

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

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

By an earlier theorem ([NZM91, Theorem 2.2]), we know that if $f(a) \equiv 0 \pmod{m}$ and $b \equiv a \pmod{m}$, then $f(b) \equiv 0 \pmod{m}$. Therefore, solutions modulo m are *independent of representatives for congruence classes.* Thus, when asking about solutions modulo m, we don't distinguish between a and b if $a \equiv b \pmod{m}$.

Example 2.2.1. $f(x) := x^2 + 1$ has two solutions modulo 5: they are $x \equiv \pm 2 \mod 5$. Since $2 \equiv 7 \pmod{5}$, we know that $f(2) \equiv f(7) \equiv 0 \pmod{5}$. We count $2 \mod 5$ and $7 \mod 5$ as the same solution.

Some polynomials can "degenerate" when reducing modulo p. For example, for integer $m > 0$, the polynomial $f(x) := mx$ is identically zero modulo m: i.e., for all $a \in \mathbb{Z}$ we have $f(a) \equiv 0 \pmod{m}$.

This leads to the following definition.

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

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

with $c_n \neq 0$; thus, the degree of f is n.

Given $m > 0$, the **degree of f modulo** m is the greatest positive i such that $c_i \not\equiv 0 \pmod{m}$. Note that $i \leq n$.

This section will end with the simplest case for analyzing solutions modulo m: linear polynomials of the form $f(x) := ax - b$.

Theorem 2.2.1. [NZM91, Theorem 2.17] *Fix integers a and b, and $m > 0$. Then the polynomial $ax - b$ has roots modulo m, i.e., the congruence*

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

has a solution modulo m, if and only if $\gcd(a, m) \mid b$. If this happens, then there are g distinct solutions modulo m, given by $c = A \cdot \frac{b}{g} + \frac{m}{g} \cdot k$, where $0 \leq k < g$ and A is a representative for the inverse of $\frac{a}{g}$ modulo $\frac{m}{g}$.

Something to note is that when $\gcd(a, m) = 1$, the theorem says there is exactly one solution modulo m, and it's given by $c = a^{-1}b \pmod{m}$ – this is already known from taking the multiplicative inverse! So this theorem shines when $\gcd(a, m) > 1$.

Proof. Let $g := \gcd(a, m)$. Then writing $a = a'g$ and $m = m'g$, we have

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

If this has a solution, then there exists $c \in \mathbb{Z}$ with $a'gc = b + m'g$, which forces $g \mid b$. On the other hand, if $g \mid b$ (say $b'g = b$) then cancellation gives

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

Since $\gcd(a', m') = 1$, it follows that a' has a multiplicative inverse modulo m', say $(a')^{-1}$; it follows that $c := (a')^{-1}b' \mod m'$ is a solution. This proves the first part.

As shown above, taking any $c \in \mathbb{Z}$ with $c \equiv (a')^{-1}b' \mod m'$ gives a solution to $ax \equiv b \pmod{m}$ once we multiply everything in $a'c \equiv b' \pmod{m'}$ by g. However, $c \equiv (a')^{-1}b' \pmod{m'}$ is the same as $c = (a')^{-1}b' + m'k$ for some $k \in \mathbb{Z}$ (here, $(a')^{-1}$ is an integer which represents the inverse of a' modulo m'). Any integer $k \in \mathbb{Z}$ gives the desired congruence; however, there are g distinct solutions modulo m when choosing $0 \leq k < g$; for if $k \geq g$, then writing $k = gq + r$ with $0 \leq r < g$, one has $c = (a')^{-1}b' + m'(gq + r) = (a')^{-1}b' + m'r + mq$, which is congruent to $(a')^{-1}b' + m'r$ modulo m. □

Throughout this course, we'll study solutions to polynomials modulo p and over $\mathbb{Q}$.

Remark. Studying solutions to $f(x) \in \mathbb{Z}[x]$ modulo various m can give information about solutions to $f(x)$ in $\mathbb{Z}$ (or even $\mathbb{Q}$). For example, if $f(x)$ has a solution $a \in \mathbb{Z}$, then $f(a) = 0$, thus $f(a) \equiv 0 \pmod{p}$ for all primes $p \in \mathbb{Z}^+$. Therefore, if there exists $p \in \mathbb{Z}^+$ such that f has no solutions modulo p, then f has no solutions in $\mathbb{Z}$!

Exercise 2.2.1 (Solutions to linear congruences)**.** Find all integer solutions to each of the following congruences. If no solution exists, then explain why.

a) $20x \equiv 4 \pmod{30}$;
b) $353x \equiv 254 \pmod{400}$;
c) $64x \equiv 83 \pmod{105}$.

Exercise 2.2.2 (Roots of $x^2 - 1$ mod prime powers)**.**

a) Show that for an odd prime power p^e, the equation $x^2 \equiv 1 \pmod{p^e}$ has two solutions mod p^e, namely ± 1.
b) Show that $x^2 \equiv 1 \pmod{2}$ has one solution, and that $x^2 \equiv 1 \pmod{4}$ has two solutions.
c) Show that for $e \geq 3$, $x^2 \equiv 1 \pmod{2^e}$ has four solutions. (*Hint:* recall that for any integer $a \in \mathbb{Z}^+$, one has $\gcd(a + 1, a - 1) = 2$).

Bonus Exercise 2.2.3 (Covering systems)**.** A **covering system** is a collection of arithmetic progressions $\{S(a_i, b_i)\}_{i \in I}$ whose union covers $\mathbb{Z}$:

$$\bigcup_{i \in I} S(a_i, b_i) = \mathbb{Z}.$$

For example, a complete residue system modulo m induces a covering system via since $S(0, m) \cup S(1, m) \cup \ldots \cup S(m - 1, m) = \mathbb{Z}$.

a) Show that the following congruences form a covering system:

$$\begin{aligned} x &\equiv 0 \pmod{2}; \\ x &\equiv 0 \pmod{3}; \\ x &\equiv 1 \pmod{4}; \\ x &\equiv 1 \pmod{6}; \\ x &\equiv 11 \pmod{12}. \end{aligned}$$

b) The following is an open problem on the parity of the moduli from a covering system.

Conjecture 2.2.2 (Odd covering problem)**.** *Every covering system has at least one odd modulus.*

2.3. **The Chinese Remainder Theorem.** This section focuses on the the following problem: given $a_1, a_2, \ldots, a_r, m_1, m_2, \ldots, m_r \in \mathbb{Z}$, is there a solution $x \in \mathbb{Z}$ to the following system of congruences:

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

If one imposes extra conditions on the moduli m_i, then the answer is yes. This is called the Chinese remainder theorem (CRT).

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

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

Furthermore, fixing a solution $x_0 \in \mathbb{Z}$, an integer y is also a solution if and only if $y \equiv x_0 \pmod{m_i}$ for each $1 \leq i \leq r$, i.e., $y \equiv x_0 \pmod{m_1 m_2 \ldots m_r}$.

Proof. We will directly construct a solution. First, let us set $m := m_1 m_2 \cdots m_r$. Observe that for each $1 \leq i \leq r$, one has $\gcd(\frac{m}{m_i}, m_i) = 1$. Thus, $\frac{m}{m_i}$ is invertible modulo m_i, so there exists $b_i \in \mathbb{Z}$ with

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

We also note that for $j \neq i$, we have $m_j \mid m$, so that $\frac{m}{m_i} \equiv 0 \pmod{m_j}$.

Let us define

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

Then for each $1 \leq i \leq r$, we have

$$x_0 = \sum_{i=1}^{r} \frac{m}{m_i} b_i a_i \equiv \left(\frac{m}{m_i} b_i \right) a_i \equiv a_i \pmod{m_i}.$$

Thus, x_0 is the desired solution.

For the last part, observe that if y is also a solution, then by definition, $y \equiv x_0 \pmod{m_i}$ for each $1 \leq i \leq r$, which is equivalent to $y \equiv x_0 \pmod{m_1 m_2 \ldots m_r}$ by [NZM91, Theorem 2.3.(3)]. □

We can use the proof to construct solutions when they exist.

Example 2.3.1. Find the least positive integer x_0 that is a solution to the system

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

Note that $5, 7, 13$ are pairwise coprime, so we can apply the proof of the CRT. To do this, we first need to find b_i satisfying the following congruences:

$$\begin{aligned} 11 \cdot 13 \cdot b_1 &\equiv 1 \pmod 7, \\ 7 \cdot 13 \cdot b_2 &\equiv 1 \pmod{11}, \\ 7 \cdot 11 \cdot b_3 &\equiv 1 \pmod{13}. \end{aligned}$$

We check that $11 \cdot 13 b_1 \equiv 24 b_1 \equiv 3 b_1 \pmod 7$, which means we will take $b_1 := 3^{-1} \pmod 7$; let $b_1 := 5$. Similarly, $7 \cdot 13 \cdot b_2 \equiv 1 \pmod{11}$ means we need $b_2 \equiv 3^{-1} \pmod{11}$, so set $b_2 := 4$. Finally, $7 \cdot 11 b_3 \equiv 1 \pmod{13}$ means $b_3 \equiv (-1)^{-1} \pmod{13}$, so take $b_3 := -1$.

Then the integer

$$x_0 := 143 \cdot 5 \cdot 5 + 91 \cdot 4 \cdot 7 + 77 \cdot (-1) \cdot 3 = 5892$$

is a solution to the system of congruences (check it!).

To get the *least positive solution,* simply divide $m := 7 \cdot 11 \cdot 13 = 1001$ into $x_0 = 5892$ using the division algorithm:

$$x_0 = m \cdot q + 887.$$

Then $887 \equiv x_0 \pmod m$, and thus 887 is the smallest positive solution.

The condition that the m_i are pairwise coprime is necessary for a solution to exist, as our next example illustrates.

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

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

Since 4 divides both 52 and 72, we can reduce both equations modulo 4 and get

$$\begin{aligned} x &\equiv 29 \equiv 1 \pmod 4, \\ x &\equiv 19 \equiv 3 \pmod 4. \end{aligned}$$

Such a solution then implies $1 \equiv 3 \pmod 4$, so that $4 \mid (3-1) = 2$, which is impossible.

However, it is possible to have a solution even when the moduli aren't pairwise coprime!

Example 2.3.3. Consider the system

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

The CRT doesn't immediately apply since $\gcd(10, 84) = 2 > 1$. However, we can break these up into more congruences based on prime factorizations. $10 = 2 \cdot 5$, and so $x \equiv 3 \pmod{10}$ implies $x \equiv 3 \equiv 1 \pmod{2}$ and $x \equiv 3 \pmod{5}$. $84 = 3 \cdot 4 \cdot 7$, and so $x \equiv 5 \pmod{84}$ implies $x \equiv 5 \equiv 2 \pmod{3}$, $x \equiv 5 \equiv 1 \pmod{4}$ and $x \equiv 5 \pmod{7}$. Since $\gcd(10, 84) = 2$, we should check that the new congruences involving 2 are compatible, and they are: $x \equiv 1 \pmod{4}$ implies $x \equiv 1 \pmod{2}$. Therefore, the new congruences to consider are:

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

Since $4, 3, 5$ and 7 are pairwise coprime, a solution can be constructed from the CRT using $m = 4 \cdot 3 \cdot 5 \cdot 7 = 420$ and solving $105b_1 \equiv 1 \pmod{4}$, $140b_2 \equiv 1 \pmod{3}$, $84b_3 \equiv 1 \pmod{5}$ and $60b_4 \equiv 1 \pmod{7}$. We can take $b_1 = 1$, $b_2 = 2$, $b_3 = -1$ and $b_4 = 2$, and produce the solution $x_0 = 105 \cdot 1 \cdot 1 + 140 \cdot 2 \cdot 2 + 84 \cdot (-1) \cdot 3 + 60 \cdot 2 \cdot 5 = 1013$.

The CRT sets up a way to "factorize" reduced residue systems modulo m based its prime power factors. Given $m > 0$, we let $R(m)$ be the "canonical" RRS modulo m.

Theorem 2.3.2. [NZM91, Theorem 2.19] *Given* ***pairwise coprime integers*** $m_1, m_2, \ldots, m_r > 0$*, define* $m := m_1 m_2 \cdots m_r$*. Then we have a bijection*

$$R(m) \to R(m_1) \times R(m_2) \times \ldots \times R(m_r)$$

via the natural map $a \mapsto (a \mod m_1, a \mod m_2, \ldots, a \mod m_r)$*, where* $a_i \mod m_i$ *denotes the remainder of* a *divided by* m_i*. In particular, we have*

$$\phi(m) = \phi(m_1)\phi(m_2) \cdots \phi(m_r).$$

Therefore, given an integer $n \in \mathbb{Z}$ *with factorization* $n = \prod_{i=1}^{r} p_i^{e_i}$*, one has*

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

Proof. Let $\varphi \colon R(m) \to R(m_1) \times R(m_2) \times \ldots \times R(m_r)$ denote the map above.

1. φ is injective: if $a, b \in \mathbb{Z}$ satisfy

$$(a \mod m_1, a \mod m_2, \ldots, a \mod m_r) = (b \mod m_1, b \mod m_2, \ldots, b \mod m_r),$$

then $a \equiv b \pmod{m_i}$ for each $1 \leq i \leq r$, and thus by [NZM91, Theorem 2.3.(3)], we have

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

i.e.,

$$a \equiv b \pmod{m_1 m_2 \cdots m_r},$$

i.e.,

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

which forces $a = b$ since $0 \leq a, b < m$.

2. φ is surjective: by the CRT! Given a tuple $(a_1 \mod m_1, a_2 \mod m_2, \ldots, a_3 \mod m_r)$ where $a_i \in \mathbb{Z}$, by the CRT there exists $a \in \mathbb{Z}$ with $a \equiv a_i \pmod{m_i}$ for each i. Thus, $\varphi(a) = (a_1 \mod m_1, a_2 \mod m_2, \ldots, a_3 \mod m_r)$ (we can assume $0 \leq a < m$).

Therefore, this bijection implies an equivalence of sizes:

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

so that

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

If $n \in \mathbb{Z}^+$ has factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, then the $p_i^{e_i}$ are pairwise coprime, so the above implies that

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

We've just shown that if $m, n \in \mathbb{Z}$ are coprime, then $\phi(mn) = \phi(m)\phi(n)$; we say that ϕ is a *multiplicative function* (more on this in Chapter 4).

In HW 2 Exercise 8 (see Exercise 2.1.12), we have shown that for prime power p^e, one has $\phi(p^e) = p^e - p^{e-1} = p^{e-1}(p-1)$. Therefore, we have a more precise formula for $\phi(n)$: writing $n = \prod_{i=1}^{r} \phi(p_i^{e_i})$ with the p_i pairwise coprime, one has

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

Example 2.3.4. Using the formulas for $\phi(n)$, we check that:

i) $\phi(18) = \phi(2 \cdot 3^2) = \phi(2)\phi(3^2) = 1 \cdot 3(3-1) = 6$;
ii) $\phi(100) = \phi(2^2 \cdot 5^2) = \phi(2^2)\phi(5^2) = 2(2-1) \cdot 5(5-1) = 40$;
iii) $\phi(210) = \phi(2 \cdot 3 \cdot 5 \cdot 7) = \phi(2)\phi(3)\phi(5)\phi(7) = 1 \cdot 2 \cdot 4 \cdot 6 = 48$.

Remark. The same natural map is also bijection $C(m) \to \prod_{i=1}^{r} C(m_i)$. We already knew that $|C(m)| = \prod_{i=1}^{r} |C(m_i)|$, though.

Solutions to polynomials modulo m. The CRT is also useful for studying solutions/roots/zeroes to polynomials modulo m.

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

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

Let us also count the number of solutions via $\phi_f(m) := \#V(f, m)$.

An analogy between $\phi_f(m)$ and the Euler phi function $\phi(m)$ is made below.

Theorem 2.3.3. [NZM91, Theorem 2.20] *Given a polynomial $f(x) \in \mathbb{Z}[x]$ and pairwise coprime $m_1, m_2, \ldots, m_r \in \mathbb{Z}^+$, define $m := m_1 m_2 \cdots m_r$. Then we have a bijection*

$$V(f, m) \to V(f, m_1) \times V(f, m_2) \times \ldots \times V(f, m_r)$$

via the natural map $a \mapsto (a \mod m_1, a \mod m_2, \ldots, a \mod m_r)$. *In particular, we have*

$$\phi_f(m) = \phi_f(m_1)\phi_f(m_2)\cdots\phi_f(m_r).$$

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

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

The key idea for proving this bijection is that if a_i is a root of $f(x)$ modulo m_i, then any solution to the simultaneous congruences $x \equiv a_i \pmod{m_i}$ via the CRT is a root of $f(x)$ modulo $m_1 m_2 \ldots m_r$.

Proof. This proof will be similar to the one for factorizing complete residue systems. In fact, the map in this theorem is a restriction of the map

$$\varphi\colon C(m) \to C(m_1) \times C(m_2) \times \ldots \times C(m_r)$$

to $V(f,m) \subseteq C(m)$ (where we take $C(m)$ to be the "canonical" CRS modulo m). If $f(a) \equiv 0 \pmod{m}$, then for each $1 \leq i \leq r$, we have $f(a) \equiv 0 \pmod{m_i}$, and so φ takes $V(f,m)$ into $V(f,m_1) \times V(f,m_2) \times \ldots V(f,m_r)$.

1. φ is injective *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 $(a_1 \mod m_1, a_2 \mod m_2, \ldots, a_3 \mod m_r)$ where for each $1 \leq i \leq r$, $a_i \in V(f,m_i)$; thus, $f(a_i) \equiv 0 \pmod{m_i}$.

   By the CRT there exists $a \in C(m)$ with $a \equiv a_i \pmod{m_i}$ for each i. Thus, since $a \equiv a_i \pmod{m_i}$, by [NZM91, Theorem 2.2] we have $f(a) \equiv f(a_i) \pmod{m_i}$, so that $f(a) \equiv 0 \pmod{m_i}$. Therefore, by [NZM91, Theorem 2.3.(3)], we have

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

   i.e.,

   $$f(a) \equiv 0 \pmod{m_1 m_2 \cdots m_r},$$

   i.e.,

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

   Therefore, our solution a from the CRT lies inside $V(f,m)$, and so the map $\varphi\colon V(f,m) \to V(f,m_1) \times V(f,m_2) \times \ldots \times V(f,m_r)$ is surjective.

We conclude from this that

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

whence we have

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

Therefore, given an integer $n \in \mathbb{Z}^+$ with factorization $n = \prod_{i=1}^{r} p_i^{e_i}$, one has

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

□

Remark. Unlike Euler's totient function $\phi(m)$, it is possible for $\phi_f(m)$ to be zero!

The above theorem that finding roots of a polynomial $f(x) \in \mathbb{Z}[x]$ modulo m amounts to finding roots modulo prime-power divisors of m, and then applying the CRT (which preserves solutions). 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'd like to determine all of its roots modulo 21. To do this, we factorize $21 = 3 \cdot 7$ and looks for roots of f modulo 3 and 7; for each of these solutions, we create pairs to apply the CRT towards.

Since 3 and 7 are small, we check directly that $f(x) \equiv 0 \pmod 3$ iff $x \equiv 1 \pmod 3$, and $f(x) \equiv 0 \pmod 7$ iff $x \equiv 2, 4 \pmod 7$. Thus, we have two cases for solutions via the CRT:

1. $x \equiv 1 \pmod 3$, $x \equiv 2 \pmod 7$. Then solving $7b_1 \equiv 1 \pmod 3$ and $3b_2 \equiv 1 \pmod 7$, we can take $b_1 := 1$ and $b_2 := 5$, to construct $x_0 := 7 \cdot 1 \cdot 1 + 3 \cdot 5 \cdot 2 = 37 \equiv 16 \pmod 2$1. We can check directly that
$$f(16) = 16^2 + 16 + 1 = 271 = 13 \cdot 21,$$
and so $f(16) \equiv 0 \pmod{21}$.
2. $x \equiv 1 \pmod 3$ and $x \equiv 4 \pmod 7$. The same work from before applies (since only the a_i changed) and we can take $b_1 := 1$ and $b_2 := 5$, and thus have solution $x_0 := 7 \cdot 1 \cdot 1 + 3 \cdot 5 \cdot 4 = 67 \equiv 4 \pmod{21}$. We double-check that
$$f(4) = 4^2 + 4 + 1 = 21,$$
and so $f(4) \equiv 0 \pmod{21}$.

Therefore, the two solutions to $f(x)$ modulo 21 are the congruence classes of 4 and 16. In particular, $f(x) \equiv (x - 4)(x - 4^2) \pmod{21}$.

Exercise 2.3.1 (Calculations with the CRT)**.** For each part, determine all integers x which satisfy each of the simultaneous congruences. If no such solution exists, then prove it.

a) $x \equiv 2 \pmod 3$, $x \equiv 3 \pmod 5$ and $x \equiv 5 \pmod 2$;
b) $x \equiv 1 \pmod 4$, $x \equiv 0 \pmod 3$ and $3x \equiv 1 \pmod 7$;
c) $5x \equiv 1 \pmod 6$, $4x \equiv 13 \pmod{15}$.

Exercise 2.3.2 (Primes in multiple arithmetic progressions)**.** In this exercise, you will prove the following "multilinear version" of Dirichlet's theorem on primes in arithmetic progressions. See Exercise 1.3.6 (or Exercise 4 from HW 2) for a statement of the original result, which you should use to prove this one.

Theorem. *Given pairwise coprime integers $m_1, m_2, \ldots, m_r \in \mathbb{Z}^+$, for any integers $a_1, a_2, \ldots, a_r \in \mathbb{Z}$ 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*
$$p \equiv a_i \pmod{m_i}.$$

Exercise 2.3.3 (Totient arithmetic)**.** This exercise will study some arithmetic properties of Euler's totient function $\phi(n)$.

a) Determine all integers $n \in \mathbb{Z}^+$ for which $\phi(n)$ is odd.

b) Show that if every prime p which divides m also divides n, then $\phi(mn) = m\phi(n)$.
c) Prove that if $\phi(mn) = \phi(n)$ and $m > 1$, then $m = 2$ and n is odd. Characterize the set of positive integers satisfying $\phi(2n) = \phi(n)$.

Exercise 2.3.4 (The totient as a divisor)**.** Determine all integers $n \in \mathbb{Z}^+$ for which $\phi(n) \mid n$.

Exercise 2.3.5 (Fibers of the totient function are finite)**.** Show that for a fixed integer $n \in \mathbb{Z}^+$, the equation $\phi(x) = n$ has a finite number of solutions.

Bonus Exercise 2.3.6 (Beyond the totient function's range)**.** This exercise explores which positive integers are not in the image of Euler's totient function $\phi(x)$.

a) Show that for odd $n \in \mathbb{Z}^+$, the equation $\phi(x) = n$ has a solution if and only if $n = 1$. Thus, the image $\phi(\mathbb{Z}^+)$ has no odd numbers except $n = 1$.
b) Show that there does not exist a solution to $\phi(x) = 14$.
c) Show that 14 is the *smallest* positive even integer not in $\phi(\mathbb{Z}^+)$. Then determine the next smallest such integer.

2.4. **Prime Power Moduli (Hensel's Lemma).** We've seen previously that studying polynomials modulo m can give information back about m. For example, if $p > 2$ is prime, then $x^2 + 1$ has a root modulo p if and only if p is a sum of squares. We'll continue to study roots of polynomials modulo m.

From [NZM91, Theorem 2.20], we say that determining solutions to a given $f \in \mathbb{Z}[x]$ modulo m was equivalent to determining solutions modulo each prime power p^e dividing m. This section studies how finding a solution modulo p determines solutions modulo each higher power p^k. This technique is called *Hensel's lemma.*

Theorem 2.6.1. [NZM91, Theorem 2.23] *Let $f(x)$ be a polynomial with integer coefficients. If a is a root of $f(x)$ modulo p^k, i.e.,*

$$f(a) \equiv 0 \pmod{p^k},$$

and *$f'(a) \not\equiv 0 \pmod{p}$**,** then there exists an integer $t \in \mathbb{Z}$, **unique modulo** p**,** for which*

$$f(a + tp^k) \equiv 0 \pmod{p^{k+1}}$$

(so $a + tp^k$ is a root of $f(x)$ modulo p^{k+1}).

Such a solution modulo p^{k+1} in Hensel's lemma is called a *lift of $a \mod p^k$*.

In practice, if we find a root a modulo p and have $f'(a) \equiv 0 \pmod{p}$, then we can apply Hensel's lemma repeatedly until we have our desired solution modulo p^e where $p^e \parallel m$.

Before proving this, let us state a new lemma from Section 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 $f(a) \equiv 0 \pmod{p^k}$, we want to lift this to a a solution $f(b) \equiv 0 \pmod{p^{k+1}}$.

We'll make an observation first. Recall from Calculus II that the Taylor series of $f(x)$ at $x = a$ is

$$f(x) = \sum_{k\geq 0} \frac{f^{(k)}(x)(x-a)^k}{k!} = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2} + \dots$$
$$= f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2} + \ldots + \frac{f^{(n)}(a)(x-a)^n}{n!},$$

where n is the degree of $f(x)$.

Our potential solution $b \in \mathbb{Z}$ will have $b \equiv a \pmod{p^k}$, and so takes the form $b = a + tp^k$ for some $t \in \mathbb{Z}$. To this end, plug in an integer $x = a + tp^k$ to get (noting $x - a = tp^k$)

$$f(a + tp^k) = f(a) + f'(a) \cdot tp^k + \frac{f''(a) \cdot t^2 p^{2k}}{2} + \ldots + \frac{f^{(n)}(a) \cdot t^n p^{nk}}{n!}.$$

Modding out by p^{k+1}, we get

$$f(a + tp^k) \equiv f(a) + f'(a) \cdot tp^k \pmod{p^{k+1}}.$$

Let's justify this: this is because for each $j \geq 2$, the term $\frac{f^{(j)}(a) \cdot t^j p^{jk}}{j!} = t^i p^{jk} \cdot \frac{f^{(j)}(a)}{j!}$ is divisible by p^k; this is not obvious since the fraction $\frac{f^{(j)}(a)}{j!}$ might not be an integer, and may have a power of p in its denominator. This follows from the fact that $\frac{f^{(j)}(a)}{j!}$ is an integer: if $c_r x^r$ is a monomial term in the sum $f(x)$, then the corresponding term in $f^{(j)}(a)$ (taking derivative j times) is

$$c \cdot r(r-1)(r-2)\cdots(r-(j-1)) \cdot a^{r-j}.$$

Since $r, r-1, r-2, \ldots, r-(j-1)$ are consecutive integers, the lemma implies that $j!$ divides this term. Therefore, it divides all terms in $f^{(j)}(a)$, thus $j! \mid f^{(j)}(a)$. In particular, $p^{k+1} \mid \frac{f^{(j)}(a) \cdot t^i p^{jk}}{j!}$ for $k \geq 2$.

So we have for any $t \in \mathbb{Z}$ that

$$f(a + tp^k) \equiv f(a) + f'(a) \cdot tp^k \pmod{p^{k+1}}.$$

We want this to be zero modulo p^{k+1}, so we want

$$f(a) + f'(a) \cdot tp^k \equiv 0 \pmod{p^{k+1}}$$

We are assuming that $f(a) \equiv 0 \pmod{p^k}$, so $p^k \mid f(a)$ and we can divide by p^k and get

$$f'(a) \cdot t \equiv -\frac{f(a)}{p^k} \pmod{p}.$$

Since we are assuming $f'(a) \not\equiv 0 \pmod{p}$, we can invert it and get

$$(*) \qquad t \equiv -(f'(a))^{-1} \cdot \frac{f(a)}{p^k} \pmod{p}.$$

Choosing a $t \in \mathbb{Z}$ which satisfies this congruence, we conclude that $a + tp^k$ is a solution to $f(x)$ modulo p^{k+1}. $\square$

Definition 2.6.1. In general, given $a, b \in \mathbb{Z}$ and $k \leq j$, if $f(a) \equiv 0 \pmod{p^k}$ and $f(b) \equiv 0 \pmod{p^j}$ and $b \equiv a \pmod{p^k}$, we say that b is a **lift** of a, or lies above a.

If a is a root of $f(x)$ modulo p^k and $f'(a) \not\equiv 0 \pmod{p}$, we say that a is a *nonsingular root modulo p^k;* otherwise, we say a is *singular.*

If a is a nonsingular root of $f(x)$ modulo p, then by Hensel's lemma, a lifts to a root a_2 modulo p^2; since $a_2 \equiv a \pmod{p}$, a_2 is also nonsingular. Apply Hensel's lemma again to get a nonsingular root a_3 modulo p^3, and so on. We obtain a recursive sequence from these solutions $a = a_1, a_2, a_3, \ldots$ via the proof of Hensel's lemma. To see this, note first that for each $k \geq 1$, we have that

$$a_{k+1} = a_k + t_k p^k,$$

where

$$t_k \equiv -A_k \cdot \frac{f(a_k)}{p^k} \pmod{p},$$

i.e.,

$$t_k p^k \equiv -A_k \cdot f(a_k) \pmod{p^{k+1}},$$

where $A_k := (f'(a_k))^{-1} \in \mathbb{Z}$ represents a multiplicative inverse modulo p. Thus, modulo p^{k+1} we have

$$(*) \qquad a_{k+1} \equiv a_k - A_k \cdot f(a_k) \pmod{p^{k+1}}.$$

Fun fact: this is analogous to Newton's method for finding roots of a real differentiable function.

Example 2.6.1. We'll solve $f(x) \equiv 0 \pmod{7^3}$ where $f(x) := x^2 + x + 47$. We'll try and lift a solution modulo 7: observe that $x^2 + x + 47 \equiv 0 \pmod{7}$ has two solution: they are $a = 1$ and $a = 5$, since $1^2 + 1 + 47 = 49 \equiv 0 \pmod{7}$ and $5^2 + 5 + 47 = 77 \equiv 0 \pmod{7}$. Furthermore, $f'(x) = 2x + 1$, and we see that both $f'(1) = 3 \not\equiv 0 \pmod{7}$ and $f'(5) = 11 \not\equiv 0 \pmod{7}$, whence it follows that 1 and 5 are nonsingular roots modulo 7.

We will lift each solution individually:

1. We can lift $a = 1$ to a solution a_2 modulo 7^2 via equation $(*)$. Our lift modulo 49 is

$$\begin{aligned} a_2 &\equiv a_1 - A_1 \cdot f(a_1) \\ &= 1 - A_1 \cdot f(1) \pmod{49}, \end{aligned}$$

where $A_1 := f'(1)^{-1}$ is the multiplicative inverse computed modulo 7. Since $f'(1) = 3$, we check that $3^{-1} \mod 7 = 5 \mod 7$. Thus, since $f(1) = 49$, we deduce that

$$a_2 \equiv 1 - 5 \cdot 0 = 1 \pmod{49}.$$

Next, we can lift $a_2 = 1$ to a solution $a_3 \bmod 7^3 = 343$ via $(*)$:

$$a_3 \equiv a_2 - A_2 \cdot f(a_2) \pmod{343},$$

i.e.,

$$a_3 \equiv 1 - A_2 \cdot f(1) \pmod{343},$$

where $A_2 := (f'(1))^{-1}$ is the multiplicative inverse modulo 7; so we already know that $A_2 = 5$ from the previous step. Since $f(1) = 49$, we deduce that

$$\begin{aligned} a_3 &\equiv 1 - 5 \cdot 49 \\ &= 1 - 245 \\ &= -244 \\ &\equiv 99 \pmod{343}. \end{aligned}$$

Therefore, $a = 99$ is a solution to x^2+x+47 modulo 343. You can plug in $a = 343$ into $x^2 + x + 47$ to double-check that it's a multiple of 343: $(343)^2 + 343 + 47 = 343 \cdot 29$.

2. We lift $a = 5 \mod 7$ to a solution a_2 modulo 49 via $(*)$:

$$a_2 \equiv 5 - A_1 \cdot f(5) \pmod{49}$$

where $A_1 := (f'(5))^{-1} \mod 7$. We compute that modulo 7, $(f'(5))^{-1} = (11)^{-1} \equiv (4)^{-1} \equiv 2 \pmod 7$, as well as $f(5) = 77 \pmod{49}$. Therefore, we have

$$\begin{aligned} a_2 &\equiv 5 - 2 \cdot 77 \\ &= -149 \\ &= -2 \pmod{49}. \end{aligned}$$

Next, we can lift $a_2 = -2$ to a solution a_3 modulo 343, via

$$a_3 = -2 - A_2 \cdot f(-2) \pmod{343}$$

where $A_2 := (f'(-2))^{-1}$ is the inverse modulo 7. Since $(f'(-2))^{-1} \equiv (f'(5))^{-1} = 2 \pmod 7$ and $f(-2) = 49$, we have

$$\begin{aligned} a_3 &\equiv -2 - 2 \cdot 49 \\ &= -100 \\ &\equiv 243 \pmod{343}. \end{aligned}$$

Thus, $a = 243$ is a solution to $x^2 + x + 47$ modulo 343; we can check that $243^2 + 243 + 47 = 343 \cdot 173$.

Therefore, our two solutions to $x^2 + x + 47$ modulo 7^3 are represented by $a = 99, 243$.

A final note: Hensel's lemma requires that a root a of $f(x)$ modulo p^k be **nonsingular** (i.e., $f'(a) \not\equiv 0 \pmod p$). However, when a is a **singular** root, it is possible for there to exist 0 or p lifts of a to a root modulo p^{k+1}. We won't worry about that for this course (the exercises cover this a little bit).

Exercise 2.6.1 (Studying Hensel's lemma)**.** Fix a polynomial $f(x) \in \mathbb{Z}[x]$ and prime $p \in \mathbb{Z}^+$, and suppose that $f(x)$ has a nonsingular root $a \in \mathbb{Z}$ modulo p. By Hensel's lemma, we can inductively lift the solution a to obtain solutions a_k modulo p^k, and these solutions are described by the formula

$$a_{k+1} \equiv a_k - A_k \cdot f(a_k) \pmod{p^{k+1}},$$

where $A_k \in \mathbb{Z}$ represents $(f'(a_k))^{-1} \mod p$.

a) Using the formula above, show directly that a_{k+1} is a lift of a_k.

b) Show that the integers A_k can be chosen independently of $k \geq 1$.
c) Suppose that a is instead a *singular root* modulo p, i.e., $f'(a) \equiv 0 \pmod{p}$. Prove that there are either 0 or p lifts of a modulo p^2.

Exercise 2.6.2 (Hensel's lemma practice)**.** Solve the congruence $x^4 + 2 \equiv 0 \pmod{27}$. Then check directly that your solution(s) works by writing it as a multiple of 27.

Exercise 2.6.3 (Solutions at the composite level)**.** With proof, solve the equation $x^2 + 5x + 24 \equiv 0 \pmod{36}$. Then check that your solution(s) works by writing it as a multiple of 36.

Bonus Exercise 2.6.4 (Programming Hensel's Lemma)**.** Create a program which does the following: given a polynomial $f(x) \in \mathbb{Z}[x]$, a prime power p^e and an integer a such that $f(a) \equiv 0 \pmod{p^e}$, it uses Hensel's Lemma to check whether a lifts to a root of $f(x)$ modulo p^{e+1}. The code also returns whether the root a is singular or not, and if it is singular, it returns the p lifts of a mod p^{e+1}.

You could also write code that does the following: given $f(x) \in \mathbb{Z}[x]$ and a prime power p^e, it determines all roots of $f(x)$ mod p^e.

2.7. **Prime modulus.** By the CRT and Hensel's lemma, the problem of finding solutions to $f(x) \equiv 0 \pmod{m}$ is reduced to finding solutions $f(x) \equiv 0 \pmod{p}$ for $p \mid m$. Unfortunately, there's no general method for finding roots modulo p.

Polynomial roots modulo m can have counterintuitive properties. For example, by the fundamental theorem of algebra, any degree n polynomial $f(x) \in \mathbb{Z}[x]$ has exactly n solutions in $\mathbb{C}$, and thus *at most* n solutions in $\mathbb{Z}$. However, modulo composite m, it's possible for a polynomial $f(x) \in \mathbb{Z}[x]$ to have more roots than its degree. For example, $f(x) := 5x^2 + 10$ has roots $0, 1, 2, 3$ and 4 modulo 5 since $f(x)$ is zero modulo 5. As another example, $g(x) := x^2 + 7x + 2$ has four solutions modulo 10: $x = 3, 4, 8, 9$.

However, like over $\mathbb{C}$ or $\mathbb{Z}$, A polynomial $f(x) \in \mathbb{Z}[x]$ cannot have a reduction modulo p with roots than its degree modulo p.

Theorem 2.7.1. [NZM91, Theorem 2.26] *Given prime $p \in \mathbb{Z}^+$ and polynomial $f(x) \in \mathbb{Z}[x]$, if $\deg(f) = n$ then $f(x)$ has at most n roots modulo p.*

Therefore, after reducing $f(x)$ modulo p, the resulting polynomial $\overline{f} := f(x) \mod p$ will have the "correct" amount of roots modulo p. In more technical language, this is because the set of congruence classes modulo p forms a *field.*

For the rest of this chapter, we will develop some basic *abstract algebra* to recontextualize our what we've learned about congruences.

2.8. **Number theory from an algebraic viewpoint.** Abstract algebra is the study of algebraic structures – this includes groups, rings, modules, fields, ...

Much of elementary number theory can be recontextualized in terms of abstract algebra. For example, a complete residue system modulo m is a **group.**

What is a group? Here's some terminology before we start. Given a set X, a binary operation on X is a map $X \times X \to X$. It can be "addition", or "multiplication", or something more abstract.

Definition 2.10.1. A set G is a **group** if there exists a binary operation $\oplus\colon G\times G\to G$ such that G and $\oplus$ have the following properties:

1. by definition, G is **closed** under $\oplus$, i.e., $\oplus$ takes G into itself;
2. $\oplus$ is **associative:** for $g,h,k\in G$, one has
$$(g\oplus h)\oplus k = g\oplus(h\oplus k);$$
3. G has an **identity** element $e := e_G\in G$: for all $g\in G$,
$$e_G\oplus g = g\oplus e_G = g;$$
4. G is closed under **inverses:** for all $g\in G$, there exists $h\in G$ with
$$g\oplus h = h\oplus g = e.$$
Note: h is often written as g^{-1}, and e_G is often written as 1.

Such an operation $\oplus\colon G\times G\to G$ is called a **group law on G.**

If G has the property such that for all $g,h\in G$ one has $g\oplus h = h\oplus g$, then G is called an **abelian** or **commutative** group. Finally, if G is a finite set, we say that G is a **finite group.** We call the size of G its **order,** and often write it as $\#G$ or $|G|$.

Remark. Given a group G, do we write g^{-1} or $-g$ for inverses? By default, we write g^{-1} and the group structure multiplicatively. However, there are some exceptions, such as groups associated with the integers under addition.

When specifying the operation on a fixed group G, we'll sometimes write $(G,\oplus)$ instead of G to emphasize the group operation.

Example 2.10.1. The integers $\mathbb{Z}$ is an abelian group under addition: so take $a\oplus b := a+b$ (the usual addition). However, $\mathbb{Z}$ is not a group under multiplication, since it's not closed under multiplicative inverses.

Definition 2.10.2. Given a group $(G,\oplus)$ and subset $H\subseteq G$, we say that H is a **subgroup** of G if $e\in H$ and H is a group under $\oplus$.

An important class of examples of finite groups is that of the integers modulo m. Given $a \mod m$ and $b \mod m$, we know that $a+b \mod m$ is also a congruence class. Thus, we have an group law on any complete residue system modulo m. To be more precise, consider the "canonical" CRS mod m, namely
$$\{0,1,2,\ldots,m-1\}.$$
Then given any two numbers a,b in this set, the sum $a+b$ will also lie in this set *once we reduce it modulo m.* Thus, this CRS mod m is closed "up to congruence modulo m."

Definition 2.10.3. Given an integer $m>0$, we define the **integers modulo** m as
$$\mathbb{Z}/m\mathbb{Z} := \mathbb{Z}_m := \{[a] : 0\le a<m\},$$
where $[a] := \overline{a} := a \mod m$ is the congruence class of a modulo m. We add elements of $\mathbb{Z}/m\mathbb{Z}$ via normal addition, and then reducing mod m: that is, $[a]\oplus[b] := [a+b \mod m]$. The identity element is $e := [0]$.

As usual, we will abuse notation and sometimes write a instead of $[a]$ or $a \mod m$ – context should be clear!

Example 2.10.2. Consider $m = 6$. Then $\mathbb{Z}/6\mathbb{Z}$ is

$$\mathbb{Z}/6\mathbb{Z} = \{[0], [1], [2], [3], [4], [5]\};$$

note that $0, 1, 2, 3, 4, 5$ forms a complete residue system modulo 6, so $\mathbb{Z}/6\mathbb{Z}$ has exactly 6 elements. We have e.g. $[3] \oplus [4] := [7] = [1]$ (this just says that $3+4 = 7 \equiv 1 \pmod 6$), and $-[a] = [-a] = [6 - a]$ ($a - a \equiv 0 \pmod 6$, and $-a \equiv 6 - a \pmod 6$).

The next definition tells us the "correct" way to interpret one group inside another.

Definition 2.10.4. Given two groups $(G, \oplus)$ and $(G', \odot)$, we say that a map $\varphi \colon G \to G'$ is a **homomorphism** if it preserves the group structure: that is, for all $g, h \in G$, one has

$$\varphi(g \oplus h) = \varphi(g) \odot \varphi(h).$$

We say that φ is an **isomorphism** if it is also a bijection. The inverse map of an isomorphism is also a homomorphism.

Remark. From the definitions, one can deduce that:

1. For all $g \in G$, one has $\varphi(g^{-1}) = \varphi(g)^{-1}$;
2. φ preserves identity elements, i.e., $\varphi(e_G) = e_{G'}$.

Example 2.10.3. Given $m > 0$, we have a natural map

$$\varphi \colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$$

defined by

$$\varphi(a) := a \mod m.$$

One checks that

$$\varphi(a + b) = (a + b) \mod m = [a \mod m] + [b \mod m] = \varphi(a) + \varphi(b).$$

This map is called **reduction modulo** m**.**

It's important to note when two groups are the same.

Definition 2.10.5. Given two groups $(G, \oplus)$ and $(G', \odot)$, we say that G and G' are **isomorphic** if there exists a bijective homomorphism $\Phi \colon G \to G'$. We write $G \cong G'$.

Example 2.10.4. Observe that $\mathbb{Z}^\times := \{1, -1\} \subseteq \mathbb{Z}$ is a group under the usual multiplication. We have an isomorphism

$$(\mathbb{Z}/2\mathbb{Z}, +) \to (\mathbb{Z}^\times, \cdot)$$

via $[0] \mapsto 1$ and $[1] \mapsto -1$ (check it!). So additive and multiplicative groups can be isomorphic.

Example 2.10.5. Consider a nonstandard CRS mod 6, such as

$$X := \{6, 1, 8, 3, 16, 11\}$$

Then X is a group under addition, from the fact that every integer $a \in \mathbb{Z}$ is represented by a number in X. For example, $11 \oplus 3 := 8$ since $11 - 3 \equiv 8 \pmod 6$. This group is isomorphic to $\mathbb{Z}/6\mathbb{Z}$ via

$$\Phi \colon X \to \mathbb{Z}/6\mathbb{Z}, a \mapsto [a \mod 6].$$

This generalizes to any CRS mod m:

Theorem 2.10.1. [NZM91, Theorem 2.46] *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 modulo m.*

Just as $\mathbb{Z}/m\mathbb{Z}$ represents complete residue systems modulo m as additive groups, we can represent *reduced* residue systems as multiplicative groups.

Definition 2.10.6. Given $m > 0$, we define the **unit group modulo** m as

$$(\mathbb{Z}/m\mathbb{Z})^\times := \mathbb{Z}_m^\times := \{[a] \in \mathbb{Z}/m\mathbb{Z} : \gcd(a, m) = 1\}.$$

Its group operation is multiplication modulo m: $[a] \cdot [b] := [ab \mod m]$. Recall that $|(\mathbb{Z}/m\mathbb{Z})^\times| = \phi(m)$.

In comparison to [NZM91, Theorem 2.46], it turns out that all reduced residue systems modulo m are also groups under multiplication and are isomorphic to one another.

Theorem 2.10.2. [NZM91, Theorem 2.47] *For integer $m > 0$, any reduced residue system modulo m is a group under multiplication mod m, and is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^\times$.*

Remark. For an arbitrary group $(G, \oplus)$, we'll often write the group law *multiplicatively:* i.e., for $g \in G$, we write e.g. $g \oplus g \oplus g =: g \cdot g \cdot g =: ggg =: g^3$. We'll often write inverses as g^{-1}. Etc.

If $(G, \oplus)$ is commutative/abelian, we'll sometimes write $+$ instead of $\oplus$, and e.g. $g \oplus g \oplus g =: g + g + g =: 3g$, inverses as $-g$, etc.

Exercise 2.10.1 (Isomorphic Additive and Multiplicative Groups)**.** Prove that the groups $(\mathbb{Z}/4\mathbb{Z}, +)$ and $((\mathbb{Z}/5\mathbb{Z})^\times, \cdot)$ are isomorphic via an explicit bijective homomorphism, describing where each element goes. (*Hint:* 2 is a primitive root modulo 5. See HW 3, Exercise 7 for a definition, or §2.8.)

Exercise 2.10.2 (Comparing $+$ and $\times$)**.**

a. Prove that the groups $(\mathbb{Z}/m\mathbb{Z}, +)$ and $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ cannot be isomorphic for $n > 1$. isomorphic for $m > 1$.
b. Characterize all group homomorphisms $(\mathbb{Z}/m\mathbb{Z}, +) \to ((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$.

Bonus Exercise 2.10.3 (Groups of matrices)**.** 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 matrices in $\mathrm{Mat}_{n\times n}(\mathbb{R})$ which are invertible. Check 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 $n = 1$.

2.11. **Groups, rings and fields.** This section is a direct continuation of Section 2.10. However, we'll study not just groups, but also **rings** and **fields,** which have two operations on them.

First, some more group theory. Here's a theorem concerning finiteness in groups.

Theorem 2.11.1. [NZM91, Theorem 2.48]

1. *If G is a group, then **cancellation** always holds: if $g, h, k \in G$ and $gh = gk$, then $h = k$.*
2. *If G is **finite,** then for any element $g \in G$, there exists a least positive integer $r \in \mathbb{Z}^+$ for which $g^r = e$ (recall e is the identity element).*
3. *If the order of g is r, then for all integers k, one has that $g^k = e \Rightarrow r \mid e$.*

Proof.

1. If
$$gh = gk$$
then
$$(g^{-1})gh = (g^{-1})gk,$$
which by associativity is
$$(g^{-1}g)h = (gg^{-1})k,$$
i.e.,
$$eh = ek,$$
i.e.,
$$h = k.$$
2. Assume G is finite, and let $g \in G$. Consider the subset
$$\{e, g, g^2, g^3, \ldots\} \subseteq G.$$
Since G is finite, so is the set $\{e, g, g^2, g^3, \ldots\}$, so for some $s, t \geq 0$ with $t > s$, we must have $g^t = g^s$. Rewritten as $g^{t-s}g^s = g^s$, by the previous part we can cancel g^s and deduce that $g^{t-s} = e$. Then by the well-ordering principle, there will exist a minimal such $r \in \mathbb{Z}^+$.
3. Prove it! This one is Exercise 4 in HW 5 (see also Exercise 2.11.2). □

This introduces the following important definition.

Definition 2.11.1. Given a group G (finite *or* infinite), 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 this holds is called the **order of g.** It is often written as $|g|$.

If $g^n \neq e$ for all $n \in \mathbb{Z}^+$, then g is said to have **infinite order.**

Elements from finite groups always have finite order.

Definition 2.11.2. A group G is said to be **cyclic** if there exists an element $g \in G$ whose powers generate the whole group:
$$\{\ldots, g^{-2}, g^{-1}, g^0 = e, g, g^2, \ldots, \} = G.$$
Such an element g is called a **generator for the group G.** We often write $G = \langle g \rangle$.

Proposition 2.11.2. *If G is finite, then $g \in G$ is a generator if and only if $|g| = \#G$.*

Proof. Prove it! □

Example 2.11.1. The (additive) group $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic of size m. The element $[1] \in \mathbb{Z}/m\mathbb{Z}$ has order m, since it's the least positive integer for which $m[1] := [m] = [0]$ (to see this: if $k \in \mathbb{Z}$ is such that $k[1] = [0]$, then this is equivalent to $k \equiv 0 \pmod{m}$, thus $m \mid k$). Thus, $\langle [1] \rangle = (\mathbb{Z}/m\mathbb{Z}, +)$.

Question: what other generators of $(\mathbb{Z}/m\mathbb{Z}, +)$ are there?

Here's an important result comparing the order of an element to its (finite) group's size. This is a special case of **Lagrange's theorem** (you can call it that in this class.)

Theorem 2.11.3. [NZM91, Theorem 2.49] *If G is a finite group, then the order of any element $g \in G$ divides $|G|$. Therefore, writing $n := |G|$, one has for all $g \in G$ one has*

$$g^n = 1.$$

Proof. (Sketch only in class?) Let $g \in G$ have order r; thus, the set $A := \{e, g, g^2, \ldots, g^{r-1}\}$ are r distinct elements in G. If these are all the elements of G, then we're done: $|g| = |G|$. If not, then there exists $h_1 \in G$ with $h_1 \neq g^k$ for all $k \in \mathbb{Z}$. If the set

$$A_1 := A \cup h_1 A := \{1, g, g^2, \ldots, g^{r-1}, h_1, h_1 g, h_1 g^2, \ldots, h_1 g^{r-1}\}$$

has $2r$ distinct elements. To check this: first, if $h_1 g^i = h_1 g^j$ for some $0 \leq i, j < r$, then cancellation implies $g^i = g^j$, which forces $i = j$; thus, $h_1 A$ has r distinct element. Then, if for $0 \leq i, j < r$ we have $g^i = h_1 g^j$, then $i \neq j$ implies $h_1 = g^{|i-j|} \in A$, a contradiction; thus, $A \cap h_1 A = \emptyset$. So the union is as large as possible, with $2r$ elements.

If $A_1 = G$, then we're done. If not, then we can continue this process until we've constructed a set

$$A_k := A \cup h_1 A \cup h_2 A \cup \ldots \cup h_k A$$

of kr elements which spans G: i.e., $A_k = G$, so that $kr = |G|$, thus $r := |g| \mid |G|$.

The second part follows from the following: if $r = |g|$, then writing $r \mid n \Rightarrow n = rk$, one has

$$g^n = (g^r)^k = e^k = e.$$ □

What's nice about [NZM91, Theorem 2.49] is that it provides an alternative proof to some earlier results. For example, it proves Euler's theorem, and thus Fermat's little theorem: if $\gcd(a, m) = 1$, then we have the congruence class $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$. Thus, the multiplicative order of $[a]$ divides the order of $(\mathbb{Z}/m\mathbb{Z})^\times$, the latter of which is $\phi(m)$. Therefore, $[a]^{\phi(m)} = [1]$; in terms of congruences, this is the same as

$$a^{\phi(m)} \equiv 1 \pmod{m},$$

which is Euler's theorem. Additionally, this shows that $[a]^{-1} = [a^{\phi(m)-1}]$.

Example 2.11.2. Consider the *additive* group $\mathbb{Z}/6\mathbb{Z} = \{[0], [1], [2], [3], [4], [5]\}$. One checks by hand the **additive orders**

$$|[0]| = 1, |[1]| = 6, |[2]| = 3, |[3]| = 2, |[4]| = 3, |[5]| = 6.$$

Each of these orders divide $|(\mathbb{Z}/6\mathbb{Z}, +)| = 6$. Is there a pattern to these orders? On the other hand, the multiplicative group $((\mathbb{Z}/6\mathbb{Z})^\times, \cdot) = \{[1], [5]\}$, and the **multiplicative order** of $[5]$ is 2. This checks out, as $|((\mathbb{Z}/6\mathbb{Z})^\times, \cdot)| = 2$.

Ring theory. Rings are a step above groups: they are much more analogous to $\mathbb{Z}$, where $\mathbb{Z}$ is not just a group under one operation, but has two operations: addition and multiplication, and they interact with each other.

Definition 2.11.3. A **ring** is a set R with the following properties.

1. R has two operations $\oplus$ and $\odot$, such that R is a commutative group under $\oplus$.
2. R has an identity $0 := 0_R$ under $\oplus$, and an identity $1 := 1_R$ under $\odot$.
3. R isn't necessarily a group under $\odot$, but it's closed under $\odot$ and $\odot$ is associative.
4. (distributive property) $\odot$ 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.,

$$r \cdot (s + t) = r \cdot s + r \cdot t.$$

Remark. One has for all $r \in R$ that

$$0 \cdot r = r \cdot 0 = 0.$$

(This will be proven in the exercises.)

Just as before, we'll often write $+$ instead of $\oplus$ and $\cdot$ instead of $\odot$.

If R satisfies $r \cdot s = s \cdot r$ for all $r, s \in R$, then we say that R is **commutative.**

If R is commutative and multiplication $\odot$ on the subset $R^\bullet$ of nonzero elements is closed under multiplicative inverses (i.e., R is also a group under $\odot$), then R is called a **field.**

Example 2.11.3. The set $\mathbb{Z}$ of integers is a commutative ring under the usual addition and multiplication. It is *not* a field since e.g. $2 \in \mathbb{Z}$ doesn't have a multiplicative inverse.

For each integer $m > 0$, "being congruent modulo m" is an equivalence relation on $\mathbb{Z}$. We can define the set $\mathbb{Z}/m\mathbb{Z}$ of equivalence classes modulo m, written as $[a]$. Then $\mathbb{Z}/m\mathbb{Z}$ is a ring under addition and multiplication of $\mathbb{Z}$ modulo m: $[a] \oplus [b] := [a + b]$, and $[a] \cdot [b] := [ab]$. This is called the *ring of integers modulo* m. This definition agrees with what we've already had for $\mathbb{Z}/m\mathbb{Z}$ – Exercise 1 on HW 5 explores this!

The subset $R^\times$ of elements in R with multiplicative inverses is called the **unit group of** R**.** While R might not be a group under $\odot$, $R^\times$ will be. This notation applies to $(\mathbb{Z}/m\mathbb{Z})^\times$, the unit group of the ring $\mathbb{Z}/m\mathbb{Z}$.

The following theorem provides a large class of examples of rings that we've seen before.

Theorem 2.11.4. [NZM91, Theorem 2.50] *For $m > 0$, the set $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m is a ring under its addition and multiplication modulo m. One has that $\mathbb{Z}/m\mathbb{Z}$ is a field if and only if m is prime.*

Proof. We'll check the second statement. If $m = p$ is prime, then all nonzero elements $[a] \in \mathbb{Z}/p\mathbb{Z}$ satisfy $\gcd(p, a) = 1$, thus $[a]$ has a multiplicative inverse. Therefore, $\mathbb{Z}/p\mathbb{Z}$ is a field.

Conversely, if $\mathbb{Z}/m\mathbb{Z}$ is a field, then for any $[a] \in \mathbb{Z}/m\mathbb{Z}$ there must exist $[b] \in \mathbb{Z}/m\mathbb{Z}$ with $[a][b] = [1]$, i.e., $ab \equiv 1 \pmod{m}$. Such a congruence implies $\gcd(a, m) = 1$, so m is coprime to all $1 \leq a < m$, which forces m to be prime. □

Definition 2.11.4. Given two rings R and S, a **ring homomorphism** from R to S is a map $\varphi\colon R \to S$ such that:

1. φ is an additive group homomorphism: for $r_1, r_2 \in R$, one has $\varphi(r_1 \oplus_R r_2) = \varphi(r_1) \oplus_S \varphi(r_2)$, as well as $\varphi(-r_1) = -\varphi(r_1)$ and $\varphi(0_R) = 0_S$.
2. φ respects multiplication: $\varphi(r_1 \odot_R r_2) = \varphi(r_1) \odot_S \varphi(r_2)$.
3. $\varphi(1_R) = 1_S$.

A bijective ring homomorphism is a **ring isomorphism.**

Example 2.11.4. For any $m > 0$, we have a natural ring homomorphism $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ via reduction mod m: $a \mapsto [a] \mod m$.

Direct products of groups and rings. Given two groups $(G, \oplus_G)$ and $(H, \oplus_H)$, we can define the **direct product group of** G **and** H as the Cartesian product

$$G \times H := \{(g, h) : g \in G, h \in H\}$$

with the group operation

$$(g_1, h_1) \cdot (g_2, h_2) := (g_1 \oplus_G g_2, h_1 \oplus_H h_2).$$

We can inductively apply this to construct products of n groups, for $n \geq 1$: these look like

$$G_1 \times G_2 \times \ldots \times G_n,$$

whose elements are n-tuples and the group law is coordinate-wise. Similar to the above, one can define a direct product of two rings R and S.

With direct products, we can "factorize" the ring $\mathbb{Z}/m\mathbb{Z}$ based on the factorization of m, à la the CRT.

Theorem 2.11.5 (The CRT on $\mathbb{Z}/m\mathbb{Z}$)**.** *Let $m \in \mathbb{Z}^+$ with a factorization $m = \prod_{i=1}^{r} p_i^{e_i}$. Then under the natural map, one has a ring isomorphism*

$$\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/p_1^{e_1} \times \mathbb{Z}/p_2^{e_2} \times \ldots \times \mathbb{Z}/p_r^{e_r}.$$

With this natural map, one also has 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

$$((\mathbb{Z}/m\mathbb{Z})^\times, \cdot) \cong ((\mathbb{Z}/p_1^{e_1})^\times, \cdot) \times (\mathbb{Z}/p_2^{e_2})^\times, \cdot) \times \ldots \times (\mathbb{Z}/p_r^{e_r})^\times, \cdot).$$

Proof. We will omit the proof of this theorem; however, the key ingredient is the fact in each case, the Chinese remainder theorem shows surjectivity. □

Example 2.11.5. One has by the above that the rings $\mathbb{Z}/60\mathbb{Z}$ and $(\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/5\mathbb{Z})$ are isomorphic, with

$$(\mathbb{Z}/60\mathbb{Z}, +) \cong (\mathbb{Z}/4\mathbb{Z}, +) \times (\mathbb{Z}/3\mathbb{Z}, +) \times (\mathbb{Z}/5\mathbb{Z}, +)$$

and

$$(\mathbb{Z}/60\mathbb{Z})^\times \cong (\mathbb{Z}/4\mathbb{Z})^\times \times (\mathbb{Z}/3\mathbb{Z})^\times \times (\mathbb{Z}/5\mathbb{Z})^\times.$$

Exercise 2.11.1 (An equivalent definition of $\mathbb{Z}/m\mathbb{Z}$)**.** This exercise will explore a "residue system-free" definition of the ring of integers modulo m. Consequently, this alternative definition also applies to the additive group $(\mathbb{Z}/m\mathbb{Z}, +)$, as well as the unit group $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$.

Fix an integer $m > 0$. Let us define a relation on $\mathbb{Z}$ as follows: say $a \sim b$ if $a \equiv b \pmod{m}$.

a) Show that $\sim$ is an *equivalence relation:* i.e., show it it reflexive, symmetric and transitive.
b) Given an integer $a \in \mathbb{Z}$, what is the equivalence class $[a]$ of a under $\sim$, explicitly?
c) Using the ring operations $+$ and $\cdot$ from $\mathbb{Z}$, show that the set $\mathbb{Z}/m\mathbb{Z}$ of equivalence classes is also a ring. How many elements does it have?
d) Show that with this definition, $\mathbb{Z}/m\mathbb{Z}$ is isomorphic as a ring to the complete residue system $C(m) := \{0, 1, 2, \ldots, m-1\}$ given in class.

Exercise 2.11.2 (Some useful group-theoretic results)**.** This exercise collects some useful facts about orders of group elements. In the following, we let G denote a group.

a) Let $g \in G$. Prove that if for $k \in \mathbb{Z}$ 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 $\phi(m)$ generators.
d) Assume that the unit group $(\mathbb{Z}/m\mathbb{Z})^\times$ is cyclic. Prove that it has $\phi(\phi(m))$ generators.

Exercise 2.11.3 (The order of a product of elements)**.**

a) Show that for any abelian group G, if $a, b \in G$ have finite order, then so does ab, and the order of ab satisfies

$$|ab| \mid \operatorname{lcm}(|a|, |b|).$$

b) Show that if $|a|$ and $|b|$ are coprime, then $|ab| = |a| \cdot |b|$.

Exercise 2.11.4 (Orders under homomorphisms)**.** In this exercise, let G and H be finite groups, and $\varphi\colon G \to H$ a homomorphism.

a) Given an element $g \in G$, show the order divisibility $|\varphi(g)| \mid |g|$.

b) As it turns out, if φ is surjective, then for all $h \in H$ one has $|h| \mid |G|$. Give an example of a nontrivial surjective group homomorphism $\varphi\colon G \to H$ where, for some $g \in G$, one has $|\varphi(g)| < |g|$.
c) Show that if φ is injective, then $|\varphi(g)| = |g|$. In particular, injective homomorphisms and isomorphisms preserve orders.

Exercise 2.11.5 (Identities of rings)**.** In the following, let R denote a ring, with its two operations written as $+$ and $\cdot$.

a) Let $0 := 0_R$ denote the additive identity of R. Show that for all $r \in R$, one has $r \cdot 0 = 0 \cdot r = 0$.
b) Let $1 := 1_R$ denote the multiplicative identity of R. Prove there exists exactly one ring homomorphism $\iota\colon \mathbb{Z} \to R$.
c) Continuing part b), show that ι is injective if and only if 1_R has infinite additive order.
d) Continuing part c), show that if ι is not injective, then *assuming that R is an integral domain,* the additive order of 1_R is prime. (For the definition of an integral domain, see Exercise 2.11.6.)

When ι is injective, we say that R has *characteristic zero.* When ι is not injective, we say that R has *positive characteristic p,* where p is the additive order of 1_R.

Exercise 2.11.6 (Integral domains)**.** Let R be a commutative ring. Say that an element $r \in R$ is a *zero divisor* if for some nonzero $s \in R$ we have $rs = 0$. We say that R is an *integral domain* if R has no nontrivial 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 $s = t$.
b) Show that a field is automatically 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 2.11.7 (A sum of reciprocals of cubes)**.** 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 $\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$.)

Bonus Exercise 2.11.8. Characterize the $m \in \mathbb{Z}^+$ for which $\mathbb{Z}/m\mathbb{Z}$ is an integral domain.

Bonus Exercise 2.11.9 (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 $\lim_{x\to 0} f(x) = 0$.

2.12. **Primitive roots and power residues.** In the last section of chapter 2, we will study the generators of $(\mathbb{Z}/m\mathbb{Z})^\times$, which are called *primitive roots modulo m.* We will also study *n'th power residues modulo p,* which are solutions to $x^n \equiv a \pmod p$ for fixed a, n, p.

Primitive roots.

Definition 2.8.1. Given $m > 0$ and $g \in \mathbb{Z}$, if $[g]$ generates $(\mathbb{Z}/m\mathbb{Z})^\times$, then we say that g is a **primitive root modulo** m**.** This condition is equivalent to the order $|[g]| = |(\mathbb{Z}/m\mathbb{Z})^\times| = \phi(m)$.

Primitive roots have important applications in cryptography via the *discrete logarithm problem.*

Remark. As noted in Section 2.11 ([NZM91, Theorem 2.49]), for any element $[g] \in (\mathbb{Z}/m\mathbb{Z})^\times$ we have the divisibility

$$|[g]| \mid \#(\mathbb{Z}/m\mathbb{Z})^\times = \phi(m).$$

This also implies Euler's theorem, that $g^{\phi(m)} \equiv 1 \pmod m$.

In Exercise 2.11.2 (HW 5, Exercise 4), you'll prove a more general form of the following.

Proposition 2.8.1. *If $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$ has order r, then $[a^k]$ has order $\frac{r}{\gcd(r,k)}$. In particular, $[a^k]$ has order r if and only if $\gcd(r, k) = 1$.*

A consequence of this result (see Exercise 2.11.2, i.e., HW 5, Exercise 4) is that if $(\mathbb{Z}/m\mathbb{Z})^\times$ has a primitive root, then it has exactly $\phi(\phi(m))$ primitive roots (I won't spoil why).

The above work begs the question: given $m \in \mathbb{Z}^+$, do we know if there exists a primitive root modulo m? As it turns out, there is a precise classification of such $m \in \mathbb{Z}^+$.

Theorem 2.8.2. [NZM91, Theorem 2.41] *There exists a primitive root modulo m if and only if $m = 1, 2, 4, p^e$ or $2p^e$ where p is an odd prime.*

We'll forgo a proof of this result: it's quite detailed, including an analysis of factoring polynomials modulo p which we haven't touched on. The next homework will have an alternate proof as an exercise.

Primitive roots mod p^2 are enough. How do we find primitive roots? The following theorem shows that when considering odd prime powers, it's a primitive root modulo p^2 is also a primitive root modulo p^e for $e \geq 2$.

Theorem 2.8.3. [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^e for all $e \geq 2$.*

Proof. Suppose that g is a primitive root modulo p^2. Let us write n as the order of g modulo p^e; we want to show that $n = \phi(p^e) = p^{e-1}(p-1)$.

Some things to note: given that $g^n \equiv 1 \pmod{p^e}$, it follows that $g^n \equiv 1 \pmod{p^2}$, whence the order of g mod p^2 divides n; since this order is $\phi(p^2) = p(p-1)$, we have $p(p-1) \mid n$. On the other hand, by [NZM91, Theorem 2.49] we know that $n \mid |(\mathbb{Z}/p^e\mathbb{Z})^\times| = \phi(p^e) = p^{e-1}(p-1)$. Therefore, we have the divisibilities

$$p(p-1) \mid n \mid p^{e-1}(p-1).$$

One checks that this implies n is of the form $n = p^{a-1}(p-1)$ for some $2 \leq a \leq e$.

Our goal is to show that $a = e$. To this end, we will apply an inductive argument. By Euler's theorem or Fermat's little theorem, we know that $g^{p-1} \equiv 1 \pmod{p}$, so we can write $g^{p-1} = 1 + pb_1$ for some $b_1 \in \mathbb{Z}$. Observe that from $g^{p-1} \not\equiv 1 \pmod{p^2}$ (since its order mod p^2 is $p(p-1)$) we know that $p \nmid b_1$.

We check with the binomial theorem that

$$\begin{aligned} g^{p(p-1)} &= (1+pb_1)^p \\ &= \sum_{k=0}^{p} \binom{p}{k} 1^{p-k} \cdot (pb_1)^k \\ &= 1 + \binom{p}{1} pb_1 + \binom{p}{2} p^2 b_1^2 + \ldots + \binom{p}{p-1} p^{p-1} b_1^{p-1} + p^p b_1^p. \end{aligned}$$

As you probably observed in Exercise 2.1.11 (HW 2 Exercise 7), one has that p divides $\binom{p}{k} := \frac{p!}{(p-k)!k!}$ when $0 < k < p$. Therefore, for $k > 1$ one has $p^3 \mid \binom{p}{k} p^k b_1^k$, so that this number

$$\begin{aligned} g^{p(p-1)} &\equiv 1 + \binom{p}{1} pb_1 \\ &= 1 + p^2 b_1 \pmod{p^3}. \end{aligned}$$

In particular, $g^{p(p-1)} \not\equiv 1 \pmod{p^3}$, which forces its order modulo p^3 to be $p^2(p-1)$, hence g is a primitive root mod p^3.

In a similar manner, one can write $g^{p(p-1)} = 1 + p^2 b_2$ for $b_2 \in \mathbb{Z}$ with $p \nmid b_2$ and show that $g^{p^2(p-1)} \not\equiv 0 \pmod{p^4}$, and thus g is a primitive root mod p^4. Continuing this process, we conclude that g is a primitive root mod p^e. □

Therefore, for prime $p > 2$, if one has a primitive root g mod p^2, then it is also a primitive root mod $p^3, p^4, \ldots$. This still leaves the question of determining primitive roots modulo p and p^2 though! Here's a proposition towards this, based off our proof of [NZM91, Theorem 2.40].

Proposition 2.8.4. *For prime $p > 2$, if g is a primitive root modulo p and $g^{p-1} \not\equiv 1 \pmod{p^2}$, then g is a primitive root modulo p^2, and thus modulo $p^3, p^4, \ldots$.*

However, there is no general formula for finding primitive roots modulo prime $p > 2$. There are some techniques you can employ, though – HW 6 will address this.

n'th power residues mod p. Next, we turn our attention to studying n'th roots modulo p.

Definition 2.8.2. Given prime $p \in \mathbb{Z}^+$ and $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$, 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.

For example, studying $x^2 + 1 \bmod p$ amounts to asking whether -1 is a *quadratic residue mod p*. $\sqrt{-1} =: i$ exists in the complex numbers, but not necessarily mod p! Another example: $x^{10} - 7$ has a real root over $\mathbb{R}$ (namely, the usual $\sqrt[10]{7}$); however, "$\sqrt[10]{7}$" might not exist modulo p – but it could!

The following theorem describes when $x^n - a$ has a root modulo p. It is called **"Euler's criterion for n'th power residues".**

Theorem 2.8.5. [NZM91, Theorem 2.37] *For prime p with $p \nmid a$, 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 $\gcd(n, p-1)$ solutions.

To prove this, we'll use the following "linear congruence theorem" repeatedly (which we covered in §2.2).

Theorem. [NZM91, Theorem 2.17] *The congruence $ax \equiv b \pmod{m}$ has solutions if and only if $\gcd(a,m) \mid b$. In this case, there are $\gcd(a,m)$ distinct solutions modulo m, given by $(\frac{a}{\gcd(a,m)} \bmod \frac{m}{\gcd(a,m)})^{-1} \cdot \frac{b}{\gcd(a,m)} + \frac{m}{\gcd(a,m)} \cdot k$, where $0 \leq k < \gcd(a,m)$.*

Proof. First, some notation. By [NZM91, Theorem 2.41], there exists a primitive root $g \bmod p$. Thus, $p \nmid a$ implies $a = g^e$ for some $e \in \mathbb{Z}$. Let us also set $d := \gcd(n, p-1)$.

First, suppose that $x^n \equiv a \pmod{p}$ has a solution; call it x_0. Since $x_0^n \equiv a \pmod{p}$, we know that $\gcd(x_0^n, p) = \gcd(a,p) = 1$, thus $\gcd(x_0, p) = 1$. Therefore, by definition of a primitive root, we can write $x_0 = g^f$ for some $f \in \mathbb{Z}$. So our congruence becomes

$$g^{fn} \equiv g^e \pmod{p},$$

thus

$$g^{fn-e} \equiv 1 \pmod{p}.$$

Since the order of g is $\phi(p) = p-1$, we have $p-1 \mid (fn - e)$ (Exercise 4 from HW 5, see also Exercise 2.11.2), so that $fn \equiv e \pmod{p-1}$. In particular, this shows that f is a solution to the *linear congruence*

$$xn \equiv e \pmod{p-1}.$$

Thus, by the linear congruence theorem, we find that $d \mid e$ (it's also visibly clear). With this fact, we check that (modulo p)

$$\begin{aligned} a^{\frac{p-1}{\gcd(n,p-1)}} &=: a^{\frac{p-1}{d}} \\ &\equiv g^{\frac{e(p-1)}{d}} \\ &= (g^{p-1})^{\frac{e}{d}} \\ &\equiv 1 \pmod{p}. \end{aligned}$$

For the other direction, suppose that $a^{\frac{p-1}{d}} \equiv 1 \pmod{p}$. Then we have $g^{\frac{e(p-1)}{d}} \equiv 1 \pmod{p}$, which forces $|g| = p-1 \mid (e \cdot \frac{(p-1)}{d})$. Dividing both sides by $\frac{p-1}{d}$ shows that $d \mid e$, which by the linear congruence theorem implies $xn \equiv e \pmod{p-1}$ has a solution. Let f be a solution: then $fn \equiv e \pmod{p-1}$, thus we have modulo p that (using Exercise 4 in HW 4)

$$a \equiv g^e \equiv g^{fn} = (g^f)^n \pmod{p}.$$

Therefore, g^f is a solution to $x^n \equiv a \pmod{p}$.

Finally, in the case where we have a solution, the linear congruence theorem tells us there are precisely $\gcd(n, p-1)$ solutions. □

Example 2.8.1. Does the polynomial $x^5 - 6$ have any solutions modulo 101? 101 is prime, so by Euler's criterion for n'th power residues, this is the case if and only if $6^{\frac{100}{\gcd(5,100)}} \equiv 1 \pmod{101}$. Since $5 \mid 100$, we have $\frac{100}{\gcd(5,100)} = 20$. Is $6^{20} \equiv 1 \pmod{101}$? It is (why?); thus, $x^5 \equiv 6 \pmod{101}$ has 5 solutions modulo 101. Fix a primitive root g mod 101 and write $6 \equiv g^e \pmod{101}$; then we have $5 \mid e$ (this follows from the proof of Euler's criterion, but will also be an exercise in the next HW).

Now, unwinding the proof of Euler's criterion, the linear congruence theorem applied to the exponent congruence $5x \equiv e \pmod{100}$ gives $\gcd(5, 100) = 5$ has $\gcd(5, 100) = 5$ solutions modulo 100: they are $\frac{e}{5} + 20k$, with $0 \leq k < 5$ (check this!). Therefore, the roots of $x^5 - 6$ modulo 101 are the distinct elements $g^{\frac{e}{5}}, g^{\frac{e}{5}+20}, g^{\frac{e}{5}+40}, g^{\frac{e}{5}+60}$ and $g^{\frac{e}{5}+80}$ – whatever g might be! For example, one can show that 2 is a primitive root mod 101 and $g^{70} \equiv 6 \pmod{101}$; thus $\frac{e}{5} = 14$ and our roots become 2^{14+20k} where $0 \leq k < 5$.

An important case of Euler's criterion for n'th power residues is the quadratic case, where $n = 2$. This is commonly referred to as "Euler's criterion". It determines when an integer $a \in \mathbb{Z}$ coprime to p is a square modulo p.

Theorem 2.8.6. [NZM91, Corollary 2.38] *For an odd prime p with $p \nmid a$, one has that a is a square modulo p, i.e., $x^2 - a$ has a root mod p, i.e., $x^2 \equiv a \pmod{p}$ has a solution, if and only if*

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

In the above theorem, if a is a square modulo p, we say that a is a **quadratic residue modulo p.** Otherwise, we say that a is a *quadratic nonresidue mod p.* We will study quadratic residues in Chapter 3.

Exercise 2.8.1 (Finding the order of a group element)**.** In parts a) and b), let G be a finite group of order n.

a) Show that for all $g \in G$, one has $|g| = n$ if and only if $g^d \neq e$ for all $1 \leq d < n$ with $d \mid n$. Thus, knowing the factorization of $|G|$ gives a way to check whether an element generates G.
b) Show that for any $g \in G$, one has $g^{\frac{n}{p}} \neq e$ for all primes $p \mid n$ if and only if g has order n. This gives a way to test whether g is a generator for G. (*Hint:* calculate the order of $g^{\frac{n}{p}}$ for suitably chosen $p \mid n$, using the formula from Exercise 2.11.2).
c) Using part b), show that 2 is a primitive root modulo 19, and is *not* a primitive root modulo 17.

Exercise 2.8.2 (Solutions via Euler's criterion)**.** Show that 2 is a primitive root modulo 61. Then determine with proof the solutions to $x^6 \equiv 6 \pmod{61}$, if they exist.

Exercise 2.8.3 (Properties of the discrete logarithm)**.** This is a continuation of Exercise 2.1.10 (see also HW 3 Exercise 8). Let g be a primitive root modulo m. Recall that there exists a *discrete logarithm mod m with base g:* for each $a \in \mathbb{Z}$ coprime to m, there exists a unique exponent $0 \leq e < \phi(m)$ with $g^e \equiv a \pmod{m}$. Then we define the discrete logarithm as $\log_g(a) := e$.

a) Show that for $x, y \in \mathbb{Z}$ coprime to m, one has $\log_g(x \cdot y) \equiv \log_g(x) + \log_g(y) \pmod{\phi(m)}$.
b) Show that there is a "change of base" formula for discrete logarithms. More precisely, given another primitive root h modulo m, show that $\log_h(g)$ is coprime to $\phi(m)$, and that for all $[a] \in (\mathbb{Z}/m\mathbb{Z})^\times$ one has
$$\log_g(a) \equiv \frac{\log_h(a)}{\log_h(g)} \pmod{\phi(m)}$$
(here, $\frac{1}{\log_h(g)}$ represents the multiplicative inverse mod $\phi(m)$).
c) Assume 3 and 5 are primitive roots modulo $4802 = 2 \cdot 7^4$, and that the discrete logarithm $\log_3(5)$ is equal to 911. Compute the discrete logarithm $\log_5(81)$.

Bonus Exercise 2.8.4 (Fermat numbers)**.** 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 of odd a'th powers of two numbers, $x^a - y^a$.)
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 known Fermat primes are $3, 5, 17, 257$ and 65537. It is conjectured that these are the *only* Fermat primes.

Bonus Exercise 2.8.5 (Products of groups)**.** The following exercise outlines a proof of [NZM91, Theorem 2.41], on when primitive roots modulo m exist. It also gives some practice with products of groups.

Let G and H be two finite abelian groups.

a) Show that the order of any element $(g, h) \in G \times H$ is equal to $\operatorname{lcm}(|g|, |h|)$. Then explain how this would generalize to a longer product $G_1 \times G_2 \times \ldots \times G_n$.

b) Given a group homomorphism $\varphi\colon K \to G \times H$, define two homomorphisms $\varphi_1\colon K \to G$ and $\varphi_2\colon K \to H$ as follows: $\varphi_1(k) := g_k$ and $\varphi_2(k) := h_k$ where $\varphi(k) = (g_k, h_k)$. Show that φ_1, φ_2 are homomorphisms, and that

$$\varphi(k) = (\varphi_1(k), \varphi_2(k)).$$

Thus, any homomorphism to a product is a product of homomorphisms.

c) Let $m \in \mathbb{Z}^+$ have the prime factorization $m = \prod_{i=1}^{r} p_i^{e_i}$. Show that for any $a \in \mathbb{Z}$, if a is coprime to m then

$$|\overline{a} \mod m| = \operatorname{lcm}\{|\overline{a} \mod p_i^{e_i}|\}_{i=1}^{r}.$$

d) Use part c) to give an alternate proof for the existence of primitive roots:

Theorem. [NZM91, Theorem 2.41] *There exists a primitive root modulo m if and only if $m = 1, 2, 4, p^e$ or $2p^e$ where p is an odd prime.*

4. Chapter 4: Some Functions of Number Theory

In this chapter, we will study several functions persistently seen in number theory (especially elementary, analytic and combinatorial number theory). We've already seen Euler's totient function $\phi(n)$; this is an *arithmetic function,* which we will study more generally in §4.2.

4.1. Greatest integer function.

In this subsection, we will study the *greatest integer function,* also called the *floor function.* We will then apply it towards a neat formula for computing the p-adic valuation of any factorial number $n!$.

Definition 4.1.1. The **floor function,** denoted by $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$, is defined as follows: given $x \in \mathbb{R}$, we let $\lfloor x \rfloor$ be the largest positive integer less than or equal to x.

Example 4.1.1.

- $\lfloor 4.7 \rfloor = 4$;
- $\lfloor 100 \rfloor = 100$;
- $\lfloor \pi \rfloor = 3$.

Auxiliary to the floor function, we can define the **fractional part** function $\{\cdot\}\colon \mathbb{R} \to \mathbb{R}$, via

$$\{x\} := x - \lfloor x \rfloor .$$

One has $0 \leq \{x\} < 1$.

The floor function has many properties – some of these are listed in the following theorem. The hope is that, once we work through a few of their proofs, we'll have better intuition for values of the floor function.

Theorem 4.1.1. [NZM91, Theorem 4.1] *Given real numbers x and y, one has the following.*

(1) $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$, $x - 1 < \lfloor x \rfloor \leq x$ *and* $0 \leq x - \lfloor x \rfloor < 1$.

(2) $\lfloor x \rfloor = \sum_{1 \leq i \leq x} 1$ *if* $x \geq 0$.
(3) $\lfloor x + m \rfloor = \lfloor x \rfloor + m$ *if* $m \in \mathbb{Z}$.
(4) $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1$.
(5) $\lfloor x \rfloor + \lfloor -x \rfloor = \begin{cases} 0 & \text{if } x \in \mathbb{Z} \\ -1 & \text{else} \end{cases}$.
(6) $\left\lfloor \frac{\lfloor x \rfloor}{m} \right\rfloor = \left\lfloor \frac{x}{m} \right\rfloor$ *if* $m \in \mathbb{Z}^+$.
(7) $-\lfloor -x \rfloor$ *is the least integer* $\geq x$.
(8) $\left\lfloor x + \frac{1}{2} \right\rfloor$ *is the closest integer to* x. *If two integers are equally near* x, *then it is the larger of the two.*
(9) $-\left\lfloor -x + \frac{1}{2} \right\rfloor$ *is the closest integer to* x. *If two integers are equally near* x, *then it is the smaller of the two.*
(10) *If* $n, a \in \mathbb{Z}^+$, *then* $\left\lfloor \frac{n}{a} \right\rfloor$ *is the number of integer integers of* a *between* 1 *and* n.

The book proves all of these, but we'll prove just a few to get a feeling for these types of proofs.

Proof. We'll prove (1), (3), (4), (6) and (10).

- (1): by definition, $\lfloor x \rfloor \leq x$. If $x < \lfloor x \rfloor + 1$ weren't true, then we'd have $\lfloor x \rfloor + 1 \leq x$, which contradicts $\lfloor x \rfloor$ being the largest integer below x. So $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$ holds. Similarly, if $x - 1 < \lfloor x \rfloor$ isn't true, then $\lfloor x \rfloor \leq x - 1$, thus $\lfloor x \rfloor + 1 \leq x$, which contradict maximality of $\lfloor x \rfloor$. Thus, $x - 1 < \lfloor x \rfloor \leq x$ also holds. Finally, the last inequality follows from the first one: subtract $\lfloor x \rfloor$ from all sides.
- (3): for all $n \in \mathbb{Z}$, one has $n \leq x$ if and only if $n + m \leq x + m$. Thus, $\lfloor x \rfloor + m$ is the largest integer $\leq x + m$.
- (4): let us write $x = \lfloor x \rfloor + \{x\}$ and $y = \lfloor y \rfloor + \{y\}$, where $0 \leq \{x\}, \{y\} < 1$. We see that

$$\lfloor x + y \rfloor = \lfloor \lfloor x \rfloor + \lfloor y \rfloor + \{x\} + \{y\} \rfloor . \tag{1}$$

  Since $\{x\}, \{y\} > 0$, it follows by Equation (1) that

$$\lfloor x + y \rfloor \geq \lfloor x \rfloor + \lfloor y \rfloor .$$

  On the other hand, since $0 \leq \{x\} + \{y\} < 2$, we have $\lfloor x \rfloor + \lfloor y \rfloor + \{x\} + \{y\} < \lfloor x \rfloor + \lfloor y \rfloor + 2$, and thus by Equation (1) we have $\lfloor x + y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$ or $\lfloor x \rfloor + \lfloor y \rfloor + 1$, so we find that

$$\lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1.$$

  This proves (4).
- (6): let us write $\lfloor x \rfloor = mq + r$, $0 \leq r < m$. Then we check that

$$\left\lfloor \frac{\lfloor x \rfloor}{m} \right\rfloor = \left\lfloor \frac{mq + r}{m} \right\rfloor = \left\lfloor q + \frac{r}{m} \right\rfloor .$$

  by (3), we have $\left\lfloor q + \frac{r}{m} \right\rfloor = q + \left\lfloor \frac{r}{m} \right\rfloor$. Then, since $0 \leq r < m$, we have $\left\lfloor \frac{r}{m} \right\rfloor = 0$, so we deduce that $\left\lfloor \frac{\lfloor x \rfloor}{m} \right\rfloor = q$. On the other hand, we can write $x = \lfloor x \rfloor + \{x\} =$

$mq + r + \{x\}$, and so

$$\left\lfloor \frac{x}{m} \right\rfloor = \left\lfloor \frac{mq + r + \{x\}}{m} \right\rfloor = q + \left\lfloor \frac{r + \{x\}}{m} \right\rfloor.$$

Since $0 \leq r \leq m-1$ and $0 \leq \{x\} <$, we have $0 \leq r + \{x\} < m$, and thus $\left\lfloor \frac{r+\{x\}}{m} \right\rfloor = 0$. We conclude that $\left\lfloor \frac{x}{m} \right\rfloor = \left\lfloor \frac{\lfloor x \rfloor}{m} \right\rfloor = q$.

- (10): let $a, 2a, 3a, \ldots, ja$ be all multiples of a in $[1, n]$. We wish to show that $\left\lfloor \frac{n}{a} \right\rfloor = j$. Observe that $(j+1)a > n$, so

$$ja \leq n < (j+1)a,$$

i.e.,

$$j \leq \frac{n}{a} < j + 1.$$

Thus, j is the largest integer below $\frac{n}{a}$, which implies that $\left\lfloor \frac{n}{a} \right\rfloor = j$. □

The next theorem describes the p-adic valuation of a factorial number. This formula is called *de Polignac's formula.*

Theorem 4.1.2. [NZM91, Theorem 4.2] *Let $p \in \mathbb{Z}^+$ be a prime. Then the largest power $p^e \mid n!$ is*

$$e = \sum_{i=1}^{\infty} \left\lfloor \frac{n}{p^i} \right\rfloor.$$

Proof. This sum is finite since eventually $p^i > n$, thus $\left\lfloor \frac{n}{p^i} \right\rfloor = 0$. We will prove this formula by induction.

Base case, $n = 1$. True. Assume the formula is true for $n-1$, $n \geq 1$: that is, de Polignac's formula works for $p^e \mid (n-1)!$. We will show that it's true for $n!$ too.

Let $f \geq 0$ be such that $p^f \parallel n$; i.e., $f = v_p(n)$. Since $n! = n \cdot (n-1)!$, we have $v_p(n!) = v_p(n) + v_p((n-1)!)$, so for this theorem to be true, we need to prove that

$$\sum_{i=1}^{\infty} \left\lfloor \frac{n}{p^i} \right\rfloor - \sum_{i=1}^{\infty} \left\lfloor \frac{n-1}{p^i} \right\rfloor = f,$$

i.e.,

$$\sum_{i=1}^{\infty} \left(\left\lfloor \frac{n}{p^i} \right\rfloor - \left\lfloor \frac{n-1}{p^i} \right\rfloor \right) = f$$

(this is a finite sum, so rearranging terms is OK). However, we check that

$$\left\lfloor \frac{n}{p^i} \right\rfloor - \left\lfloor \frac{n-1}{p^i} \right\rfloor = \begin{cases} 1 & \text{if } p^i \mid n \\ 0 & \text{if } p^i \nmid n \end{cases}.$$

To see this: note that if $p^i \mid n$, then $p^i \nmid (n-1)$, and so the floor of $\frac{n-1}{p^i}$ rounds down an integer by [NZM91, Theorem 4.1.(10)]; thus the difference is 1. If $p^i \nmid n$, then both terms round down, and their floors are equal by [NZM91, Theorem 4.1.(10)]. Therefore, for each $1 \leq k \leq j$, we have $\left\lfloor \frac{n}{p^i} \right\rfloor - \left\lfloor \frac{n-1}{p^i} \right\rfloor = 1$, and the rest of the terms in the sum are zero. We conclude that the sum is j. □

Example 4.1.2. How much does 11 divide (2024!)? Noting that $11^4 = 14641 > 2024$, by de Polignac's formula we know the 11-adic valuation of (2024!) is $\left\lfloor \frac{2024}{11} \right\rfloor + \left\lfloor \frac{2024}{121} \right\rfloor + \left\lfloor \frac{2024}{1331} \right\rfloor$. We check (e.g. using [NZM91, Theorem 4.1.(10)] or a calculator) that this sum is $184 + 16 + 1 = 201$; thus, $11^{201} \parallel (2024)!$.

While de Polignac's formula might seem like a strange result, it has a lot of application in combinatorial-flavored number theory and recreational mathematics. A bonus exercise in the HW will apply it towards understanding the factorization of binomial coefficients $\binom{n}{k}$.

Exercise 4.1.1 (Using iteration to compute valuations)**.** Using de Polignac's formula, for any prime $p \in \mathbb{Z}^+$ and $n \in \mathbb{Z}^+$ one can compute the p-adic valuation[1] of $n!$ via

$$v_p(n!) = \sum_{k \geq 1} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

This exercise gives an iterative way to compute each term $\left\lfloor \frac{n}{p^k} \right\rfloor$ in the sum.

a) Prove that for each $k \geq 1$, one has

$$\left\lfloor \frac{n}{p^k} \right\rfloor = \left\lfloor \frac{\left\lfloor \frac{n}{p^{k-1}} \right\rfloor}{p} \right\rfloor.$$

Therefore, computing $\left\lfloor \frac{n}{p} \right\rfloor$ lets us compute $\left\lfloor \frac{n}{p^2} \right\rfloor = \left\lfloor \frac{\left\lfloor \frac{n}{p} \right\rfloor}{p} \right\rfloor$, which then lets us compute $\left\lfloor \frac{n}{p^3} \right\rfloor = \left\lfloor \frac{\left\lfloor \frac{n}{p^2} \right\rfloor}{p} \right\rfloor$, and so on.

b) Using the technique in part a), compute the 5-adic valuation of 3443!.

Exercise 4.1.2 (A factorial fraction)**.**

a) Show that for $n \in \mathbb{Z}^+$, one has $(n!)^2 \mid (2n)!$.

b) Show additionally that $\frac{(2n)!}{(n!)^2}$ is even; equivalently, show that the 2-adic valuation of $\frac{(2n)!}{(n!)^2}$ is positive.

Bonus Exercise 4.1.3 (Kummer's Theorem)**.** This exercise will give a formula for computing the power of a prime which divides a binomial coefficient.

Fix a prime $p \in \mathbb{Z}^+$.

a) Show that every integer $n \geq 0$ can be uniquely written in "base p", with the form

$$n = n_0 + n_1 p + n_2 p^2 + \ldots + n_r p^r,$$

where the $n_i \in \mathbb{Z}$ each satisfy $0 \leq n_i < p^i$, and $n_r \neq 0$.

b) With notation as above, let $S(n) := S_p(n) := n_0 + n_1 + \ldots + n_r$ be the sum of the base p digits of n. Then use de Polignac's formula to prove the following theorem.

[1]Recall that the *p-adic valuation of n,* denoted $v_p(n)$, is the largest power of p which divides n. This definition extends to rational numbers $\frac{a}{b}$ via $v_p\left(\frac{a}{b}\right) := v_p(a) - v_p(b)$.

Theorem (Kummer)**.** *Let $p \in \mathbb{Z}^+$ be a prime. Then for all integers $0 \leq m \leq n$, one has*

$$v_p\left(\binom{n}{m}\right) = \frac{S(m) + S(n-m) - S(n)}{p-1}.$$

4.2. **Arithmetic functions.** This subsection will study functions of number-theoretic interest related to divisors of integers. For now, let us define a very general class of functions.

Definition 4.2.1. Any function $f\colon \mathbb{Z}^+ \to \mathbb{C}$ is called an *arithmetic function.*

Let us define several arithmetic functions of particular number-theoretic interest.

Definition 4.2.2. Given an integer $n \in \mathbb{Z}^+$, we will define the following arithmetic functions:

1. $d(n)$ is the number of positive divisors of n;
2. $\sigma(n)$ is the sum of positive divisors of n;
3. For $k \geq 0$, $\sigma_k(n)$ is the sum of k'th powers of positive divisors of n;
4. $\omega(n)$ is the number of distinct primes dividing n;
5. $\Omega(n)$ is the number of primes dividing n *counting multiplicity.*

Example 4.2.1.

- $d(30) = d(2 \cdot 3 \cdot 5) = 8$ (divisors are $1, 2, 3, 5, 6, 10, 15, 30$).
- $\sigma(28) = \sigma(2 \cdot 2 \cdot 7) = 1 + 2 + 4 + 7 + 14 + 28 = 56$.
- $\sigma_3(6) = 1^3 + 2^3 + 3^3 + 6^3 = 1 + 8 + 27 + 216 = 252$.
- $\omega(100) = \omega(2^2 \cdot 5^2) = 2$.
- $\Omega(100) = \Omega(2 \cdot 2 \cdot 5 \cdot 5) = 4$.

Each of these functions can be rewritten in the following way. We will use the notation $\sum_{d|n}$ to denote a sum over the positive divisors of n.

1. $d(n) = \sum_{d|n} 1$;
2. $\sigma(n) = \sum_{d|n} d$;
3. $\sigma_k(n) = \sum_{d|n} d^k$;
4. $\omega(n) = \sum_{p|n} 1$;
5. $\Omega(n) = \sum_{p|n} v_p(n) = \sum_{p^a|n, a\geq 1} 1$.

Here is another formula for $d(n)$.

Theorem 4.2.1. [NZM91, Theorem 4.3] *For $n \in \mathbb{Z}^+$, one has*

$$d(n) = \prod_{p^e \| n} (e+1).$$

Proof. Let us write the prime factorization of n as

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

Any divisor $d \mid n$ has the form

$$d = \prod_{i=1}^{r} p_i^{f_i}$$

where each $0 \leq f_i \leq e_i$. There are $e_i + 1$ choices for f_i, thus the number of divisors of n is equal to

$$\prod_{i=1}^{r}(e_i + 1),$$

which proves the result. □

Example 4.2.2. One has $d(30) = d(2 \cdot 3 \cdot 5) = 2^3 8$. One also has $d(1,000,000) = d(10^6) = d(2^6 \cdot 5^6) = 7 \cdot 7 = 49$.

Multiplicative functions. Recall the following definition from §2.3.

Definition 4.2.3. If an arithmetic function f satisfies $f(ab) = f(a)f(b)$ for all coprime $a, b \in \mathbb{Z}^+$, then f is said to be *multiplicative*. If $f(ab) = f(a)f(b)$ is true for *all* positive integers, then f is said to be *totally multiplicative*.

Thus, multiplicative functions are determined by their values on prime powers, and totally multiplicative functions by their values on primes.

Example 4.2.3. By our theorem we've just proved, $d(n)$ is multiplicative, since $d(p^e) = e + 1$. Euler's totient function $\phi\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ is also multiplicative, and thus given $n = \prod_{p^e \| n} p^e$, one has

$$\phi(n) = \prod_{p^e \| n} \phi(p^e).$$

We also defined a function $\phi_f\colon \mathbb{Z}^+ \to \mathbb{Z}_{\geq 0}$ which counted solutions modulo n for a fixed polynomial $f \in \mathbb{Z}[x]$; this was also multiplicative (by the CRT).

The next result gives a way to construct new multiplicative functions from other ones – it will also give us a way to prove that some of our distinguished arithmetic functions are multiplicative.

Theorem 4.2.2. [NZM91, Theorem 4.4] *Let $f\colon \mathbb{Z}^+ \to \mathbb{C}$ be a multiplicative function, and set*

$$F(n) := F_f(n) := \sum_{d|n} f(d).$$

Then F is a multiplicative function.

Proof. Let $a, b \in \mathbb{Z}^+$ be coprime; we need to show that $F(ab) = F(a)F(b)$. To do this, we will analyze the indexing set in $F(ab) := \sum_{d|ab} f(d)$.

Since a and b are coprime, for any divisor $d \mid ab$, we can write $d = \gcd(a, d)\gcd(b, d)$ (prove it for yourself!). In this case, let can write $d_1 := \gcd(a, d)$ and $d_2 := \gcd(b, d)$. On the other hand, given any two divisors $d_1 \mid a$ and $d_2 \mid b$, we have $d_1 d_2 \mid ab$. Thus, there is a bijection between the set of positive divisors of ab, and pairs d_1, d_2 of positive divisors of a and b, respectively. Therefore, as indexing sets they're the same, which

means we can make the following calculations (noting that f is multiplicative)

$$\begin{aligned} F(ab) &:= \sum_{d|ab} f(d) \\ &= \sum_{d_1|a, d_2|b} f(d_1 d_2) \\ &= \sum_{d_1|a, d_2|b} f(d_1) f(d_2) \\ &= \left(\sum_{d_1|a} f(d_1) \right) \left(\sum_{d_2|b} f(d_2) \right) \\ &= F(a)F(b). \end{aligned}$$

We conclude that F is multiplicative. □

Remark. This theorem could have been used to show that $d(n)$ is multiplicative, since $d(n) := \sum_{d|n} 1 =: F_1(n)$ and the constant function $1\colon \mathbb{Z}^+ \to \mathbb{C}$ via $1(a) := 1$ is multiplicative.

Theorem 4.2.3. [NZM91, Theorem 4.5] *The function $\sigma(n)$ is multiplicative, and*

$$\sigma(n) = \prod_{p^e \| n} \sigma(p^e) = \prod_{p^e \| n} \frac{p^{e+1} - 1}{p - 1}.$$

Proof. By definition,

$$\sigma(n) := \sum_{d|n} d =: F_\iota(d)$$

where $\iota\colon \mathbb{Z}^+ \to \mathbb{C}$ is the identity function, $\iota(d) := d$. Since ι is multiplicative, so is $\sigma(n)$ by [NZM91, Theorem 4.4].

In general, one has the polynomial factorization

$$(1 + x + x^2 + \ldots + x^n)(x - 1) = x^n - 1,$$

so that for positive integers $a \neq 1$ one has

$$1 + a + a^2 + \ldots + a^n = \frac{a^n - 1}{a - 1}.$$

Therefore, since p^e only has the divisors $1, p, p^2, \ldots, p^e$, we have $\sigma(p^e) = 1 + p + p^2 + \ldots + p^e = \frac{p^n - 1}{p - 1}$. □

The following gives a new algebra-combinatorial identity for $\phi(n)$.

Theorem 4.2.4. [NZM91, Theorem 4.6] *For each $n \in \mathbb{Z}^+$, one has*

$$\sum_{d|n} \phi(d) = n.$$

Proof. Since $\phi(n)$ is a multiplicative function, so is

$$F(n) := \sum_{d|n} \phi(d)$$

by [NZM91, Theorem 4.4]. We want to show that for the identity function $\iota(n) := n$, one has $F(n) = \iota(n)$. Since $\iota(n) := n$ is also multiplicative, it suffices to show that $F(p^e) = \iota(p^e)$.

Since $\phi(p^f) = p^e - p^{f-1}$ for $f \geq 1$, and since the divisors of p^e are exactly $1, p, p^2, \ldots, p^e$, we make the following calculations:

$$\begin{aligned} F(p^e) &:= \sum_{d|p^e} \phi(d) \\ &= \sum_{0 \leq f \leq e} \phi(p^f) \\ &= 1 + \sum_{1 \leq f \leq e} \phi(p^f) \\ &= 1 + \sum_{1 \leq f \leq e} (p^f - p^{f-1}) \\ &= 1 + (p-1) + (p^2 - p) + \ldots + (p^e - p^{e-1}) \\ &= 1 - 1 + p^e \\ &= p^e. \end{aligned}$$

Thus, $F(p^e) = p^e$, which by multiplicativity of F and ι concludes our proof. □

Exercise 4.2.1 (Fibers of some arithmetic functions)**.** In HW 4 Exercise 3, we showed that $\phi(x) = n$ has a finite number of solutions.

a) Show that $\sigma(x) = n$ has a finite number of solutions.
b) Show that $d(x) = n$ has an *infinite* number of solutions if $n > 1$.

Exercise 4.2.2 (Liouville's lambda function)**.** Define an arithmetic function $\lambda \colon \mathbb{Z}^+ \to \mathbb{C}$ via

$$\lambda(n) := (-1)^{\Omega(n)}.$$

This is called *Liouville's lambda function.*

a) Show that λ is totally multiplicative.
b) Show that

$$\sum_{d|n} \lambda(d) = \begin{cases} 1 & \text{if } n \text{ is a perfect square} \\ 0 & \text{otherwise.} \end{cases}$$

(*Hint:* for the case where n is a perfect square, fix a prime $p \mid n$ and write $n = p^e m$ with $p \nmid m$. Then show that $\sum_{d|n} \lambda(d) = \sum_{d|m} \lambda(d)$.)

Bonus Exercise 4.2.3. It is not hard to show that $\phi(n) < n$ for all integers $n > 1$. However, with the formula

$$\phi(n) = \prod_{p^e \| n} p^{e-1} \cdot (p-1),$$

it would seem that $\phi(n)$ should also be smaller than

$$n = \prod_{p^e \| n} p^{e-1} \cdot p$$

by a "consistently small amount." This exercise quantifies that: we will show that $\phi(n)$ is bigger than $n^{1-\epsilon}$ for any $\epsilon > 0$, provided that n is sufficiently large with respect to ϵ.

a) Let $f : \mathbb{Z} \to \mathbb{C}$ be a multiplicative arithmetic function such that

$$\lim_{p^k \to \infty} f(p^k) = 0$$

as p^k ranges through all prime powers. Prove that

$$\lim_{n \to \infty} f(n) = 0.$$

Thus, convergence to zero for a multiplicative function $f(x)$ can be checked using prime powers.

b) Show that for any $\epsilon > 0$, for all sufficiently large $n \in \mathbb{Z}^+$ one has

$$n^{1-\epsilon} < \phi(n) < n.$$

(*Hint:* for the inequality $n^{1-\epsilon} < \phi(n)$, it will suffice by part a) to show that

$$\lim_{p^k \to \infty} \frac{p^{k(1-\epsilon)}}{\phi(p^k)} = 0.)$$

Bonus Exercise 4.2.4 (Finite subgroup of a field's unit group)**.** This exercise will prove the following result:

Theorem. *For any field F, one has that any finite subgroup G of $F^\times$ is cyclic.*

We will prove this the following way. Let us set $n := |G|$. Define for each $d \mid n$ the subset

$$G_d := \{g \in G : |g| = d\}.$$

We will show that $G_n \neq \emptyset$, which forces G to be cyclic.

a) Assume that $G_d \neq \emptyset$. Show that $|G_d| = \phi(d)$ by considering the roots of the polynomial $x^d - 1$ over F and applying the following generalization of [NZM91, Theorem 2.26]:

Proposition. *Over a field F, any nonzero polynomial $f \in F[x]$ has at most* $\deg(f)$ *roots.*

b) Using the fact that G is a disjoint union of the sets G_d (which are either empty or have size $\phi(d)$), prove using [NZM91, Theorem 4.6] that for each $d \mid n$ one must have $|G_d| = \phi(d)$. Show that this implies the theorem.

4.3. **The Möbius inversion formula.** Last section, we created new arithmetic functions from old ones: given $f\colon \mathbb{Z}^+ \to \mathbb{C}$, the function

$$F(n) := F_f(n) := \sum_{d|n} f(d)$$

is also arithmetic. If f is multiplicative, then so is F [NZM91, Theorem 4.4]. We used this to show that e.g. $\sigma(n)$ is multiplicative.

In this section, we will see how to "undo" the construction of F, and produce f from it. This turns out to have several nifty applications in number theory.

Definition 4.3.1. Define the *Möbius function* $\mu\colon \mathbb{Z}^+ \to \mathbb{C}$ via

$$\mu(n) := \begin{cases} (-1)^{\omega(n)} & \text{if } n \text{ is squarefree} \\ 0 & \text{otherwise.} \end{cases}$$

(recall that $\omega(n)$ is the number of distinct prime divisors of n.)

Example 4.3.1.

- $\mu(10) = (-1)^2 = 1$;
- $\mu(30) = (-1)^3 = -1$;
- $\mu(100) = \mu(2^2 \cdot 5^2) = 0$;
- $\mu(1) = (-1)^0 = 1$.

As it turns out, μ is multiplicative, and the corresponding function $F := F_\mu$ for μ has a concise formula:

Theorem 4.3.1. [NZM91, Theorem 4.7] *The function μ is multiplicative, and*

$$F_\mu(n) := \sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. It is not too hard to check that μ is multiplicative, i.e., $\gcd(a, b) = 1 \Rightarrow \mu(ab) = \mu(a)\mu(b)$.

For the second part: since μ is multiplicative, so is F by [NZM91, Theorem 4.4]. Thus, it's enough to check that the formula above holds for 1 and prime powers p^e. Clearly $F(1) = \mu(1) = 1$; on the other hand,

$$\begin{aligned} F(p^e) &= \mu(1) + \mu(p) + \mu(p^2) + \ldots + \mu(p^e) \\ &= \mu(1) + \mu(p) \\ &= 1 + -1 \\ &= 0. \end{aligned}$$

This proves the equation. □

Now we can prove the Möbius inversion formula.

Theorem 4.3.2. [NZM91, Theorem 4.8] *If $f, F\colon \mathbb{Z}^+ \to \mathbb{C}$ are two arithmetic functions satisfying*

$$F(n) = \sum_{d|n} f(d),$$

then one has

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right).$$

Proof. The proof follows from a clever indexing trick. Observe that the RHS is

$$\begin{aligned}
\sum_{d|n} \mu(d) F\left(\frac{n}{d}\right) &= \sum_{d|n} \mu(d) \cdot \left(\sum_{k|\frac{n}{d}} f(k) \right) \\
&= \sum_{d|n} \mu(d) \cdot \left(\sum_{k \geq 1 : dk|n} f(k) \right) \\
&= \sum_{d,k \geq 1 : dk|n} \mu(d) f(k) \\
&= \sum_{d,k \geq 1 : dk|n} f(d) \mu(k) \\
&= \sum_{d|n} f(d) \cdot \left(\sum_{k \geq 1 : dk|n} \mu(k) \right) \\
&= \sum_{d|n} f(d) \cdot \left(\sum_{k|\frac{n}{d}} \mu(k) \right)
\end{aligned}$$

By the previous theorem [NZM91, Theorem 4.7], we have that $\sum_{k|\frac{n}{d}} \mu(k) = 0$ when $\frac{n}{d} > 1$, so this equation becomes

$$\sum_{d|n} f(d) \cdot \left(\sum_{k|\frac{n}{d}} \mu(k) \right) = f(n).$$

We conclude that

$$\sum_{d|n} \mu(d) F\left(\frac{n}{d}\right) = f(n). \qquad \square$$

The following theorem shows when an arithmetic function $F \colon \mathbb{Z}^+ \to \mathbb{C}$ must be F_f for some $f \colon \mathbb{Z}^+ \to \mathbb{C}$.

Theorem 4.3.3. [NZM91, Theorem 4.9] *If for functions $f, F \colon \mathbb{Z}^+ \to \mathbb{C}$ we have*

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right),$$

then

$$F(n) = \sum_{d|n} f(d),$$

i.e., $F = F_f$.

Proof. Let's follow our nose and check out $\sum_{d|n} f(d)$:

$$\begin{aligned}\sum_{d|n} f(d) &= \sum_{d|n}\left(\sum_{k|d}\mu(k)F\left(\frac{d}{k}\right)\right)\\ &= \sum_{d|n}\left(\sum_{k|d}\mu\left(\frac{d}{k}\right)F\left(k\right)\right).\end{aligned}$$

Here, we noted that the set of divisors k of d, corresponds to the set of divisors $\frac{d}{k}$ of d. Now, for each $k \mid d$, group together all terms in the double sum which involve $F(k)$; they're each in a product with terms of the form $\mu\left(\frac{d}{k}\right)$ where $k \mid d \mid n$, i.e., $\frac{d}{k} \mid \frac{n}{k}$; such terms $\mu\left(\frac{d}{k}\right)$ are the same as $\mu(r)$ where $r \mid \frac{n}{k}$. Therefore, our sum can be rewritten as

$$\begin{aligned}\sum_{d|n}\left(\sum_{k|d}\mu\left(\frac{d}{k}\right)F\left(k\right)\right) &= \sum_{k|n}\sum_{r|\frac{n}{k}}\mu(r)F(k)\\ &= \sum_{k|n}F(k)\left(\sum_{r|\frac{n}{k}}\mu(r)\right).\end{aligned}$$

By [NZM91, Theorem 4.7], we have $\sum_{r|\frac{n}{k}}\mu(r) = 0$ for $r > 1$ and 1 otherwise ($r = 1$ if and only if $k = n$). Therefore, our sum becomes

$$\sum_{k|n}F(k)\left(\sum_{r|\frac{n}{k}}\mu(r)\right) = F(n).$$

Thus, we have shown that

$$\sum_{d|n} f(d) = F(n),$$

which proves that $F = F_f$. □

Example 4.3.2. We can use [NZM91, Theorem 4.9] to re-prove that $\phi(p^e) = p^e - p^{e-1}$ when $e \geq 1$ (this is Exercise 8 in HW 2). We know by [NZM91, Theorem 4.6] in §4.2 that

$$\sum_{d|n}\phi(d) = n,$$

and thus the function $F(n) := n$ satisfies $F = F_\phi$. Thus, Möbius inversion implies

$$\phi(n) = \sum_{d|n}\mu(d)F\left(\frac{n}{d}\right),$$

i.e.,

$$\phi(n) = \sum_{d|n}\mu(d)\cdot\frac{n}{d},$$

i.e.,

$$\phi(n) = n \cdot \sum_{d|n} \frac{\mu(d)}{d}.$$

The inner function $g(d) := \frac{\mu(d)}{d}$ is multiplicative (check it!), thus so is $\phi(n)$ by [NZM91, Theorem 4.4], §4.2 – note that $\phi(n) = F_{\frac{\mu(n)}{n}}$.

In particular, we check that

$$\begin{aligned}
\phi(p^e) &= p^e \cdot \sum_{d|p^e} \frac{\mu(d)}{d} \\
&= p^e \left(1 + \frac{\mu(p)}{p} + \frac{\mu(p^2)}{p^2} + \ldots + \frac{\mu(p^e)}{p^e}\right) \\
&= p^e \left(1 + \frac{\mu(p)}{p}\right) \\
&= p^e \left(1 - \frac{1}{p}\right) \\
&= p^e - p^{e-1}.
\end{aligned}$$

Exercise 4.3.1 (A twist on the $\phi(d)$ sum)**.**

a) Calculate the sum

$$\sum_{d|n} \mu(d)\phi(d)$$

for $n = 5, 6, 7, 8, 9, 10$.

b) Prove that for all even $n \in \mathbb{Z}^+$ one has

$$\sum_{d|n} \mu(d)\phi(d) = 0.$$

Bonus Exercise 4.3.2 (Formula for cyclotomic polynomials)**.** For each $n \geq 1$, let us define the *n-cyclotomic polynomial* $\Phi_n(x) \in \mathbb{Z}[x]$ recursively: for $n = 1$ we set $\Phi_1(x) := x - 1$, and for $n > 1$ we define $\Phi_n(x)$ as the unique monic[2] irreducible[3] polynomial in $\mathbb{Z}[x]$ that divides $x^n - 1$ and does not divide $x^k - 1$ for $1 \leq k < n$. As it turns out, one has

$$x^n - 1 = \prod_{d|n} \Phi_d(x).$$

Each $\Phi_n(x)$ is irreducible, and its roots are of the form $e^{2\pi i \frac{a}{n}}$ where $\gcd(a, n) = 1$; these roots are called *roots of unity,* and have many applications in algebraic number theory. The first few cyclotomic polynomials are $\Phi_1(x) = x - 1, \Phi_2(x) = x + 1, \Phi_3(x) = x^2 + x + 1, \Phi_4(x) = x^2 + 1, \Phi_5(x) = x^4 + x^3 + x^2 + x + 1$ and $\Phi_6(x) = x^2 - x + 1$.

Use the Möbius inversion formula to prove the following formula for $\Phi_n(x)$:

[2]A *monic* polynomial has a leading term coefficient of 1.

[3]A polynomial is *irreducible* if it is not a product of two polynomials with lower degree.

Theorem. *For each $n \geq 1$, one has*

$$\Phi_n(x) = \prod_{d|n} (x^{\frac{n}{d}} - 1)^{\mu(d)}.$$

(*Hint:* Fix $x \in \mathbb{R}$ with $|x| \neq 1$, so that $x^n - 1 \neq 0$ and $\phi_d(x) \neq 0$ for $d \mid n$; then take complex logarithms. Conclude that the relation is true by the identity theorem from complex analysis.)

Bonus Exercise 4.3.3 (Möbius inversion formula to abelian groups)**.** This exercise gives a more general version of arithmetic functions and the Möbius inversion formula, and an alternative proof for Bonus Exercise 4.3.2.

Let G be an abelian group, written multiplicatively. Call any map $f\colon \mathbb{Z}^+ \to G$ a *G-arithmetic function.*

a) Mimic the proof of the usual Möbius inversion formula [NZM91, Theorem 4.8] to prove the following "G-Möbius inversion formula" for G-arithmetic functions:

Theorem (G-Möbius inversion formula)**.** *For G-arithmetic functions $f, F\colon \mathbb{Z}^+ \to G$ with*

$$F(n) = \prod_{d|n} f(d),$$

one has

$$f(n) = \prod_{d|n} F\left(\frac{n}{d}\right)^{\mu(d)}.$$

b) Use the G-Möbius inversion formula to give an alternative proof of Bonus Exercise 4.3.2:

Theorem. *For each $n \geq 1$, one has*

$$\Phi_n(x) = \prod_{d|n} (x^{\frac{n}{d}} - 1)^{\mu(d)}.$$

(*Hint:* Take $G := \mathbb{Q}(x)^\times$ as the unit group of the fraction field of $\mathbb{Z}[x]$, and apply the G-Möbius inversion formula to the relation $x^n - 1 = \prod_{d|n} \Phi_d(x)$.)

Bonus Exercise 4.3.4 (The ring of arithmetic functions)**.** In this exercise, you will explore an algebraic structure on the set $\mathcal{A}$ of arithmetic functions.

Given two arithmetic functions $f, g\colon \mathbb{Z}^+ \to \mathbb{C}$, their *convolution* is the function $(f * g)\colon \mathbb{Z}^+ \to \mathbb{C}$ defined by

$$(f * g)(n) := \sum_{d|n} f(d) g\left(\frac{n}{d}\right).$$

a) For an arithmetic function $f\colon \mathbb{Z}^+ \to \mathbb{C}$, we have previously defined $F(n) := F_f(n) = \sum_{d|n} f(d)$. What is $\mu * F_f$?
b) Show that convolution is commutative: for any two arithmetic functions f and g, one has $f * g = g * f$.
c) show that if f and g are multiplicative functions, then so is $f * g$.

d) Show that convolution is associative: for arithmetic functions f, g and h, one has $(f * g) * h = f * (g * h)$.

e) Define a function $I\colon \mathbb{Z}^+ \to \mathbb{C}$ via

$$I(n) := \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that for any arithmetic function f, one has $f * I = I * f = f$.

f) Given an arithmetic function f with $f(1) \neq 0$, define a new arithmetic function $f^{-1}\colon \mathbb{Z}^+ \to \mathbb{C}$ recursively: set $f^{-1} := \frac{1}{f(1)}$, and for $n > 1$ define

$$f^{-1}(n) := \frac{-1}{f(1)} \sum_{d|n, d<n} f\left(\frac{n}{d}\right) f^{-1}(d).$$

Show that $f * f^{-1} = I$.

The above steps show that the subset of $\mathcal{A}$ of functions f with $f(1) \neq 0$ is a *commutative group* under the convolution operation. However, there is a larger ring structure on $\mathcal{A}$ to consider.

g) For two arithmetic functions $f, g\colon \mathbb{Z}^+ \to \mathbb{C}$, define their sum $f + g\colon \mathbb{Z}^+ \to \mathbb{C}$ as their pointwise sum:

$$(f + g)(n) := f(n) + g(n).$$

Note that $f + g$ is an arithmetic function.

h) Show that $\mathcal{A}$ is a commutative group under the sum operation above.

i) Show that the distributive law holds for $+$ and $*$: more precisely, for arithmetic functions f, g and h, one has

$$f * (g + h) = f * g + f * h.$$

The conclusion is that $\mathcal{A}$ is a *commutative ring* under addition and convolution; it is called the **Dirichlet ring**. By the Möbius inversion formula, every element $f \in \mathcal{A}$ is a multiple of the Möbius function μ (see part a)). What other properties does $\mathcal{A}$ have as a ring?

3. Chapter 3: Quadratic Reciprocity and Quadratic Forms

3.1. **Quadratic residues.** Previously, we studied solutions to polynomials $f(x) \in \mathbb{Z}[x]$ modulo $m > 0$, and reduced it to studying solutions modulo prime powers p^e (by the CRT), and then reduced it to studying solutions modulo p (by Hensel's lemma). In §2.8, we proved Euler's criterion for n'th power residues, to determine when a polynomial of the form $f(x) = x^n - a \in \mathbb{Z}[x]$ has solutions modulo p.

Theorem. [NZM91, Theorem 2.37] *For prime $p \in \mathbb{Z}^+$ and $a \in \mathbb{Z}$ with $\gcd(a, p) = 1$, 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, it has $\gcd(n, p-1)$ *solutions.*

We ended §2.8 with a special case, known as THE "Euler's criterion."

Theorem. [NZM91, Corollary 2.38] *For any prime* $p \in \mathbb{Z}^+$ *and integer* $a \in \mathbb{Z}$ *with* $\gcd(a, p) = 1$*, the congruence*

$$x^2 \equiv a \pmod{p}$$

has a solution if and only if

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

If there is a solution, then there are two solutions if p *is odd, and one solution if* $p = 2$*.*

Proof. If we set $b := a^{\frac{p-1}{2}}$, then

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

(by Fermat's little theorem). Thus, b is a root of $x^2 - 1$ modulo p, which by [NZM91, Lemma 2.10] forces $b \equiv \pm 1 \pmod{p}$. □

In this chapter, we will study solutions to the polynomial $x^2 - a$ modulo p, reducing it to studying $x^2 - 2, x^2 + 1$ and $x^2 - q$ modulo p, where q is an odd prime. This will culminate in the classical theorem of **quadratic reciprocity** and its supplements. There are currently 345 documented proofs of quadratic reciprocity, see here.

Let's define quadratic residues and Legendre symbols.

Definition 3.1.1. For integers $m \in \mathbb{Z}^+$ and $a \in \mathbb{Z}$ such that $\gcd(a, m) = 1$, 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).

Quadratic residues mod m are simply squares mod m.

Example 3.1.1.

- 1 is a quadratic residue modulo *any* $m > 0$, since $x^2 - 1$ has root 1 mod m.
- Is 5 a QR modulo 7? 7 is small, so we can compute all the squares mod 7: $1^2 = 1, 2^2 = 4, 3^2 \equiv 2, 4^2 \equiv (-3)^2 = 3^2 \equiv 2, 5^2 \equiv (-2)^2 = 4, 6^2 \equiv 1$. These were all the congruence classes mod 7, so 5 is a QNR modulo 7.
- Is -1 a quadratic residue mod 101? Since $101 \equiv 1 \pmod{4}$, we know that -1 is a square modulo 101 – this is by [NZM91, Theorem 2.12], where $x^2 + 1$ has roots mod p iff $p = 2$ or $p \equiv 1 \pmod{4}$.

Remark. In general, if $a \equiv b \pmod{m}$ and a is a quadratic residue mod m, then so is b. Thus, we only check whether $a \in \mathbb{Z}$ is a quadratic residue *up to congruence class mod* m (e.g., -1 and 100 are the same mod 100 in our quadratic residue checks).

Definition 3.1.2. For odd prime $p \in \mathbb{Z}^+$, we define the **Legendre symbol of** a **modulo** p as follows:

$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{if } a \text{ is a } \textbf{quadratic residue} \text{ modulo } p, \\ -1 & \text{if } a \text{ is a } \textbf{quadratic nonresidue} \text{ modulo } p, \\ 0 & \text{if } p \mid a. \end{cases}$$

Example 3.1.2. Based on our previous examples:

- For all $m > 0$, $\left(\frac{1}{m}\right) = 1$.
- $\left(\frac{5}{7}\right) = -1$.
- $\left(\frac{-1}{101}\right) = 1$; more generally, for any prime $p = 2$ or $p \equiv 1 \pmod 4$, we have $\left(\frac{-1}{p}\right) = 1$. Similarly, for $p \equiv -1 \pmod 4$, one has $\left(\frac{-1}{p}\right) = -1$.

Here are some things to keep in mind when calculating Legendre symbols:

Theorem 3.1.1. [NZM91, Theorem 3.1] *For odd prime $p \in \mathbb{Z}^+$, one has 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 terms are equal to 0. Suppose then that $\gcd(a, p) = 1$. Then this follows from Euler's criterion:
   (a) If $x^2 - a$ has a root mod p, then $a^{\frac{p-1}{2}} \equiv 1 \pmod p$ by Euler's criterion, and $\left(\frac{a}{p}\right) = 1$ by definition (say b is a root of $x^2 - a$ mod p: then a is a square mod p, namely b^2).
   (b) If $x^2 - a$ has no roots mod p, then 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 $x^2 - 1$ mod p, which forces $a^{\frac{p-1}{2}} \equiv -1 \pmod p$ by [NZM91, Lemma 2.10].

The other parts follow from part 1.:

2. $$\left(\frac{a}{p}\right) \cdot \left(\frac{a}{p}\right) = a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} = (ab)^{\frac{p-1}{2}} = \left(\frac{ab}{p}\right).$$
3. Clear from a previous argument, but also from
$$\left(\frac{a}{p}\right) = a^{\frac{p-1}{2}} \equiv b^{\frac{p-1}{2}} = \left(\frac{b}{p}\right).$$
4. $$\left(\frac{a^2}{p}\right) = (a^2)^{\frac{p-1}{2}} = a^{p-1} \equiv 1.$$
   We also have
$$\left(\frac{a^2 b}{p}\right) = (a^2 b)^{\frac{p-1}{2}} = (a^2)^{\frac{p-1}{2}} \cdot b^{\frac{p-1}{2}} \equiv b^{\frac{p-1}{2}} = \left(\frac{b}{p}\right).$$

5. Second part follows from [NZM91, Theorem 2.12]. □

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 $\left(\frac{3}{11}\right) = 1$.
- $\left(\frac{10000000000000000000001}{5}\right) = \left(\frac{1}{5}\right) = 1$.
- $\left(\frac{96}{97}\right) = \left(\frac{-1}{97}\right) = -1$, since $97 \equiv -1 \pmod{4}$.

In the next section, we will prove quadratic reciprocity, and see how this simplifies Legendre symbol calculations.

Exercise 3.1.1 (Two times nonsquares makes a square)**.** Prove that for any odd prime $p \in \mathbb{Z}^+$, if $a, b \in \mathbb{Z}$ are not squares modulo p then their product ab is a square modulo p.

Exercise 3.1.2 (Primitive roots and quadratic residues)**.** Let $p \in \mathbb{Z}^+$ be an odd 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 using positive integers between** 1 **and** 18**, listed in increasing order.** (*Hint:* 2 is a primitive root modulo 19.)
c) How many quadratic residues are there modulo an odd prime $p \in \mathbb{Z}^+$? What about quadratic nonresidues?

Exercise 3.1.3 (When is 2 a square mod p?)**.** Here is a result which describes when 2 is a quadratic residue modulo odd primes.

Proposition. *For all primes $p > 2$, one has that*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

Prove the second equality.

Bonus Exercise 3.1.4 (A concise formula for the number of roots mod p)**.** Let $a \in \mathbb{Z}$ be an integer, and set $f(x) := x^2 - a$. Show that for any prime $p \in \mathbb{Z}^+$, the number of roots of $f(x)$ modulo p is $1 + \left(\frac{a}{p}\right)$.

3.2. **Quadratic reciprocity.** This section will focus on proving the classical theorem of quadratic reciprocity, which relates the Legendre symbols of two odd primes with respect to one another. As mentioned previously, there are 345 known 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] *If $p, q \in \mathbb{Z}^+$ are distinct odd primes, then*

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

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. For example, at first glance calculating $\left(\frac{3}{101}\right)$ may seem tough, but since $101 \equiv 1 \pmod 4$, we have by quadratic reciprocity that

$$\left(\frac{3}{101}\right) = \left(\frac{101}{3}\right) = \left(\frac{2}{3}\right) = -1;$$

thus, we know that 3 is not a square modulo 101.

Proof of quadratic reciprocity. This proof is credited to George Rousseau [Rou91]. This proof only uses the Chinese remainder theorem, Wilson's Theorem and Euler's Criterion.

In this proof, for $n \in \mathbb{Z}^+$ we will write $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$. Thus, our unit groups are written as $\mathbb{Z}_n^\times = (\mathbb{Z}/n\mathbb{Z})^\times$. We will sometimes use brackets, such as $[k]$, to denote a congruence class (where the modulus should be clear from context).

The plan of this proof is to split $\mathbb{Z}_{pq}^\times$ into two halves (one half being denoted H_1), where every element $[k] \in \mathbb{Z}_{pq}^\times$ is such that either $[k] \in H_1$ or $[-k] \in H_1$. We will also split $\mathbb{Z}_q^\times$ into halves in $\mathbb{Z}_{pq}^\times$ (one half being denoted H_2), and produce an equality which relates the product of elements in H_1 and H_2 (this is Equation (2)). Finally, we'll simplify these product and the Legendre symbols $\left(\frac{p}{q}\right)$ and $\left(\frac{q}{p}\right)$ will appear, and we'll deduce quadratic reciprocity from this.

Let us define our halves,

$$H_1 := \left\{[k] \in \mathbb{Z}_{pq}^\times : 1 \le k < \frac{pq}{2}\right\}$$

and

$$H_2 := \left\{(a,b) \in \mathbb{Z}_p^\times \times \mathbb{Z}_q^\times : 1 \le b < \frac{q}{2}\right\}.$$

By the CRT, we have a natural group isomorphism

$$\Phi\colon \mathbb{Z}_{pq}^\times \xrightarrow{\sim} \mathbb{Z}_p^\times \times \mathbb{Z}_q^\times.$$

Thus, for all $(a,b) \in \mathbb{Z}_p^\times \times \mathbb{Z}_q^\times$, there exists a *unique* $1 \le k < pq$ with $k \equiv a \pmod p$ and $k \equiv b \pmod q$, i.e., $\Phi([k]) = (a,b)$.

(1) If $1 \le k < \frac{pq}{2}$, then
$$\Phi([k]) = (a,b)$$
and there's nothing to note.

(2) If $\frac{pq}{2} \le k < pq$, then $1 \le pq - k < \frac{pq}{2}$, and $\Phi([pq-k]) = \Phi([-k]) = -\Phi([k]) = -(a,b)$. (Note that $[pq-k]$ is in H_1.)

Thus, for every element $(a,b) \in H_2$, there exists unique $1 \leq k < pq$ such that either $k \in H_1$ or $pq-k \in H_1$. Therefore, **as a product of elements in** $(\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$**,** we have

$$\prod_{(a,b)\in H_2} (a,b) = \epsilon \cdot \prod_{k \in H_1} (k,k), \tag{2}$$

where $\epsilon \in \{\pm 1\}$ (ϵ is a product of the negative signs for the pairs $(a,b) \in H_2$ which which correspond to $1 \leq k < pq$ with $pq-k \in H_1$); keep in mind that in these pairs (a,b) and (k,k), the first coordinate is mod p, and the second coordinate is mod q. Thus, we've split up H_2 in $\mathbb{Z}_{pq}^\times$ using H_1.

We will simplify each side of (2) individually, and then equate them to deduce quadratic reciprocity. We'll first simplify the left hand side.

Before we begin, 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. We calculate

$$\prod_{(a,b)\in H_2} (a,b) = \prod_{\substack{1\leq a<p,\\ 1\leq b<\frac{q}{2}}} (a,b) = ((p-1)!^Q, Q!^{p-1}),$$

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);

$$= ((p-1)!^Q, Q!^{2P}),$$

by definition of $P = \frac{p-1}{2}$;

$$= ((-1)^Q, Q!^{2P}),$$

by Wilson's theorem modulo p;

$$= ((-1)^Q, ((q-1)!(-1)^Q)^P)),$$

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 homework problem);

$$= ((-1)^Q, ((-1)^{Q+1})^P),$$

by Wilson's theorem modulo q;

$$= ((-1)^Q, (-1)^{PQ+P}).$$

Thus, we've shown that

$$\prod_{(a,b)\in H_2} (a,b) \equiv ((-1)^Q, (-1)^{PQ+P}). \tag{3}$$

Next we will simplify the right hand side of (2), 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

$$\prod_{k\in H_1} k = \prod_{\substack{1\le k<\frac{pq}{2}:\\ \gcd(k,pq)=1}} k = \left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ p\mid k}} k\right)\cdot\left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1},$$

the latter equality coming from the fact that we can divide by the integers divisible by q, to be left with a product of integers coprime to q;

$$= \left(\prod_{0<k<p} k\cdot\prod_{p<k<2p} k\cdot\prod_{2p<k<3p} k\cdots\prod_{(Q-1)p<k<Qp} k\cdot\prod_{Qp<k<\frac{pq}{2}} k\right)\cdot\left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1},$$

where we've explicitly written out our product of integers coprime to p (which are the non-multiples of p) – note that the blue product might have less terms involved, and only goes up to $\frac{pq}{2}$);

$$= \left(\underbrace{(p-1)!\cdot(p-1)!\cdot(p-1)!\cdots(p-1)!}_{Q \text{ times}}\cdot\prod_{Qp<k<\frac{pq}{2}} k\right)\cdot\left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1},$$

since each product $\prod_{jp<k<(j+1)p} k$ in the numerator for $0\le j<Q$ is congruent to $(p-1)!$ mod p;

$$\left((p-1)!^Q\cdot\prod_{Qp<k<\frac{pq}{2}} k\right)\cdot\left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1} = \left((p-1)!^Q\cdot P!\right)\cdot\left(\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k\right)^{-1},$$

which follows from the fact that every integer $Qp<k<\frac{pq}{2}$ has the form $k=\ell+Qp$, with $1\le\ell\le\frac{p-1}{2}$ – this is also a HW problem;

$$= \frac{(p-1)!^Q\cdot P!}{q\cdot 2q\cdot 3q\cdots Pq},$$

since $\prod_{\substack{1\le k<\frac{pq}{2}:\\ q\mid k}} k \equiv q\cdot 2q\cdot 3q\cdots Pq \mod p$ – part of this is also a HW problem;

$$= \frac{(p-1)!^Q\cdot P!}{q^P\cdot P!} = \frac{(p-1)!^Q}{q^P} \equiv \frac{(-1)^Q}{q^P},$$

by Wilson's theorem modulo p;

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

by Euler's criterion for the Legendre symbol, see [NZM91, Theorem 3.1];

$$= (-1)^Q \cdot \left(\frac{q}{p}\right).$$

Thus, we conclude by our 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 (2): this equation

$$\prod_{(a,b) \in H_2} (a,b) = \epsilon \cdot \prod_{k \in H_1} (k,k)$$

now becomes

$$((-1)^Q, (-1)^{PQ+P}) = \left(\epsilon \cdot (-1)^Q \cdot \left(\frac{q}{p}\right), \epsilon \cdot (-1)^P \cdot \left(\frac{p}{q}\right)\right).$$

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 $1 - \epsilon \cdot \left(\frac{q}{p}\right) = 0$ (by e.g. [NZM91, Theorem 1.1.(5)]) and thus

$$1 = \epsilon \cdot \left(\frac{q}{p}\right),$$

so that

$$\epsilon = \left(\frac{q}{p}\right).$$

A similar argument on the second coordinate equality shows that

$$(-1)^{PQ+P} = \epsilon \cdot (-1)^P \cdot \left(\frac{p}{q}\right),$$

so 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. □

Example 3.2.1. Does the polynomial $f(x) := x^2 - 7$ have a root over $\mathbb{Z}/43\mathbb{Z}$? We can use Legendre symbols to determine this.

By quadratic reciprocity, we have (since $7 \equiv 3 \pmod 4$ and $43 \equiv 3 \pmod 4$)

$$\left(\frac{7}{43}\right) = -\left(\frac{43}{7}\right).$$

Then we compute

$$-\left(\frac{43}{7}\right) = -\left(\frac{1}{7}\right) = -1.$$

We conclude that 7 is **not** a square modulo 43; thus, $x^2 - 7$ has no roots in $\mathbb{Z}/43\mathbb{Z}$.

We almost have a formula to compute $\left(\frac{a}{p}\right)$ for any integer $a \in \mathbb{Z}$; when a is a product of odd primes with, we can invoke the formula for quadratic reciprocity (and $\left(\frac{-1}{p}\right)$ if $a < 0$) to compute this symbol. However, what do we do when a is even? We have the following result, often referred to as a "supplementary law for quadratic reciprocity."

Theorem 3.2.2. [NZM91, Theorem 3.3] *For odd prime $p \in \mathbb{Z}^+$, one has*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

We won't prove this law.

For convenience, we list one more "supplementary law", which we saw in §3.1.

Theorem. [NZM91, Theorem 3.1.5] *One has for odd prime $p \in \mathbb{Z}^+$ 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's 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 *not* a quadratic residue modulo 29.

In the next section, we will define a generalization of the Legendre symbol (called the *Jacobi symbol*) which can make our quadratic residue checks even simpler.

Exercise 3.2.1. This exercise fills in some steps from our proof of quadratic reciprocity. Following the notation in our proof, let $p, q > 2$ be 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 $\frac{pq}{2}$.

Exercise 3.2.2. Using quadratic reciprocity and/or its supplemental laws, prove that for any odd prime $p \in \mathbb{Z}^+$, one has

$$\left(\frac{-2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 3 \pmod 8, \\ -1 & \text{if } p \equiv 5, 7 \pmod 8. \end{cases}$$

Exercise 3.2.3.

a) List the squares modulo 7, and then the non-squares.

b) Determine all primes $p \in \mathbb{Z}^+$ such that $x^2 - 7$ has a root modulo p. (*Hint:* your final answer should include several different congruency conditions on p.)

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.** 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 don't have to factorize* a.

For example, we already know that if $n \in \mathbb{Z}^+$ has a prime factorization $n = \prod_{i=1}^r p_i^{e_i}$, then

$$\left(\frac{n}{p}\right) = \prod_{i=1}^r \left(\frac{p_i}{p}\right)^{e_i}.$$

However, if we'd like to compute e.g. $\left(\frac{100101}{5001}\right)$, how do we even begin to factor 100101?

Definition 3.3.1. Let $m \in \mathbb{Z}^+$ be an odd number; write its factorization as $m = \prod_{i=1}^r p_i^{e_i}$. Then for $a \in \mathbb{Z}$, we define the **Jacobi symbol of a mod m** as

$$\left(\frac{a}{m}\right) := \prod_{i=1}^r \left(\frac{a}{p_i}\right)^{e_i}$$

(So this factorizes in the denominator). We also define $\left(\frac{a}{1}\right) := 1$.

Note that when m is prime, the Jacobi symbol is simply the Legendre symbol.

Example 3.3.1.

- $\left(\frac{7}{15}\right) = \left(\frac{7}{3}\right) \cdot \left(\frac{7}{5}\right) = \left(\frac{1}{3}\right) \cdot \left(\frac{2}{5}\right) = 1 \cdot (-1) = -1$.
- $\left(\frac{3}{35}\right) = \left(\frac{3}{5}\right) \cdot \left(\frac{3}{7}\right) = -1 \cdot (-1) = 1$.
- $\left(\frac{5}{375}\right) = \left(\frac{5}{3}\right) \cdot \left(\frac{5}{5}\right)^3 = 0$.

Remark. 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 tell us this if the numerator m is composite.** For example, $\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right) \cdot \left(\frac{2}{5}\right) = (-1) \cdot (-1) = 1$, but 2 is not a square modulo 15.

So what is the utility of the Jacobi symbol? As we'll see, it satisfies similar laws that the Legendre symbol follows – *without having to factorize the numerator.*

Theorem 3.3.1. [NZM91, Theorem 3.6] *For odd numbers $m, n > 1$, one has for all $a, b \in \mathbb{Z}$ the following:*

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. *consequently, 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 (completely multiplicative in the denominator), as well as the corresponding properties for the Legendre symbol. For example, let us factorize m as $m = \prod_{i=1}^r p_i^{e_i}$. We'll prove 1.: we see that

$$\left(\frac{ab}{m}\right) := \prod_{i=1}^t \left(\frac{ab}{p_i}\right)^{e_i};$$

however, each $\left(\frac{ab}{p_i}\right) = \left(\frac{a}{p_i}\right) \cdot \left(\frac{b}{p_i}\right)$ by properties of the Legendre symbol (see [NZM91, Theorem 3.1]), so this becomes

$$\prod_{i=1}^{t} \left(\frac{ab}{p_i}\right)^{e_i} = \prod_{i=1}^{t} \left(\frac{a}{p_i}\right)^{e_i} \cdot \left(\frac{b}{p_i}\right)^{e_i} = \prod_{i=1}^{t} \left(\frac{a}{p_i}\right)^{e_i} \cdot \prod_{i=1}^{t} \left(\frac{b}{p_i}\right)^{e_i} =: \left(\frac{a}{m}\right) \cdot \left(\frac{b}{m}\right).$$

We'll also prove 3. Since $a \equiv b \pmod{m}$, for each $1 \leq i \leq r$ we have $a \equiv b \pmod{p_i}$, and so the Legendre symbol $\left(\frac{a}{p_i}\right) = \left(\frac{b}{p_i}\right)$. Therefore,

$$\left(\frac{a}{m}\right) := \prod_{i=1}^{t} \left(\frac{a}{p_i}\right)^{e_i} = \prod_{i=1}^{t} \left(\frac{b}{p_i}\right)^{e_i} =: \left(\frac{b}{m}\right).$$

The rest are proven in the book. □

What's remarkable is that the Jacobi symbol also satisfies the "quadratic reciprocity law" and its supplements.

Theorem 3.3.2. [NZM91, Theorem 3.7 and 3.8] *If $m, n > 1$ are odd numbers and $\gcd(m, n) = 1$, then*

$$\left(\frac{m}{n}\right) = \left(\frac{n}{m}\right) \cdot (-1)^{\frac{m-1}{2} \cdot \frac{n-1}{2}} = \begin{cases} \left(\frac{n}{m}\right) & \text{if } m \equiv 1 \pmod 4 \text{ } \textbf{\textit{or}} \text{ } n \equiv 1 \pmod 4 \\ -\left(\frac{n}{m}\right) & \text{if } m \equiv 3 \pmod 4 \text{ } \textbf{\textit{and}} \text{ } n \equiv 3 \pmod 4. \end{cases}$$

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'll forgo the proofs of the quadratic reciprocity laws for the Jacobi symbol – they're not too pleasant.

Example 3.3.2. We'd like to know 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'd need to factor 851 first. However, with the Jacobi symbol, we can do the following (since both

are odd numbers):

$$\begin{aligned}\left(\frac{851}{1013}\right) &= -\left(\frac{1013}{851}\right) \text{ Jacobi QR, since } \gcd(1013, 851) = 1 \text{ – we know this since 1013 is prime!}\\ &= \left(\frac{162}{851}\right)\\ &= \left(\frac{2}{851}\right)\cdot\left(\frac{81}{851}\right)\\ &= -\left(\frac{81}{851}\right) \text{ supplementary Jacobi QR: } 851 \equiv 3 \pmod 8\\ &= -\left(\frac{9}{851}\right)^2 \text{ since } 3 \nmid 851, \text{as 3 doesn't divide the sum of digits of 851}\\ &= -1.\end{aligned}$$

We thus conclude that $x^2 - 851$ has no roots modulo 1013.

Exercise 3.3.1. Compute the following Legendre/Jacobi symbols.

a) $\left(\frac{51}{71}\right)$;
b) $\left(\frac{-35}{97}\right)$;
c) $\left(\frac{1001}{9907}\right)$, where 9907 is prime.

Exercise 3.3.2. Determine whether the polynomial $x^4 - 36$ has a root modulo the prime $p = 5077$.

5. Chapter 5: Some Diophantine Equations

Introduction to Chapter 5. Up to this point, we have answered several questions about solutions to polynomials **modulo** m**.** Questions about solutions in modular arithmetic can be hard, but always "finite" in the sense that there are only finitely many solutions to any congruence in any number of variables. Such questions about solutions to modular congruences are called *local* questions.

Now, we turn our attention towards *global* problems. For the remainder of this class, we are interested in finding solutions not just to *modular* equations, but **Diophantine** equations – so we're looking for solutions in $\mathbb{Z}$, not a fixed $\mathbb{Z}/m\mathbb{Z}$!

Definition 5.0.1. Given $n \in \mathbb{Z}^+$ and a set X, we use $X^n := \underbrace{X \times X \times \ldots \times X}_{n \text{ times}}$. A **Diophantine equation** is a polynomial equation in n variables, for example $f(x_1, x_2, \ldots, x_n) \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$.

An **integral solution** is a point $(a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n$ with

$$f(a_1, a_2, \ldots, a_n) = 0.$$

Example 5.0.1. Here are some examples of Diophantine equations:

- Define $f(x, y) \in \mathbb{Z}[x, y]$ by

$$f(x, y) := x^2 + y^2 - 1.$$

What are its integral solutions? We're looking for $(a, b) \in \mathbb{Z}^2$ with

$$f(a, b) = 0,$$

i.e.,

$$a^2 + b^2 = 1.$$

We have that $(\pm 1, 0)$ and $(0, \pm 1)$ are integral solutions (one can show that these are the only integral solutions).

- Let
$$g(x, y) := x - 1.$$
One can show there are infinitely many integral solutions: they are of the form $(1, b)$ where $b \in \mathbb{Z}$.
- 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 $a_i = \pm 1$ in the correct way.
- Let
$$k(x, y) := y^2 - (x^3 - x).$$
An integral solution (a, b) to $k(x, y)$ satisfies
$$b^2 = a^3 - a.$$
We have solutions $(0, 0), (1, 0)$ and $(-1, 0)$.

One more definition, to account for the different types of solutions.

Definition 5.0.2. Let us recall that

- $\mathbb{Z}$ is the set of integers;
- $\mathbb{Q}$ is the set of *rational numbers;*
- $\mathbb{R}$ is the set of *real numbers* (such as $\pi, e^1, \ln(2)$);
- $\mathbb{C}$ is the set of *complex numbers* (such as i, $\pi + 3i, e^{2\pi i/7}$).

(Ask class who knows that $\mathbb{R}$ is, then what $\mathbb{C}$ is.) One has $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

Let R be one of the rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$. Then given a polynomial $f(x_1, x_2, \ldots, x_n) \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$, we say a solution $(a_1, a_2, \ldots, a_n)$ is **over** R if $a_1, a_2, \ldots, a_n \in R$, i.e., $(a_1, a_2, \ldots, a_n) \in R^n$.

More precisely:

- Solutions over $\mathbb{Z}$ are *integral* solutions;
- solutions over $\mathbb{Q}$ are *rational* solutions;
- solutions over $\mathbb{R}$ are *real* solutions;
- solutions over $\mathbb{C}$ are *complex* solutions.

It might not be surprising that an integral solution to a polynomial might not exist, but it can still have real/complex solutions, such as for the polynomial $\ell(x, y) := x^2 + y^2 + 1$.

Example 5.0.2. Re-investigating our previous examples for other solutions:

- $$f(x,y) := x^2 + y^2 - 1$$
  defines an equation for the *unit circle.* Its solutions over $\mathbb{R}$ draws a picture: (picture of unit circle). This has an infinite amount of rational solutions (as we'll see in §5.3).
- $$g(x,y) := x - 1$$
  draws a vertical line: (picture of vertical line at $x = 1$).
- $$k(x,y) := y^2 - (x^3 - x)$$
  can be drawn like this: (draw picture of elliptic curve, two components, x-intercepts at $x = -1, 0, 1$). This is an example of an **elliptic curve.**

Each of the pictures above are examples of **plane curves over** $\mathbb{R}$ (solutions to polynomials in two variables).

Definition 5.0.3. Given a polynomial in two variables, $f(x,y) \in \mathbb{Z}[x,y]$, we say that f defines a **plane curve over** $\mathbb{R}$**.** Written as C, we use notation

$$C : f(x,y) = 0$$

to say that C is the *graph* of $f(x,y)$ over $\mathbb{R}$. Thus, the curve C is the set of real solutions.

Example 5.0.3. To practice notation:

- We have the unit circle
$$S : x^2 + y^2 = 1.$$
- We have a line
$$L : y = 2x + 7.$$
- We have an elliptic curve
$$E : y^2 = x^3 + x + 1.$$

(Try and draw these! You can also use Desmos to help draw them.)

How many solutions are there? A polynomial $f \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$, when reduced modulo m, has a *finite* amount of solutions modulo m. However, it can have infinitely many integral solutions. Furthermore, there is no "one size fits all" algorithm to determine whether any such f has a solution or not ("undecidability" and Hilbert's 10th problem). Similarly, there is no known algorithm to determining rational points on general Diophantine equations.

In this chapter, we'll study several different kinds of Diophantine equations (including Fermat equations, plane curves and specifically elliptic curves) and study their integral and rational points. Sometimes, we'll be able to use our "local" techniques (modular arithmetic) to say something about our "global" solutions (integral and rational).

For the next three exercises, it is recommended that you employ the "global-to-local" technique: if you have an integral solution to a Diophantine equation, then you have a solution to an equation *modulo any* $m > 0$.

Exercise 5.0.1. Show that the Diophantine equation

$$x^2 + y^2 = 9z + 6$$

has no integral solutions.

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.

5.1. **The equation $ax+by = c$.** In this section, we will study the "simplest" nontrivial case of Diophantine equations: namely, where $f(x, y)$ is a **line:**

$$L : ax + by = c$$

where $a, b, c \in \mathbb{Z}$.

Example 5.1.1. Some lines include:

- Consider
$$L_1 : x + y = 1.$$
What are the solutions to f over $\mathbb{Z}$? We have
$$y = 1 - x,$$
so infinitely many integral solutions of the form $(n, 1 - n)$.
- Consider
$$L_2 : 2x + 4y = 5.$$
This has *no integral solutions:* if such a solution $(a, b) \in \mathbb{Z}^2$ exists, then
$$2a + 4b = 5,$$
but reducing mod 2 shows that $0 \equiv 1 \pmod 2$, contradiction. (This is an example of using modular arithmetic to say something about integral solutions.)

What's nice is that decidability of solutions to linear equations is completely known, by the "linear Diophantine theorem". This one is for lines in $\mathbb{R}^2$; §5.2 covers lines in $\mathbb{R}^n$, but we won't go over that in this class.

Theorem 5.1.1. [NZM91, Theorem 5.1] *Fix integers $a, b, c \in \mathbb{Z}$ where $a \neq 0$ or $b \neq 0$; set $g := \gcd(a, b)$. Then the line*

$$L : ax + by = c$$

has an integral solution if and only if $g \mid c$. When $g \mid c$, L 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 $\left(x_1 + k \cdot \frac{b}{g}, y_1 - k \cdot \frac{a}{g}\right)$ where $k \in \mathbb{Z}$.

Remark. This theorem is identical to the "linear congruence theorem" from §2.2 ([NZM91, Theorem 2.17]), where $ax \equiv c \pmod b$ has a solution if and only if $\gcd(a, b) \mid c$.

Proof. $\Rightarrow$: suppose that $(x_1, y_1) \in \mathbb{Z}^2$ satisfies

$$ax_1 + by_1 = c.$$

Then clearly g divides the left hand side, thus $g \mid c$.

$\Leftarrow$: suppose that $g = \gcd(a, b) \mid c$; let us write $c = gk$ for some $k \in \mathbb{Z}$ (precisely, $k = \frac{c}{g}$). We know that $\gcd(a, b)$ can be written as a $\mathbb{Z}$-linear combination of a and b: let us write

$$ax_0 + by_0 = g$$

for some $x_0, y_0 \in \mathbb{Z}$. Multiply both sides by $k = \frac{c}{g}$ to get

$$a(kx_0) + b(ky_0) = gk = c,$$

so that (kx_0, ky_0) is an integral solution to 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 by g, we have

$$\frac{a}{g}(x_1 - x_2) = -\frac{b}{g} \cdot (y_1 - y_2) \tag{4}$$

(star this equation). Thus, $\frac{a}{g} \mid \frac{b}{g} \cdot (y_1 - y_2)$; since $\frac{a}{g}$ and $\frac{b}{g}$ are coprime, this implies that $\frac{a}{g} \mid (y_1 - y_2)$, and thus

$$y_1 - y_2 = k \cdot \frac{a}{g}$$

for some $k \in \mathbb{Z}$, i.e.,

$$y_2 = y_1 - k \cdot \frac{a}{g}.$$

On the other hand, plugging this into (4), we have

$$\frac{a}{g}(x_1 - x_2) = -\frac{b}{g} \cdot k \cdot \frac{a}{g},$$

so that

$$x_1 - x_2 = -k \cdot \frac{b}{g},$$

i.e.,

$$x_2 = x_1 + k \cdot \frac{b}{g}.$$

This concludes our proof. $\square$

A lingering question is: given a line $L : ax + by = c$ where $a, b, c \in \mathbb{Z}$ and $a \neq 0$ or $b \neq 0$, if $\gcd(a, b) \mid c$, then how do we find our "first solution" $(x_1, y_1) \in \mathbb{Z}^2$? Based on our proof, we can do the following:

1. Find a solution $(x_0, y_0) \in \mathbb{Z}^2$ to the "GCD line"
$$ax + by = \gcd(a, b);$$
2. multiply both sides by $\frac{c}{\gcd(a,b)}$ and 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 is
$$(x_1, y_1) := \left(x_0 \cdot \frac{c}{\gcd(a, b)}, y_0 \cdot \frac{c}{\gcd(a, b)}\right),$$
where (x_0, y_0) is a solution to the "GCD line" $ax + by = \gcd(a, b)$; therefore, by the theorem all other solutions have the form
$$\left(x_0 \cdot \frac{c}{\gcd(a, b)} + k \cdot \frac{b}{\gcd(a, b)}, y_0 \cdot \frac{c}{\gcd(a, b)} - k \cdot \frac{a}{\gcd(a, b)}\right),$$
where $k \in \mathbb{Z}$ (Keep track of this formula! Or the simplified one from the theorem). We can use e.g. the Euclidean or Blankinship's algorithm to determine the "GCD solution" (x_0, y_0).

Example 5.1.2. We'd like to determine all integral solutions to the line
$$6x + 9y = 21,$$
if they exist. Here, $a = 6$ and $b = 9$, so $\gcd(a, b) = 3$, and since $3 \mid 21$ we conclude there are infinitely many solutions.

To characterize these solutions, we need to find one solution. We'll first find a solution to the "GCD line"
$$6x + 9y = 3.$$
Since a, b are small, it's possible to just "see" what a solution can be. We see that $x_0 = 2$ and $y_0 = -1$ works. Therefore, a solution to the original line $L : 6x + 9y = 21$ is
$$(x_1, y_1) = \left(x_0 \cdot \frac{c}{\gcd(a, b)}, y_0 \cdot \frac{c}{\gcd(a, b)}\right) = (2 \cdot 7, -1 \cdot 7) = (14, -7).$$
(On your own HW or exams, double-check that this is a solution!) Therefore, every other solution has the form
$$(x_2, y_2) = \left(x_1 + k \cdot \frac{b}{g}, y_1 - k \cdot \frac{a}{g}\right) = (14 + k \cdot 3, -7 - k \cdot 2) = (14 + 3k, -7 - 2k)$$
for any $k \in \mathbb{Z}$.

Example 5.1.3. How many integral points does the line
$$L : 216x + 135y = 100$$
have? We need to compute $\gcd(135, 216)$.

We'll use Blankinship's algorithm: we compute that

$$\left[\begin{array}{cc|cc} 216 & & 1 & 0 \\ 135 & & 0 & 1 \end{array}\right] \xrightarrow{A\mapsto A-B\cdot 1} \left[\begin{array}{cc|cc} 81 & & 1 & -1 \\ 135 & & 0 & 1 \end{array}\right]$$
$$\xrightarrow{B\mapsto B-A\cdot 1} \left[\begin{array}{cc|cc} 81 & & 1 & -1 \\ 54 & & -1 & 2 \end{array}\right]$$
$$\xrightarrow{A\mapsto A-B\cdot 1} \left[\begin{array}{cc|cc} 27 & & 2 & -3 \\ 54 & & -1 & 3 \end{array}\right].$$

We deduce that $\gcd(216, 135) = 27 = 3^3$. Since $3 \nmid 100$, we conclude that this line has no integral points.

Exercise 5.1.1.
Show that the line

$$L : ax + by = c$$

has an integral point if and only if for any integer $n \in \mathbb{Z}$, the line

$$L_n : ax + by = na + c$$

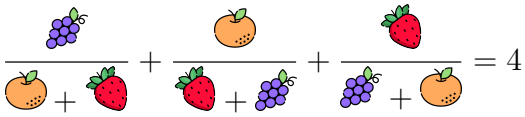
has an integral point. Briefly argue that this still holds if we replace na with nb in L_n.

Remark: Thus, in determining whether a "Diophantine line" has an integral point, one can reduce c either modulo a or b if it makes the problem simpler – note, however, that this changes your set of solutions (but in a predictable way).

Exercise 5.1.2. Determine whether the following Diophantine equations have integral solutions. If they do, give a complete description of their solutions (don't forget to show your work for calculating GCD's of larger numbers).

a) $10x - 7y = 17$.
b) $903x + 731y = 60$.
c) $mx + (m+1)y = 10$, where $m > 1$ is a fixed positive integer.
d) $(n-1)x + (n+1)y = 4573$, where $n > 1$ is a fixed odd integer.

Bonus Exercise 5.1.3. $> 99\%$ of people can't solve this! Find all  ,  ,  $\in \mathbb{Z}^+$ with



(see e.g. https://mathoverflow.net/questions/227713/estimating-the-size-of-solutions-of-a-diophantine-equation).

Bonus Exercise 5.1.4. $99.\overline{9}\%$ of people cannot solve this one! For  $\in \mathbb{Z}$ with  ≥ 3, find all  ,  ,  $\in \mathbb{Z}^+$ with

5.3. **Pythagorean triangles.** A special class of Diophantine equations are the **Fermat equations**

$$F_n : x^n + y^n = z^n.$$

Do we know when such equations have integral solutions? When $n = 1$, this is a linear equation and it's easy to create solutions. When $n \geq 3$, this has no "nontrivial" solutions (where $abc = 0$) – this is known as **Fermat's last theorem,** which is one of the most important results in mathematics (and it wasn't proven for at least 300 years). We'll talk more about it after we learn about elliptic curves.

What about $n = 2$? As it turns out, the equation

$$F_2 : x^2 + y^2 = z^2$$

has infinitely many integral solutions. We'll focus on this case today.

Given a right triangle with side lengths $a, b, c \in \mathbb{R}$ (draw one with side lengths a, b, c), we always have by the Pythagorean theorem that

$$a^2 + b^2 = c^2.$$

So it's easy to construct points on F_2. But do we know examples of *integral* side lengths that would give solutions to F_2? For example, $(3, 4, 5)$, or $(5, 12, 13)$, etc. We call integral solutions $(a, b, c) \in \mathbb{Z}^3$ where $a, b, c > 0$ **Pythagorean triples.**

Question: like with Diophantine lines, can we create infinitely many integral solutions from one solutions? **Yes:** if $(a, b, c) \in \mathbb{Z}^3$ satisfies

$$a^2 + b^2 = c^2,$$

then for any $k \in \mathbb{Z}$ one also has

$$(ka)^2 + (kb)^2 = (kc)^2,$$

so that (ka, kb, kc) is also a solution. For example, the solution $(3, 4, 5)$ implies other "multiple" solutions like $(6, 8, 10), (9, 12, 15)$, etc. All such multiple solutions create *similar* triangles (right triangles with equal angles).

We've answered our original question! However, we can ask a new one. Solutions like (ka, kb, kc) are "redundant" in a sense, and we'd like to know "minimal" solutions for each class of similar triangles, such as $(3, 4, 5)$. We notice that $(3, 4, 5)$, unlike the solutions $(3k, 4k, 5k)$, is that its coordinates are *pairwise coprime,* i.e., $\gcd(3, 4, 5) = 1$. We will call such a solution to F_2 a **primitive solution.** It's clear that a primitive solution is not a "multiple" of another solution.

New question: how many primitive solutions are there to F_2? This has a complete solution.

Theorem 5.3.1. [NZM91, Theorem 5.5] *The* ***positive*** *primitive solutions to*

$$F_2 : x^2 + y^2 = z^2$$

with y ***even*** *are precisely*

$$x = r^2 - s^s, y = 2rs, z = r^2 + s^2$$

where $r, s \in \mathbb{Z}$ *with* $r > s > 0$, r *and* s *have opposite parity and* $\gcd(r, s) = 1$.

Remark. Let's the explain the hypotheses in this theorem:

- x, y, z can be taken as positive: this is because each term is squared in the equation (keep in mind that we're eliminating redundancy in our choice of solutions, just as with the primitive assumption – we're trying to look at the most interesting solutions).
- y is even: first, if x and y are both odd, then $x^2 \equiv y^2 \equiv 1 \pmod 4$. Thus, $x^2 + y^2 \equiv 2 \pmod 4$, forcing $z^2 \equiv 2 \pmod 4$, which is impossible since 2 is not a square mod 4. Thus, they have opposite parity, and since the equation is symmetric in x and y, we can assume ("without loss of generality") that y the odd one.

We'll need a lemma before proving this.

Lemma 5.3.2. *If $u, v \in \mathbb{Z}^+$ are coprime and 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 must show that if a prime power $p^e \parallel u$, then e is even (symmetric in u and v, so just prove it for u). We know that $p^e \mid uv = a^2$; however, since $\gcd(u, v) = 1$, we know that $p \nmid v$, and thus $p^e \parallel a^2$. This forces e to be even. □

Proof of Theorem. (Here, we'll abuse notation and assume that (x, y, z) denotes our solution to $x^2 + y^2 = z^2$.)

Assume (x, y, z) is a positive primitive solution to $x^2+y^2 = z^2$ (so $x, y, z > 0$), with y even. Then $x \equiv z \pmod 2$, so they have the same parity: since x and y have different parity, x and z must be odd. Thus, $z + x$ and $z - x$ are even. Thus,

$$x^2 + y^2 = z^2$$

implies that

$$y^2 = z^2 - x^2 = (z + x) \cdot (z - x),$$

so that

$$\left(\frac{y}{2}\right)^2 = \left(\frac{z+x}{2}\right) \cdot \left(\frac{z-x}{2}\right).$$

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 $d \mid \gcd(x, z)$. But since $\gcd(x, y, z) = 1$ and $x^2 + y^2 = z^2$, we know that $\gcd(x, z) = 1$. Thus, $\frac{z+x}{2}$ and $\frac{z-x}{2}$ are also coprime. Since they are coprime and their product $\left(\frac{y}{2}\right)^2$ is a perfect square, they are both also perfect squares by the Lemma. We can write

$$\frac{z+x}{2} = r^2$$

and

$$\frac{z-x}{2} = s^2$$

for some $r, s \in \mathbb{Z}^+$.

Since $\frac{z+x}{2}$ and $\frac{z+x}{2}$ are coprime, so are r and s. Also, since $z+x > z-x > 0$, we have $r > s$. Furthermore, r and s have opposite parity: this is because e.g. $r + s \equiv r^2 + s^2 = z \equiv 1 \pmod 2$. Finally, directly adding the two squares shows that $r^2 + s^2 = z$. □

There's an interesting idea in this proof that generalizes to some other Diophantine equations: factoring solutions over the real or complex numbers. Here, we turned

$$x^2 + y^2 = z^2$$

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 this implies $(x, y) = (3, \pm 5)$ (see §9.9). This general technique/philosophy of lifting integral solutions to solutions over larger rings R is extremely useful, and is used in e.g. Fermat's last theorem.

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.

Exercise 5.3.2 (Prime difference of squares)**.** Determine all primes $p \in \mathbb{Z}^+$ such that the equation

$$x^2 - y^2 = p$$

has integral solutions.

5.6. **Rational points on curves.** (finish §5.3, last thoughts) In this section, we will begin to study the question of finding **rational** points on Diophantine plane curves. While this is "easier" than finding integral points, it can still be a difficult problem – in fact, much of modern number theory is interested in finding rational points to Diophantine equations. This area of number theory is called **arithmetic geometry.**

Recall these definitions from last week.

Definition 5.6.1. A **plane curve,** or **algebraic curve,** is the set C of points $(x, y) \in \mathbb{R}^2$ that are solutions to a fixed polynomial $f(x, y) \in \mathbb{R}[x, y]$: that is, C is defined by the equation

$$f(x, y) = 0.$$

We will write $C(\mathbb{R})$ for the set of *real* solutions $(x, y) \in \mathbb{R}^2$. We will also write $C(\mathbb{Z})$ for the *integral* solutions, $C(\mathbb{Q})$ for the *rational* solutions and $C(\mathbb{C})$ for the *complex* solutions.

Remark. Our polynomials $f(x, y)$ are often defined over $\mathbb{Z}$, i.e., $f(x, y) \in \mathbb{Z}[x, y]$. Sometimes, we'll also study $f(x, y) \in \mathbb{Q}[x, y]$.

Example 5.6.1.

- The curve
$$C_1 : x^2 + y^2 = 1$$
is the unit circle centered at the origin $(0, 0)$. (Draw a picture.) We have $C(\mathbb{Z}) = \{(0, 1), (1, 0), (-1, 0), (0, -1)\}$. By our last class on Pythagorean triples, we know that $\#C(\mathbb{Q}) = \infty$.
- The curve
$$C_2 : x^2 + y^2 = -1$$
has no real points: $C(\mathbb{R}) = \emptyset$; thus, $C(\mathbb{Z}) = C(\mathbb{Q}) = \emptyset$. However, $C(\mathbb{C})$ is an infinite set.
- The curve
$$E : y^2 = x^3 - 2x + 2$$
is an elliptic curve. (Draw it: horseshoe, $(1, 1), (1, -1)$ and $(-1.7693, 0)$.) In fact, one has $E(\mathbb{Z}) = \{(1, -1)\}$ and $\#E(\mathbb{Q}) = \infty$ – not at all obvious!

Definition 5.6.2. We sometimes will use C_f to denote a curve defined by $f(x, y) = 0$.

For a multi-variable polynomial $f(x_1, x_2, \ldots, x_n)$, we can write it as a sum of monomials, that are products of the variables:
$$f(x_1, \ldots, x_n) = \sum_{i_1, \ldots, i_n \geq 0} a_{i_1, \ldots, i_n} x_1^{i_1} \cdots x_n^{i_n}.$$
We say that the degree of f, written $\deg(f)$, is equal to d if d is the largest exponent $i_1 + \ldots + i_n$ for any monomial with $a_{i_1, \ldots, i_n} \neq 0$. For example, $\deg(y^2 - x^3 - xy) = 3$, $\deg(x^5 - yz^2 + xyz^4) = 6$ and $\deg(x^2 - yz + 1) = 2$. For a variable x_i, we say that $\deg_{x_i}(f(x_1, \ldots, x_n)) = d$ if d is the largest exponent which appears for x_i in f.

We will sometimes write $\deg(C_f) := \deg(f)$ for the degree of C_f.

- If $\deg(C_f) = 1$, we call C_f a *line.*
- If $\deg(C_f) = 2$, we call C_f a *conic,* or *quadratic*; such curves include circles, ellipses, parabolas or hyperbolas – or an empty set, such as $x^2 + y^2 = -1$.
- If $\deg(C_f) = 3$, we say C_f is *cubic.* This includes elliptic curves.
- If $\deg(C_f) = 4$, we say C_f is *quartic.*

Example 5.6.2.

- $C_3 : xy^4 = x^3y^2 + 1$ has degree 5;
- $C_2 : xy + y = x^3y$ has degree 4;
- $E : y^2 = x^3 - 2x + 2$ has degree 3.

We're interested in finding integral and rational points to plane curves. For conics (degree 2) defined over $\mathbb{Q}$, there's a really nice way to construct infinitely many rational points from a single point: the idea is that if a point on a conic is rational, then a line through that point gives another rational point on the conic.

Example 5.6.3. We're interested in finding all rational points on the conic
$$C : x^2 + 5y^2 = 1$$
(this is an ellipse). We visibly see the point $(1, 0)$. We'll use this to construct all other rational points.

Let $m \in \mathbb{Q}$ be any rational number. Consider the line through $(1,0)$ with slope m:

$$L : y - y_1 = m(x - x_1),$$

i.e.,

$$L : y = m(x-1).$$

Let's find the points of intersection of L and C; we expect two points since $\deg(L) \cdot \deg(C) = 2$ (though it can be one point, twice-intersected!). One of these points is $(1,0)$; **the other point should also be rational since the line and conic are rational.**

Let $(x_2, y_2) \in L \cap C$. Then both

$$(x_2, y_2) \in L \Rightarrow y_2 = m(x_2 - 1) \tag{5}$$

and

$$(x_2, y_2) \in C \Rightarrow x_2^2 + 5y_2^2 = 1. \tag{6}$$

We'll plug (5) into (6) and solve for x_2. I'll just write $x = x_2$ (to emphasize that we're "solving for x"):

$$\begin{aligned} x^2 + 5(m(x-1))^2 = 1 &\Rightarrow x^2 + 5m^2(x-1)^2 - 1 = 0 \\ &\Rightarrow x^2 + 5m^2(x^2 - 2x + 1) - 1 = 0 \\ &\Rightarrow x^2 + 5m^2x^2 - 10m^2x + 5m^2 - 1 = 0 \\ &\Rightarrow (1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1) = 0 \\ &\Rightarrow \dots \\ &\Rightarrow (x-1) \cdot [(5m^2+1)x - (5m^2 - 1)] = 0. \end{aligned}$$

Thus, the roots of this polynomial are $x = x_1 = 1$, and $x = x_2 = \frac{5m^2-1}{5m^2+1}$. How did we factorize this? You can do this several ways:

- Set $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1) = (ax - b)(cx - d)$, and solve for a, b, c, d in terms of m;
- Divide $x - 1$ into $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1)$ using long division, noting that $x - 1$ has to divide it since 1 must be a root (as $(1,0)$ is in $L \cap C$);
- Avoid factorizing, and instead use the quadratic formula to get the roots of $(1 + 5m^2)x^2 - 10m^2x + (5m^2 - 1)$.

In any case, to get the y-coordinate y_2, simply plug in x_2 back into the line:

$$y_2 = m(x_2 - 1) = m\left(\frac{5m^2 - 1 - (5m^2 + 1)}{5m^2 + 1}\right) = \frac{-2m}{5m^2 + 1}.$$

Therefore, we've created a *new rational point*

$$(x_1, y_1) = \left(\frac{5m^2 - 1}{5m^2 + 1}, \frac{-2m}{5m^2 + 1}\right)$$

just with $(1,0)$ and a number $m \in \mathbb{Q}$ (using the line through $(1,0)$ with slope m). Also note that (using the line equation)

$$m = \frac{y_2}{x_2 - 1}.$$

On the other hand, given any rational point $(x_2, y_2) \in C(\mathbb{Q})$ other than $(1, 0)$, the line through $(1, 0)$ and (x_2, y_2) has slope $m = \frac{y_2}{x_2 - 1}$.

The above example illustrates a **bijection** between $\mathbb{Q}$ and $C(\mathbb{Q}) \setminus \{(1, 0)\}$. We say that such a curve is *parametrizable.* In fact, in our example above, if for $m \in \mathbb{Q}$ we write $m = \frac{r}{s}$, then the corresponding solution

$$(x_2, y_2) = \left(\frac{5r^2 - s^2}{5r^2 + s^2}, \frac{-2rs}{5r^2 + s^2} \right).$$

This looks a lot like our parametrization of primitive solutions to the Fermat curve $F_2 : x^2 + y^2 = z^2$. (However, this does not characterize all primitive solutions on $x^2 + 5y^2 = 1$.)

Let's recap our technique in the example above. The idea was that, given a conic and a rational point P on it, we could construct the other rational points using all rational lines through P.

Let's run through what we did on the previous example, for $C_f : x^2 + 5y^2 = 1$.

"Algorithm" for producing many rational points on C_f:

- Given a conic C_f and a point (x_1, y_1) on C_f, choose $m \in \mathbb{Q}$ and constructed a line $L : y = mx + b$.
- A point (x_2, y_2) in the intersection $L \cap C$ satisfies both $y_2 = mx_2 + b$ and $f(x_2, y_2) = 0$.
- Substituting in $y_2 = mx_2 + b$ to $f(x_2, y_2) = 0$ gives $f(x_2, mx_2 + b) = 0$.
- The one-variable polynomial $g(x) := f(x, mx + b)$ is degree two in x (since f was degree two in x), with x_1 as a root; **since $g(x)$ has rational coefficients and x_1 is rational, this forces the other root x_2 to be rational.**
  - Find the other root x_2 using e.g. the quadratic formula on $g(x)$, or
  - the fact that $(x - x_1)$ divides $g(x)$.
- Plugging in x_2 to the line, we get $y_2 = mx_2 + b$.
- Thus we produce a point $(x_2, y_2) \in C_f(\mathbb{Q})$.

Are there any issues with applying this technique to other curves?

First observation: our method could fail if $g(x) := f(x, mx + b)$ has no rational roots besides x_1. For example, if we had $x_1 = 1$ and $g(x) = x^3 - 1 = (x-1)(x^2+x+1)$. This could happen if $\deg(f(x, y)) > 2$. But if f is quadratic (i.e., C_f is a conic), then $f(x, mx+b)$ is at most quadratic in x. **Fact:** A quadratic polynomial in x has a rational root if and only if both roots are rational (for example, use the quadratic formula to see this).

So if $\deg_x(g(x)) = 3$, this method can fail. However, for cubic curves C_f (such as elliptic curves), we'll use a modified version to produce rational points (called the "chord and tangent method").

Second observation: this method could fail to produce new rational points is if $g(x)$ has no new rational points, i.e., $g(x) = (x - x_1)^d$. This implies that (x_1, y_1) is a *singular point* on C_f.

Definition 5.6.3. Given a polynomial $f(x, y) \in \mathbb{R}[x, y]$, a *singular point* on C_f is a point $(x_0, y_0) \in C_f$ **on** C_f such that

$$\left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} = 0$$

and

$$\left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} = 0.$$

If (x_0, y_0) does not satisfy both conditions, then it is said to be *nonsingular.*

Example 5.6.4. Singular points look weird, and often have "self-intersection" issues.

- For example,
$$C_1 : y^2 = x^3$$
looks like this: (Draw picture). The singular point is $(0, 0)$, and is a "cusp."
- The curve
$$C_2 : y^2 = x^3 + x^2$$
also has a singular point at $(0, 0)$; this one is called a *node.*
- The curve
$$C_3 : (x - 1)^4 = (y - 2)^2$$
has a singular point at $(1, 2)$ (also a cusp).

As it turns out, our method generalizes to any **nonsingular conic.**

This can be called the "rational points on nonsingular conics" theorem.

Theorem 5.6.1. [NZM91, Page 255] *Let C_f be a nonsingular conic defined over $\mathbb{Q}$ (i.e., $f(x, y) \in \mathbb{Q}[x, y]$). If $C_f(\mathbb{Q}) \neq \emptyset$, then $\#C_f(\mathbb{Q}) = \infty$.*

Here is a more precise description of this theorem. Fix $(x_1, y_1) \in C_f(\mathbb{Q})$. Let m_0 be the tangent slope of C_f at (x_1, y_1), and m_1, m_2 the roots of the quadratic-in-y polynomial $f(1, y)$. Then for any $m \in \mathbb{Q}$ not equal to m_0, m_1 and m_2, there exists a rational point $(x_2, y_2) \in L \cap C$ different from (x_1, y_1), where L is the line through (x_1, y_1) with slope m.

The proof is sketched out on page 255 [NZM91]; however, I've omitted it since it's a bit geometric.

Rational points on cubics. Next, we shift our attention to studying $C_f(\mathbb{Q})$ where C_f is a cubic curve, i.e., $\deg(f(x, y)) = 3$. As noted above, our technique for constructing rational points on conics might fail for cubics. However, we can often create a new rational point with two fixed rational points.

Example 5.6.5. Consider the cubic curve $C : y^2 = x^3 + 17$. We visibly see several rational points: $(-1, \pm 4)$, $(-2, \pm 3)$ and $(2, \pm 5)$.

Let's try and construct more rational points, in analogy to the example with conics. Let us fix points $P_1 = (-1, 4)$ and $P_2 = (2, 5)$. Let L be the line through P_1 and P_2: then L has slope

$$m = \frac{5 - 4}{2 - (-1)} = \frac{1}{3},$$

and an equation for it is

$$L : y - 4 = \frac{1}{3}(x + 1),$$

i.e.,

$$L : y = \frac{1}{3}x + \frac{13}{3}.$$

Let's study the intersection $L \cap C$. If we substitute our line equation for y into the equation for C, we get

$$(\frac{1}{3}x + \frac{13}{3})^2 = x^3 + 17.$$

Expanding this out, we get that

$$x^3 - \frac{1}{9}x^2 - \frac{26}{9}x - \frac{16}{9} = 0.$$

Thus (multiplying everything by 9), we are looking for the roots of

$$9x^3 - x^2 - 26x - 16 = 0.$$

We already know that -1 and 2 are roots (since $(-1, 4)$ and $(2, 5)$ lie in $L \cap C$). 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: long division shows us that

$$9x^3 - x^2 - 26x - 16 = (x + 1(x - 2)(x + \frac{8}{9}) = 0.$$

Therefore, we have our third x-coordinate: take $x_3 := -\frac{8}{9}$. Plugging this back into L shows we can take $y_3 := \frac{109}{27}$. Thus, we conclude that

$$\left(\frac{-8}{9}, \frac{109}{27}\right) \in C(\mathbb{Q}).$$

This method is called the **chord method** (for producing rational points).

There's another way which will construct rational points (hopefully new!). This way is called the **tangent method,** and uses a single point like our conics example: this time, it uses the *tangent line* through P_1.

Fix $P_1 := (-1, 4) \in C(\mathbb{Q})$. The tangent slope m_0 of C at P_1 is computed via the implicit derivative evaluation $m_0 := \left.\frac{dy}{dx}\right|_{P_1}$. From $y^2 = x^3 + 17$, we see that (taking implicit derivatives w.r.t. x)

$$2y \cdot \frac{dy}{dx} = 3x^2,$$

and thus

$$\frac{dy}{dx} = \frac{3x^2}{2y}.$$

In particular,

$$m_0 = \frac{3}{8}.$$

The tangent line of C at P_1 is

$$L : y - y_1 = m_0(x - x_1),$$

i.e.,

$$L : y = \frac{3}{8}x + \frac{35}{8}.$$

Plugging this expression for y into $f(x, y)$, we get that

$$\left(\frac{3}{8}x + \frac{35}{8}\right)^2 = x^3 + 17,$$

so that (after much simplification)

$$64x^3 - 9x^2 - 210x - 137 = 0.$$

We know that $x + 1$ divides this; in fact, since our line was *tangent* to C at P_1, we know that $(x+1)^2 = x^2+2x+1$ divides it! P_1 is on $L \cap C$ "twice." Thus, long division shows us the final factor:

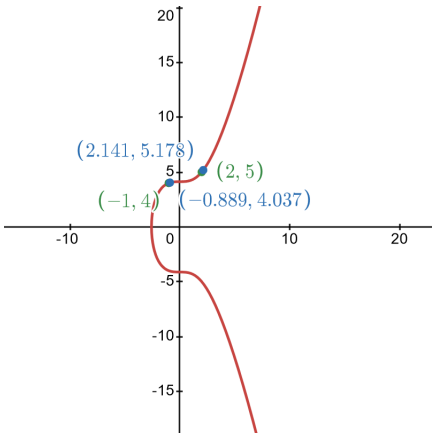
$$64x^3 - 9x^2 - 210x - 137 = (x+1)^2(x - \frac{137}{64}) = 0.$$

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 conclude that

$$\left(\frac{137}{64}, \frac{2651}{512}\right)$$

is a rational point on $C(\mathbb{Q})$.

To graph this curve: note that $\left(\frac{-8}{9}, \frac{109}{27}\right) \approx (-.89, 4.0)$ and $\left(\frac{137}{64}, \frac{2651}{512}\right) \approx (2.1, 5.2)$. Here's a graph:



The above curve is an example of an **elliptic curve.** Our methods for creating rational points is called the **chord and tangent method.**

One way our construction of rational points on elliptic curves differs from that of a nonsingular *conic,* is that this procedure does not necessarily produce infinitely many distinct rational points. Unlike nonsingular conics C, where $C(\mathbb{Q}) \neq \emptyset \Rightarrow \#C(\mathbb{Q}) = \infty$, elliptic curves will always have $E(\mathbb{Q}) \neq \emptyset$ (in *projective space*) but can also have $\#E(\mathbb{Q}) < \infty$. We'll study $E(\mathbb{Q})$ more in §5.7.

Projective space. The usual x, y-plane over $\mathbb{R}$, denoted by $\mathbb{R}^2$, is an example of an *affine space.* As it turns out, algebraic curves can have points on them which don't lie in affine space, but lie in the ambient **projective space;** such points are "invisible" over $\mathbb{R}^2$.

Let's define this algebraically first. We will define the **projective plane** as a quotient set under an equivalence relation.

Let's define an equivalence relation on $\mathbb{R}^3 \setminus \{(0,0,0)\}$. For any nonzero $(a,b,c) \in \mathbb{R}^3$ and any nonzero $\lambda \in \mathbb{R}$, we will say $(a,b,c) \sim (\lambda a, \lambda b, \lambda c)$. Then the projective plane $\mathbb{P}_2(\mathbb{R})$ is defined as the quotient set

$$\mathbb{P}_2(\mathbb{R}) := \mathbb{R}^3 \setminus \{(0,0,0)\}/((a,b,c) \sim (\lambda a, \lambda b, \lambda c)).$$

For example, in $\mathbb{P}_2(\mathbb{R})$ we have $(1,2,3) \sim (2,4,6) \sim (\pi, 2\pi, 3\pi) \sim (100,200,300) \sim \ldots$.

Given an element of $\mathbb{P}_2(\mathbb{R})$, which is an equivalence class $[(a,b,c)]$, we will write $[a:b:c]$ instead.

Points of the form (a,b,c) and $(\lambda a, \lambda b, \lambda c)$ are on the same line in $\mathbb{R}^3$ through the origin. Thus, **our points in $\mathbb{P}_2(\mathbb{R})$ are in correspondence with lines in $\mathbb{R}^3$ through the origin.**

The real plane $\mathbb{R}^2$ lives inside projective space $\mathbb{P}_2(\mathbb{R})$, via the map

$$(a,b) \mapsto [a:b:1].$$

There are points in $\mathbb{P}_2(\mathbb{R})$ that aren't on $\mathbb{R}^2$; such points are of the form $[a:b:0]$, and are called **points at infinity.** They "sit above" parallel lines, and are vanishing points (at the end of the horizon).

(Draw a picture of a road, two parallel lines which, from the viewer's perspective, converge at the horizon.)

Fact: curves in $\mathbb{R}^2$, if they go off in either direction, can actually "converge" to a point at infinity if they have one. For example, $y = x^2$ has the point at infinity $[0:1:0]$ (see below).

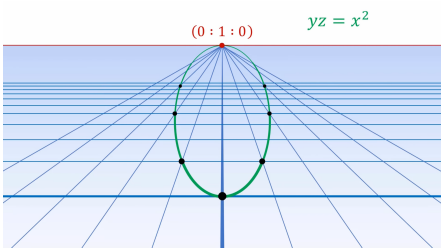


FIGURE 1. The parabola $y = x^2$ in $\mathbb{P}^2(\mathbb{R})$. Picture from here.

I'll share a video on Carmen which goes into more detail about projective geometry.

What is the upshot? Elliptic curves have a point at infinity – usually this is $[0:1:0]$ (depends on the equation). Taking this as our group law identity element makes calculations simpler. It also lets us work with vertical lines in $\mathbb{R}^2$ if they come up during the chord and tangent method (the point at infinity is on this line).

Definition 5.6.4. A polynomial $F(X,Y,Z) \in \mathbb{R}[X,Y,Z]$ is **homogeneous** if every monomial in F has the same degree.

(Capital variables are often used for homogeneous polynomials.)

Example 5.6.6. $F(X,Y,Z) := XY^3 - Z^4$ is homogeneous: $\deg(XY^3) = 4 = \deg(Z^4)$. However, $G(X,Y,Z) := XY + XZ - Y$ is not homogeneous.

Given a polynomial $f(x,y) \in \mathbb{R}[x,y]$ of degree $d \geq 1$, we can **homogenize** f by introducing three variables: the *homogenization of f,* denoted F, is

$$F(X,Y,Z) := f(\frac{X}{Z}, \frac{Y}{Z}) \cdot Z^d$$

The point is to make each monomial term in $f(x,y)$ have equal total degree ("homogeneity=the same").

Realistically: to compute a homogenization $F(X,Y,Z)$ of degree d $f(x,y)$, you can just multiply all terms by an appropriate power of Z until all the monomial degrees are $\deg(f(x,y))$.

Conversely: given a polynomial $F(X,Y,Z) \in \mathbb{R}[X,Y,Z]$, we can **dehomogenize** it by setting $Z = 1$.

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
$$h(x,y) = y^2 - x^3 - 2x.$$

Upshot: given an **affine** curve C_f (so a curve in $\mathbb{R}^2$), we get a corresponding **projective** curve C_F (in $\mathbb{P}_2(\mathbb{R})$). C_F is the set of solutions in $\mathbb{P}_2(\mathbb{R})$ to

$$F(X,Y,Z) = 0$$

where we plug in equivalence class points $[a:b:c]$ instead of (a,b,c).

Question: why is plugging in equivalence classes to $F(X,Y,Z)$ well-defined? This is because $F(X,Y,Z)$ is homogeneous, and so

$$F(\lambda a, \lambda b, \lambda c) = \lambda^d F(a,b,c),$$

where $d = \deg(f) = \deg(F)$. Thus,

$$F(a,b,c) = 0$$

if and only if

$$F(\lambda a, \lambda b, \lambda c) = 0$$

for any $\lambda \neq 0 \in \mathbb{R}$, and so

$$F([a:b:c]) = 0$$

makes sense algebraically.

Remark. Since $\mathbb{R}^2 \subseteq \mathbb{P}_2(\mathbb{R})$ via $(a,b) \mapsto [a:b:1]$, the points of an affine curve C_f live inside its homogenization C_F by the same map. However, C_F might include "points at infinity" of the form $[a:b:0]$.

Definition 5.6.5. If C_F is a projective curve, then a point $[a : b : c]$ on C_F is a *singular point* if all **three** partials are zero:

$$\left.\frac{\partial F}{\partial X}\right|_{[a:b:c]} = 0,$$

$$\left.\frac{\partial F}{\partial Y}\right|_{[a:b:c]} = 0,$$

and

$$\left.\frac{\partial F}{\partial Z}\right|_{[a:b:c]} = 0.$$

Remark. For a 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 – however, there may be singular points at infinity on C_F (of the form $[a : b : 0]$).

Irreducible curves. We need one more term before defining an elliptic curve.

Definition 5.6.6. For a ring R (such as $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$), we say that a polynomial $f(x_1, \ldots, x_n) \in R[x_1, \ldots, x_n]$ is *irreducible over* R if it does not factorize as two polynomials $g, h \in R[x_1, \ldots, x_n]$ of strictly smaller degree. In such a case, we call the corresponding curve C_f an *irreducible curve over* R.

When $f = g \cdot h$ over R, we say that g *divides* f *over* R, and write $g \mid f$.

Irreducible is in analogy to an integer being prime.

Example 5.6.8. We have:

- $f_1(x, y) := x^2 - y^2$ is *reducible* over $\mathbb{R}$, as $f = (x + y)(x - y)$.
- $f_2(x) := x^2 + x + 1$ is *irreducible* over $\mathbb{R}$ since its roots are complex (quadratic formula); but, it's reducible over $\mathbb{C}$.
- $f_3(x, y) := x^2 + y^2$ is *irreducible* over $\mathbb{R}$; however, since $f_2 = (x + iy)(x - iy)$, it is *reducible* over $\mathbb{C}$.
- $f_4(x, y) := x^3 + y^3 - z^3$ is *irreducible* over $\mathbb{C}$.

Fact: if f and g are polynomials in $\mathbb{Z}[x, y]$ and $g \mid f$ over $\mathbb{R}$ or $\mathbb{C}$, then we have $C_g(\mathbb{Z}) \subseteq C_f(\mathbb{Z})$, as well as $C_g(\mathbb{Q}) \subseteq C_f(\mathbb{Q})$ and $C_g(\mathbb{R}) \subseteq C_f(\mathbb{R})$. Therefore, if you're given a curve C_f over $\mathbb{Z}$ and you can find a factorization of f into a smaller polynomial g, then integral/rational/real solutions of g are also integral/rational/real solutions of f.

For example: from the above, we have $x^2 - y^2 = (x + y)(x - y)$, and so any integral point on the line $y = x$ is also an integral point on the curve $x^2 - y^2$.

Therefore, we are interested in studying *irreducible nonsingular cubics defined over* $\mathbb{Q}$. Such curves are called **elliptic curves.**

Exercise 5.6.1. This exercise determines when certain plane curves are nonsingular.

a) Show that for any polynomial $f(x) \in \mathbb{Z}[x]$ and for any integer $n \geq 2$, the curve

$$C : y^n = f(x)$$

in $\mathbb{R}^2$ 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

$$C : y^2 = (x - \alpha)(x - \beta)(x - \gamma)$$

where α, β, γ are 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.

In particular, when a cubic polynomial $f(x) \in \mathbb{Q}[x]$ has no repeated roots, the cubic curve defined by $y^2 = f(x)$ is nonsingular, and in fact is an elliptic curve.

Exercise 5.6.2. 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, the previous exercise might help.)

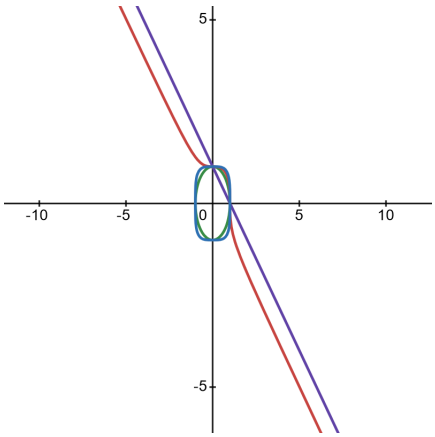


FIGURE 2. The Fermat curves F_1, F_2, F_3 and F_4 in $\mathbb{R}^2$.

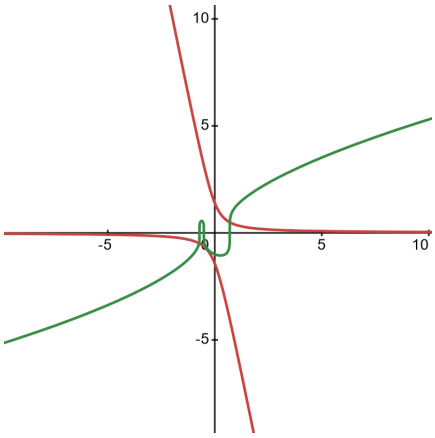


FIGURE 3. The curves C_1 (hyperbola) and C_2 (hyperelliptic) in $\mathbb{R}^2$.

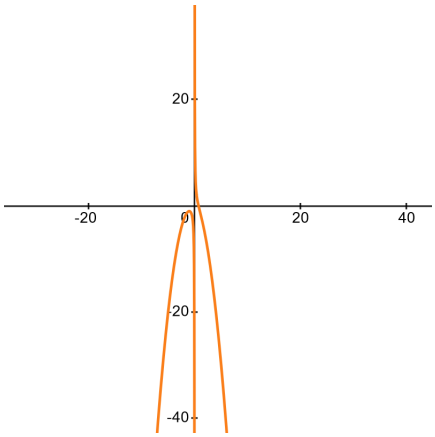


FIGURE 4. The curve C_3, dehomogenized and pictured in $\mathbb{R}^2$.

Exercise 5.6.3. Parametrize the rational points on the hyperbola

$$H : x^2 - 2y^2 = 1$$

with the point $(1, 0) \in C(\mathbb{Q})$.

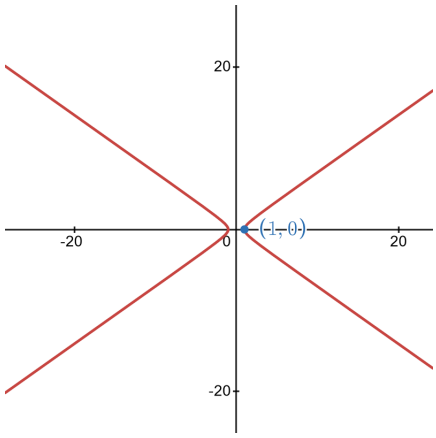


FIGURE 5. The hyperbola $H : x^2 - 2y^2 = 1$.

5.7. **Elliptic curves.** In this section, we will study the *arithmetic of elliptic curves.* We will analyze their group law and the structure of their group, its connection to classic Diophantine geometry from Greek antiquity, its application towards 21st century mathematics (such as Fermat's last theorem), and its unique behavior when reduced modulo primes p (and possibly its connection to cryptography).

Disclaimer, added after teaching this class: this is similar to "Chapter zero" of my notes on "the arithmetic of elliptic curves." Thus, if you are reading these notes to learn about elliptic curves, I suggest that you read those instead, as they are more polished.

Definition 5.7.1. The most general definition we have is: an **elliptic curve defined over** $\mathbb{Q}$**,** written $E/\mathbb{Q}$, is a nonsingular irreducible cubic curve $E_{f(x,y)}$ where $f(x, y) \in \mathbb{Q}[x, y]$. A *projective elliptic curve defined over* $\mathbb{Q}$ is a nonsingular cubic curve $E_{F(X,Y,Z)}$ where $F(X, Y, Z) \in \mathbb{Q}[X, Y, Z]$ is homogeneous.

Recall that we can make an affine elliptic curve E projective by homogenizing $f(x, y) \rightsquigarrow F(X, Y, Z)$. We'll sometimes write E_H for the "homogenization" of E.

Remark. "Irreducibility" is only a necessary assumption for affine curves, since a nonsingular projective curve must be irreducible (by e.g. Bézout's theorem).

Definition 5.7.2. More practical definition: an elliptic curve over $\mathbb{Q}$ in **short Weierstrass form** is a curve E defined by the equation

$$E : y^2 = x^3 + Ax + B$$

where $A, B \in \mathbb{Q}$ and $\Delta := \Delta_E := -16(4A^3 + 27B^2) \neq 0$. We call Δ_E the **discriminant** of E.

The homogenization of E above is the curve E_H in $\mathbb{P}_2(\mathbb{R})$ defined by

$$E_H : Y^2Z = X^3 + AXZ^2 + BZ^3.$$

Question: what are the points at infinity on $E_{/\mathbb{R}}$, i.e., what are the points on $E_H : Y^2Z = X^3 + AXZ^2 + BZ^3$ of the form $[a : b : 0]$? Plugging this into E_H, we see that

$$0 = a^3,$$

and thus $a = 0$. Therefore, $[a : b : 0] = [0 : b : 0] = [0 : 1 : 0]$. Therefore, any elliptic curve has exactly one point at infinity. Thus, the homogenization can be described as

$$E_H(\mathbb{R}) = \{[x : y : 1] \in \mathbb{P}_2(\mathbb{Q}) : y^2 = x^3 + Ax + B\} \cup \{[0 : 1 : 0]\}.$$

When talking about points on an elliptic curve of the form $E_{/\mathbb{Q}} : y^2 = x^3+Ax+B$, we will include the point at infinity, denoted $O := [0 : 1 : 0]$, in our rational solutions, and write $O \in E(\mathbb{Q})$, since O can be expressed with rational coordinates. Thus, **elliptic curves of the form $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Q}$ always have at least one rational point, namely O.**

As mentioned last week, the upshot to working with the projective version is that we now have an extra point $[0 : 1 : 0]$ on E, which lets us use the chord and tangent method even when vertical lines appear in the calculations.

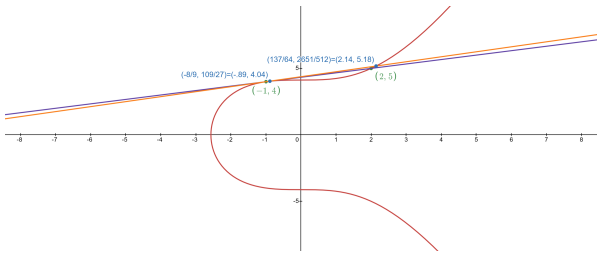
Example 5.7.1. Consider the equation from last week,

$$E : y^2 = x^3 + 17.$$

To make sure it's an elliptic curve, we need to check that $\Delta := -16(4A^3 + 27B^2) \neq 0$. We have $A = 0$ and $B = 17$, and so $\Delta = -16 \cdot (27 \cdot 17^2) \neq 0$. Thus, E is indeed an elliptic curve (and defined over $\mathbb{Q}$).

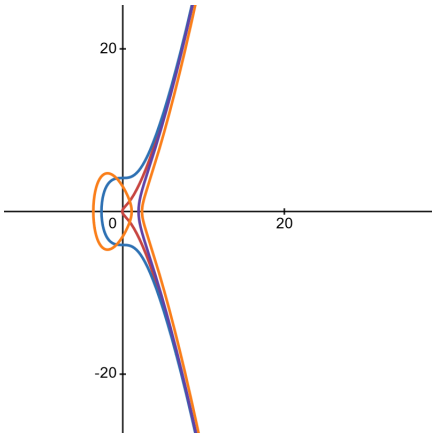
We spotted a few points on it previously: $(-1, \pm 4), (-2, \pm 3)$ and $(2, \pm 5) \in E(\mathbb{Q})$. Using the chord through $(-1, 4)$ and $(2, 5)$, we constructed the point $\left(-\frac{8}{9}, \frac{109}{27}\right) \in E(\mathbb{Q})$ ($\approx (-.89, 4.04)$). Using the tangent through $(-1, 4)$, we constructed the point $\left(\frac{137}{64}, \frac{2651}{512}\right) \in E(\mathbb{Q})$ ($\approx 2.14, 5.18$).

Graphing E in $\mathbb{R}^2$, it looks like this:



It's sort of shaped like a horseshoe.

In general, an elliptic curve in $\mathbb{R}^2$ looks like a horseshoe, a "pinched horseshoe" or some stretch/squash of either two:



The picture of E has one or two connected components. These correspond to whether $x^3 + Ax + B$ has one or three real roots (thus, one or three x-intercepts). For an elliptic curve $E : y^2 = f(x)$, the real roots α of $f(x)$ give x-intercepts of $E(\mathbb{R})$, via the points $(\alpha, 0)$.

The group law on an elliptic curve: an example. We will show that for an elliptic curve $E_{/\mathbb{Q}}$, the "chord and tangent method" can be used to turn the set $E(\mathbb{Q})$ of rational points into a group! This also applies to $E(\mathbb{R})$. First, we'll do an example.

Example 5.7.2. We will consider the curve $E : y^2 = x^3 - 7x + 10$. Its discriminant is $\Delta = -21248 \neq 0$, so E is an elliptic curve. Its graph has one connected component.

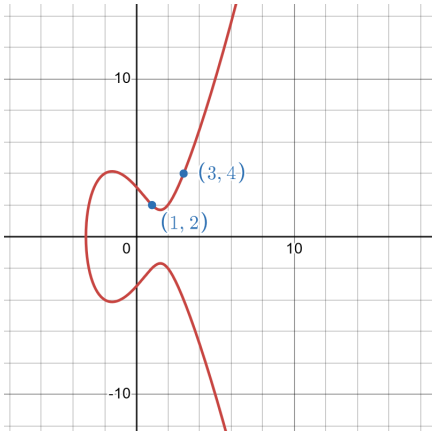


FIGURE 6. The elliptic curve $E : y^2 = x^3 - 7x + 10$.

We can check that both $P_1 := (1, 2)$ and $P_2 := (3, 4)$ are points on $E(\mathbb{R})$. Their sum, written $P_1 \oplus P_2$, is determined by a "chord and tangent method" applied twice.

1. Step 1: let $L_1 := L_{P_1,P_2}$ be the line through P_1 and P_2. Then its slope is $m = \frac{4-2}{3-1} = 1$, and thus it has the equation
$$L_1 : y - y_1 = m(x - x_1),$$
i.e.,
$$L_1 : y = x + 1.$$

Let's analyze the intersection $L_1 \cap E$. To do this, we'll 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 $x = 1, 3$ or -3. We knew $x = 1$ and $x-3$ were already roots of this polynomial, since $P_1, P_2 \in L_1 \cap E$. Thus, let us set $x_3 := -3$. Taking $y_3 := x_3 + 1 = -2$, we conclude that $R := P_1 * P_2 = (-3, 2)$ is on E and collinear to P_1 and P_2.

2. Step 2: consider the line through R and $O := [0 : 1 : 0]$. What does it look like? Let us write it as

$$L_2 := L_{R,O} := ax + by = c.$$

Homogenizing it 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 $R \in L$, we have $a \cdot -3 = c$. Thus,

$$L_2 : ax = -3a;$$

we can divide by a, and get a new equation for the same line

$$L_2 : x = -3.$$

Let's analyze $L_2 \cap E$. Plug in $x = -3$ into E and get an equation

$$y^2 = (-3)^3 - 7 \cdot (-3) + 10 = 4.$$

Thus, two points on $L_2 \cap E$ are $(-3, \pm 2)$ (only two points appear since the third point is the point at infinity). Then we take $P_1 \oplus P_2 := (-3, -2)$.

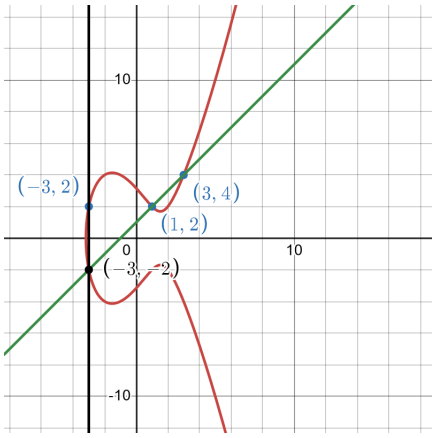


FIGURE 7. The elliptic curve $E : y^2 = x^3 - 7x + 10$.

Let's find out how to add $P_1 := (1, 2)$ to itself; this is written as $P_1 \oplus P_1$, or simply $2P_1$. We'll use the tangent method in the first step.

1. Step 1: let L_3 be the *tangent* line through P_1:
$$L_3 : y = m_0(x - 1) + 2$$
where m_0 is the tangent slope of E at P_1. We can compute it:
$$\frac{d}{dx}[y^2 = x^3 - 7x + 10] \Rightarrow \frac{dy}{dx} = \frac{3x^2 - 7}{2y},$$
and so
$$m_0 = \left.\frac{dy}{dx}\right|_{(1,2)} = \frac{-4}{4} = -1.$$
Thus,
$$L_3 : y = 3 - x.$$
Plug in $y = 3 - x$ into $y^2 = x^3 - 7x + 10$ and solve for x:
$$\begin{aligned} y^2 = x^3 - 7x + 10 &\Rightarrow (3 - x)^2 = x^3 - 7x + 10 \\ &\Rightarrow x^3 - x^2 - x + 1 = 0 \\ &\Rightarrow x^2(x - 1) - (x - 1) = 0 \\ &\Rightarrow (x^2 - 1)(x - 1) = 0 \\ &\Rightarrow (x - 1)^2(x + 1) = 0 \end{aligned}$$
$((x-1)$ appears twice as expected, since $(1, 2)$ has multiplicity two on the line). Thus, we can take $x_3 := -1$ and $y_3 := 3 - x_3 = 4$, and deduce that
$$R := P_1 * P_1 := (-1, 4).$$
2. Take the line $L_4 := L_{4,R,O}$ through R and O. As observed before, it should be the vertical line through R; so take
$$L_4 : x = -1.$$

Then the third point of intersection on L_4 is $(-1,-4)$ (simply note that we know $y=4$ is already a root of $y^2=(-1)^3-7(-1)+10$, and thus the other root must be negative this). We conclude that the sum of P_1 with itself is $(-1,-4)$.

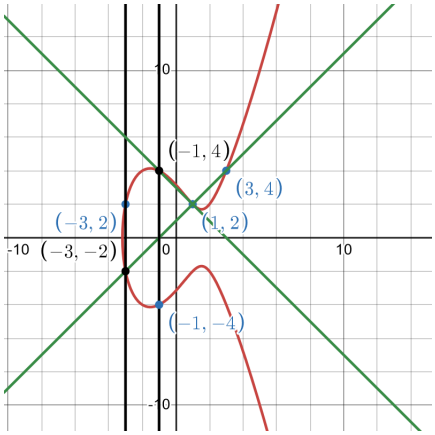


FIGURE 8. The elliptic curve $E: y^2=x^3-7x+10$.

The group law. Now we'll explain the general group law for *any* elliptic curve (in Weierstrass form or otherwise).

1. The group law, general form:
   (1) Fix a rational *inflection point* $O \in E(\mathbb{Q})$ (i.e., where the tangent line to E at O has intersection multiplicity 3; if O is affine, then this is where concavity on E changes, like in Calculus 1).
   (2) Given two points $P_1, P_2 \in E(\mathbb{Q})$, define their **sum** $P_1 \oplus P_2$ as follows:
   (3) First, take the line L_{P_1,P_2} through P_1 and P_2 (if $P_1=P_2$, take the tangent line to E at P_1). It will intersect the curve at a **rational** third point (possibly a point at infinity!). Call this third point $R := P_1 * P_2$.
   (4) Then take the line $L_{R,O}$; it will intersect the curve at another third **rational** point, which is what is our sum, $P_1 \oplus P_2$.

2. The group law, short Weierstrass form: in practice, our elliptic curve usually has the equation $E: y^2=x^3+Ax+B$ (here, E is said to be in **(short) Weierstrass form**).
   (1) Fix $O := [0:1:0] \in E(\mathbb{Q})$ (this is an inflection point).
   (2) Given $P_1, P_2 \in E(\mathbb{Q})$, define $P_1 \oplus P_2$ as follows.
   (3) Take the line L_{P_1,P_2}. It intersects the curve at a third point, say $R := P_1 * P_2$.
   (4) **The line through R and O is just the *vertical* line through R;** it intersects the curve at a third point, which is our sum $P_1 \oplus P_2$.

Remark. An upshot to working with elliptic curves in short Weierstrass form: in step (2), if R is affine, we can write $R=(x_3,y_3)$. Then the sum $P_1 \oplus P_2$ is just $(x_3,-y_3)$!

If $P_1 * P_2 = O$, then step (3) returns $P_1 \oplus P_2 = O$, since the line through O and O intersects E only at O (this is the **key property** of inflection points that we need: tangent line at itself intersects the elliptic curve 3 times at itself).

So here are things we'll keep in mind before our calculations **(bookmark this!):**

1. Elliptic curves of the form $E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B$ have one point at infinity.
2. For an elliptic curve $E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B$, the point at infinity $O = [0 : 1 : 0]$ is an inflection point: **thus, the line through O and O intersects E only at O.**
3. The tangent line L to E at a point $P \in E$ "contains P twice" (i.e., P has multiplicity two on L).
4. The point at infinity $[0 : 1 : 0]$ *lies on every vertical line in* $\mathbb{R}^2$*:* to see this, note that a vertical line has an equation $L : ax = c$; homogenizing gives $L_H : aX = cZ$, and $[0 : 1 : 0]$ is clearly on this.

Theorem 5.7.1 (Elliptic curve group law)**.** *Given an elliptic curve $E_{/\mathbb{Q}}$, the chord and tangent method described above makes $E(\mathbb{Q})$ a group. (The same is true if you replace $\mathbb{Q}$ with $\mathbb{R}$.)*

Proof. We need to check that $E(\mathbb{Q})$ satisfies the properties of an abelian group (recall §2.10, 2.11 for the definition of a group).

1. $E(\mathbb{Q})$ is **closed** under $\oplus$, i.e., $\oplus$ takes $E(\mathbb{Q})$ to itself.
2. $E(\mathbb{Q})$ is **abelian**: $\forall P_1, P_2 \in E(\mathbb{Q})$, $P_1 \oplus P_2 = P_2 \oplus P_1$.
3. $E(\mathbb{Q})$ has an **identity element:** the point at infinity $O = [0 : 1 : 0]$.
4. $E(\mathbb{Q})$ has additive **inverses:** for $P \in E(\mathbb{Q})$, there exists $Q \in E(\mathbb{Q})$ with
$$P \oplus Q = O$$
(we write $-P := Q$).
5. $\oplus$ is **associative:** $\forall P_1, P_2, P_3 \in E(\mathbb{Q})$,
$$(P_1 \oplus P_2) \oplus P_3 = P_1 \oplus (P_2 \oplus P_3).$$

Let's prove it:

1. Since P_1 and P_2 are rational, so is the third point R on L_{P_1,P_2}; and since O is also rational, so is the third point on $L_{O,R}$, which is $P_1 \oplus P_2$.
2. It's true since for $P_1, P_2 \in E(\mathbb{Q})$, the line through P_1 and P_2 is the same as the line through P_2 and P_1.
3. Given any $P \in E(\mathbb{Q})$, we must show that
$$P \oplus O = P.$$
Consider the line $L_{P,O}$; it goes through a third point $R \in E(\mathbb{Q})$. Then consider the line $L_{R,O}$: this is the same as $L_{P,O}$! Thus, the third point is $P \oplus O = P$.
4. We want to find $Q \in E(\mathbb{Q})$ such that the "two line" process returns the second third point O. That is, we want Q such that the line $L_{P,Q}$ has third point R, where $L_{R,O}$ has third point O.

   Let's try $Q := R$, i.e., take our inverse candidate to be the third point on $L_{P,O}$. Then we must show that $P \oplus R = O$:
   (a) the first line is $L_{P,R}$, which intersects E at the third point O.
   (b) The second line is $L_{O,O}$, the tangent line to E at O; since O is an inflection point, the third point is O, and thus $P \oplus Q = O$.

5. It's sort of tedious to check associativity, so here's an example picture of it:

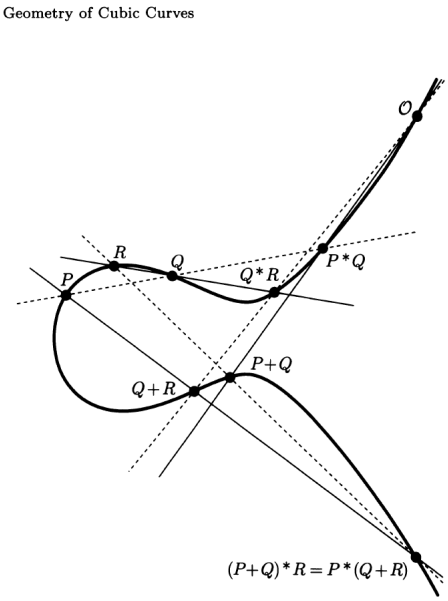


FIGURE 9. Example of associativity on an elliptic curve [ST15].

□

Remark. From here on out, given a point $P \in E(\mathbb{Q})$ and $n \in \mathbb{Z}^+$, 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 $0P := O$.

The group $E(\mathbb{Q})$ is often called the **Mordell-Weil group of E over $\mathbb{Q}$.** We will discuss its structure in the next class.

In the meantime, here is a useful group theory fact concerning the sum of 3 *collinear* points on an elliptic curve.

Proposition 5.7.2. *Let $E_{/\mathbb{Q}}$ be an elliptic curve. If $P_1, P_2, P_3 \in E(\mathbb{Q})$ are collinear, then one has*

$$P_1 \oplus P_2 \oplus P_3 = O.$$

Proof. We'd like to show that

$$P_1 \oplus P_2 = -P_3.$$

Let's go through the chord and tangent method for $P_1 \oplus P_2$ and see what happens.

1. The initial line through P_1 and P_2 passes through a third point, $R := P_1 * P_2$. However, P_1, P_2 and P_3 are on the same line, so $R = P_3$.
2. The second line through O and $R = P_3$ also contains the point $P_1 \oplus P_2$ on E.

With the above in mind, we'll show that $(P_1 \oplus P_2) \oplus P_3 = O$.

1. The line through $P_1 \oplus P_2$ and P_3 intersects the curve at a third point $R' = (P_1 \oplus P_2) * P_3$; by the above step, $R' = O$.
2. The line through O and O intersects E only at O (inflection point).
3. Thus, $(P_1 \oplus P_2) \oplus P_3 = O$, i.e., $P_1 \oplus P_2 \oplus P_3 = O$. □

Example 5.7.3. Let's try another example. We'll analyze the cubic curve

$$E : y^2 = x^3 + 5x.$$

Its discriminant $\Delta = -8000 \neq 0$, so it's an elliptic curve.

Any points we see? I spot $P := (0, 0)$. The x-intercepts are at the roots of $x^3 + 5x = x(x^2 + 5)$. There's only one real root, so one connected component.

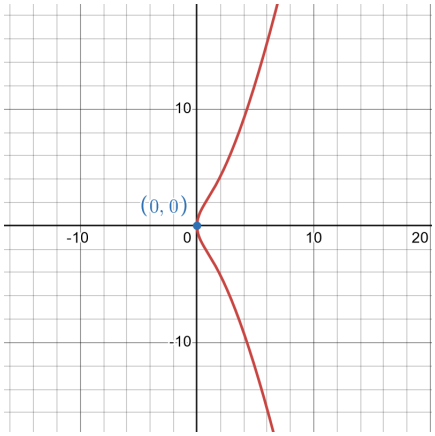


FIGURE 10. The elliptic curve $E : y^2 = x^3 + 5x$.

Let's compute $2P := P \oplus P$. Note that in the first step of the chord and tangent method, we need the tangent line to E at P. It looks like a vertical line, and the third point on it should be $O := [0 : 1 : 0]$.

Let L be the tangent line to E at P; its slope is $m_0 = \frac{3 \cdot (0)^2 + 5}{2 \cdot 0} = \frac{5}{0}$, which is undefined! This means the tangent line is vertical. Thus, it has the form

$$L : ax = c$$

where $a \neq 0$. By our "bookmarked" facts before the group law theorem, we know $O = [0 : 1 : 0]$ is on every vertical line, so it's on L. So the tangent line has O on it, as well as P with multiplicity 2. Thus, P, P and O are collinear, so by the collinearity theorem we have

$$P \oplus P \oplus O = O,$$

i.e.,

$$2P = O.$$

Thus, P has order 2 as a group element of $E(\mathbb{Q})$.

The Mordell-Weil group. What do we know about the Mordell-Weil group $E(\mathbb{Q})$, beyond it being abelian? There is a remarkable theorem about the group structure of $E(\mathbb{Q})$, due to Louis Mordell and later generalized to certain higher-dimensional algebraic equations ("abelian varieties") and fields larger than $\mathbb{Q}$ by André Weil.

Theorem 5.7.3 (The Mordell-Weil theorem)**.** *For an elliptic curve $E_{/\mathbb{Q}}$, its Mordell-Weil group is a* ***finitely generated abelian group:*** *there exist $P_1, P_2, \ldots, P_n \in E(\mathbb{Q})$ such that for any $P \in E(\mathbb{Q})$, one has*

$$P = a_1P_1 \oplus a_2P_2 \oplus \ldots \oplus a_nP_n$$

for some $a_1, a_2, \ldots, a_n \in \mathbb{Z}$.

In particular, by the structure theorem for finitely generated abelian groups (from abstract algebra), one has

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \times E(\mathbb{Q})[\text{tors}],$$

where $r \geq 0$ and $E(\mathbb{Q})[\text{tors}]$ is the subgroup of $E(\mathbb{Q})$ of points with finite order.

In the theorem above, r is called the **rank of E over $\mathbb{Q}$,** and $E(\mathbb{Q})[\text{tors}]$ is the **torsion subgroup of E over $\mathbb{Q}$.**

Example 5.7.4. In our example $E : y^2 = x^3 + 5x$ from last class, we observed that $(0, 0)$ is a point of order two. As it turns out, $E(\mathbb{Q})[\text{tors}] = \{O, (0, 0)\}$. In fact, this elliptic curve has rank one over $\mathbb{Q}$: the point $(20, 90) \in E(\mathbb{Q})$ has infinite order, and we have

$$E(\mathbb{Q}) = \langle (20, 90) \rangle \times \langle (0, 0) \rangle \cong \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

Thus, any point $P \in E(\mathbb{Q})$ has the form

$$P = a(20, 90) \oplus b(0, 0)$$

for some $a, b \in \mathbb{Z}$. For example, $P := \left(\frac{1}{4}, -\frac{9}{8}\right) \in E(\mathbb{Q})$, and we have $P = (20, 90) - (0, 0)$.

Torsion points. The above is our first example of a **torsion point.** These are points with finite order, as group elements of $E(\mathbb{Q})$.

Definition 5.7.3. 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 $E(\mathbb{R})[\text{tors}]$ be the set of *real* torsion points $P \in E(\mathbb{R})$, and $E[\text{tors}] := E(\mathbb{C})[\text{tors}]$ the set of all torsion points on E. We have $E(\mathbb{Q})[\text{tors}] \subseteq E(\mathbb{R})[\text{tors}] \subseteq E(\mathbb{C})[\text{tors}]$. As it turns out, we always have

$$E(\mathbb{R})[\text{tors}] \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{R}/\mathbb{Z},$$

A torsion point $P \in E$ is called an **N-torsion point** if $NP = O$. The **N-torsion subgroup of E** is

$$E[N] := \{P \in E(\mathbb{C}) : NP = O\}.$$

$E(\mathbb{Q})[N]$ is the subgroup of $\mathbb{Q}$-rational N-torsion points.

Here's a pictorial example of a torsion point and its multiples on an elliptic curve:

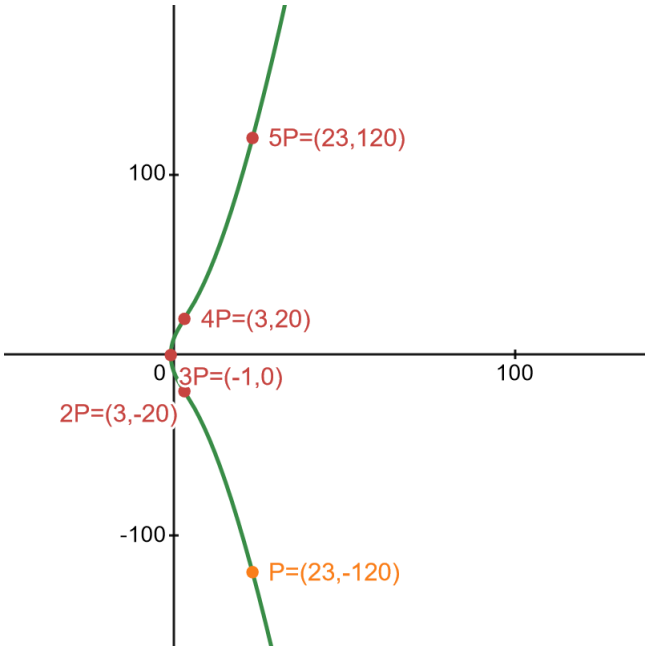


FIGURE 11. The elliptic curve $E : y^2 = x^3 + 93x + 94$, with NP for $P = (23, -120)$ and $1 \leq N \leq 5$. **(take tangent line at P for $2P$, but then take chords between P and $(n-1)P$ to get nP!)**

Example 5.7.5.

1. Every elliptic curve $E_{/\mathbb{Q}}$ has a fixed rational point $O \in E(\mathbb{Q})$ which is the identity element; this is trivially a torsion point, as it has order 1.
2. We've just shown that the elliptic curve $E_1 : y^2 = x^3 + 5x$ has an order 2 torsion point: $(0, 0) \in E(\mathbb{Q})[2]$. As it turns out,
$$E(\mathbb{Q})[2] = \{O, (0, 0)\};$$
however,
$$E[2] = \{O, (0, 0), (\pm i\sqrt{5}, 0)\}.$$
3. The elliptic curve $E_{/\mathbb{Q}} : y^2 = x^3 + 17$ has $E(\mathbb{Q})[\text{tors}] = \{O\}$. However, its Mordell-Weil group is infinite! We saw previously that $(-1, 4) \in E(\mathbb{Q})$. Therefore, the order of $(-1, 4)$ is infinite – thus, you can add it to itself an infinite amount of times to create an infinite amount of rational points on E!
4. In contrast to these, the elliptic curve $E_{/\mathbb{Q}} : y^2 = x^3 + 7$ has *no* rational points besides O; its Mordell-Weil group $E(\mathbb{Q})$ is trivial. (HW 9 proves it has no integral points.)

A particularly nice result is known about rational torsion points for elliptic curves in short Weierstrass form: they are *integral.*

Theorem 5.7.4 (Nagell-Lutz)**.** *Consider an elliptic curve*
$$E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B.$$

If $P \in E(\mathbb{Q})[\text{tors}]$ is a nonzero torsion point, then P is integral. Furthermore, if $2P \neq O$, then writing $P = (x, y)$, one has $y \mid 4A^3 + 27B^2 = \frac{\Delta}{-16}$.

An interesting question to ask is what rational torsion points an elliptic curve $E_{/\mathbb{Q}}$ can have. This has been settled by Barry Mazur (my PhD advisor's PhD advisor).

Theorem (Mazur's Theorem)**.** *$E_{/\mathbb{Q}}$ be an elliptic curve. Then $E(\mathbb{Q})[\text{tors}]$ is one of the following 15 finite abelian groups, up to group 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 $\mathbb{Q}$-torsion group of some elliptic curve $E_{/\mathbb{Q}}$.

It is an active area of research to determine what $E(F)[\text{tors}]$ can be over fields F *larger than* $\mathbb{Q}$. For example, if $F = \mathbb{C}$, then the full torsion group appears:

$$E(\mathbb{C})[\text{tors}] \cong \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}.$$

But when F is a finite extension of $\mathbb{Q}$, $E(F)[\text{tors}]$ is finite (this is a major result due to Loïc Merel). However, there are less precise classification results for $E(F)[\text{tors}]$ à la the work of Mazur. Full results are only known when $F/\mathbb{Q}$ is trivial, quadratic or cubic.

The rank of an elliptic curve. The torsion group of an elliptic curve over $\mathbb{Q}$ is "completely known," in the sense of Mazur's theorem, and their sizes are bounded by 16. What do we know about the rank r?

In contrast to torsion groups over $\mathbb{Q}$, **it is unknown whether ranks are bounded over $\mathbb{Q}$.** Every couple of years, a new elliptic curve gets discovered with a higher rank than the last known largest rank.

Here's a history of discovered ranks (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).

It is not known whether there exist elliptic curves over $\mathbb{Q}$ with a rank higher than 28, see https://web.math.pmf.unizg.hr/ duje/tors/rankhist.html for sources ("history of elliptic curves rank records"). It's hard to come up with these. Elkies' example for rank 28 was the following (written in "general Weierstrass form"):

$$E : y^2 + xy + y = x^3 - x^2 - 20067762415575526585033208209338542750930230312178956502x$$
$$+ 3448161179503055646703298569039072037485594435931918036126600829629193944873224342 9$$

(57 and 83 digits, respectively). To show this rank was at least 28, Elkies constructed 28 linearly independent rational points of infinite order (all equally complicated).

Fermat's last theorem. To wrap this chapter on Diophantine equations up, let's talk about one of the most important results in mathematics.

Theorem 5.7.5 (Fermat's last theorem)**.** *There are no positive integral solutions to the equation*

$$x^n + y^n = z^n$$

when $n \geq 3$.

When $n = 1$, $x+y = z$ is a linear equation, which is very easy to come up with integral solutions for. When $n = 2$, $x^2 + y^2 = z^2$ has integral solutions given by Pythagorean triples, which we've already covered.

However, $n \geq 3$ is new. 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.

This was proven properly 300+ years later, in 1994. The proof uses high level arithmetic geometry, which would take years of study to understand. A key idea in the proof is to show that an integral solution (a, b, c) to $x^n + y^n = z^n$ implies a certain elliptic curve $E_{a,b,c}$ has properties which cannot exist.

Let's explore this a bit. Suppose for contradiction there is a positive integral solution:

$$a^n + b^n = c^n,$$

with $a, b, c \in \mathbb{Z}^+$. Define a *Frey curve* as

$$E_{a,b,c} : y^2 = x(x - a^n)(x - b^n).$$

In 1986, Ken Ribet showed $E_{a,b,c}$ is **never modular.** However, in 1994 Andrew Wiles (building on many others' work) showed that an elliptic curve $E : y^2 = x(x-A)(x-B)$ is **always modular** when $A, B \in \mathbb{Q}$. Contradiction!

What does modular mean? That is something to learn in a second number theory course! *Fin.*

Exercise 5.7.1. For the elliptic curve

$$E_{/\mathbb{Q}} : y^2 = x^3 + 17,$$

given its points $P_1 := (-2, 3)$, $P_2 := (-1, 4)$ and $P_3 := (2, 5)$, prove the following.

a) $-2P_1 = (8, 23)$.
b) $P_2 \oplus P_3 = \left(-\frac{8}{9}, -\frac{109}{27}\right)$ (which is $\approx (-0.889, -4.037)$).

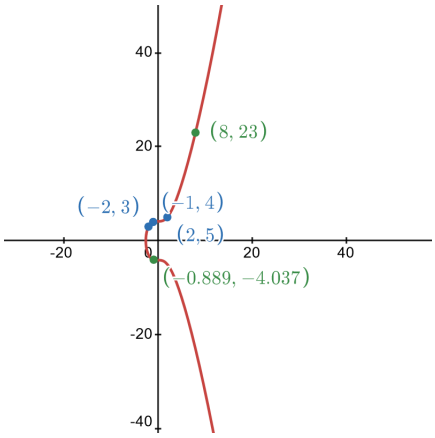


FIGURE 12. The elliptic curve $E : y^2 = x^3 + 17$.

Exercise 5.7.2. This exercise concerns the arithmetic of the elliptic curve $E : y^2 = x^3 - 5x + 1$.

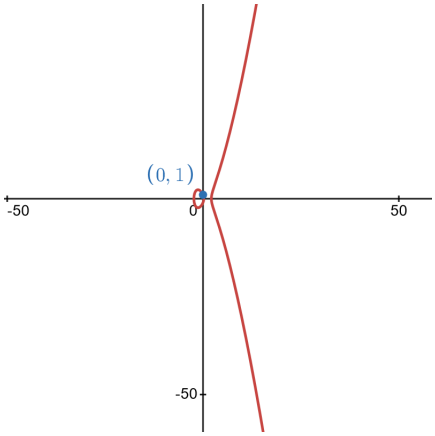


FIGURE 13. The elliptic curve $E : y^2 = x^3 - 5x + 1$.

a) We have the point $P = (0, 1) \in E(\mathbb{Q})$. Show that $2P := P \oplus P = \left(\frac{25}{4}, -\frac{117}{8}\right)$.
b) Suppose that $\alpha \in \mathbb{R}$ satisfies

$$\alpha^3 - 5\alpha + 1 = 0.$$

Then $Q := (\alpha, 0) \in E(\mathbb{R})$. Show that $2Q := Q \oplus Q = O$, where $O := [0 : 1 : 0]$.

Exercise 5.7.3 (An elliptic curve not in Weierstrass form)**.** This exercise explores some arithmetic with elliptic curves not in Weierstrass form.

Consider the cubic curve

$$E : x^3 + y^3 = 1.$$

a) Write down the homogenization E_H of E. Prove that that $O := [1 : -1 : 0]$ is the only point at infinity on E_H.

b) Assume that O is a nonsingular inflection point on E_H. Show that E_H is nonsingular. Thus, E_H is a *projective* elliptic curve. (*Hint:* de-homogenize E_H to end up with E again, and check E for singular points.)

c) Assume that $x^3 + y^3 - 1$ is irreducible over $\mathbb{Q}$; thus, E is an elliptic curve over $\mathbb{Q}$. Prove that for any point $P = (a, b) \in E(\mathbb{R})$ with $a \neq b$, the inverse of P is

$$-P = (b, a).$$

(*Hint:* what *three* points does the line through (a, b) and (b, a) pass through on E?)

d) (Extra credit) Explain why E has no positive rational solutions.

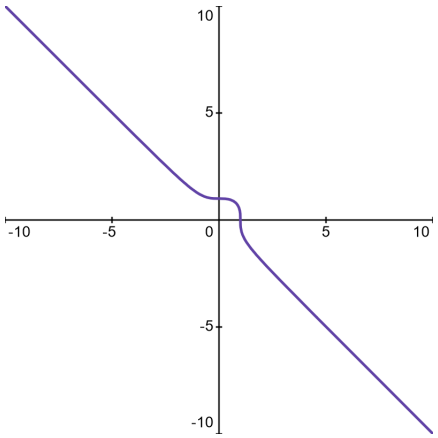


FIGURE 14. The elliptic curve $E : x^3 + y^3 = 1$.

Exercise 5.7.4. This exercise investigates the behavior of the number of points of the elliptic curve

$$E : y^2 = x^3 + x$$

modulo primes $\ell \in \mathbb{Z}^+$. 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 $\ell \in \mathbb{Z}^+$, we will write $\mathbb{F}_\ell := \mathbb{Z}/\ell\mathbb{Z}$. We will also use $E(\mathbb{F}_\ell)$ to denote the set of points on E modulo ℓ. (*Reminder:* we always include $O := [0 : 1 : 0]$ as a point in $E(\mathbb{F}_\ell)$.)

a) For primes $\ell = 3, 7, 11$, explicitly compute by hand the set of points $(x_0, y_0) \in \mathbb{F}_\ell^2$ with $y_0^2 \equiv x_0^3 + x_0 \pmod{\ell}$.

b) Prove that for any prime $\ell \equiv 3 \pmod 4$, one has

$$|E(\mathbb{F}_\ell)| = \ell + 1.$$

(*Hint:* if $y_0^2 \equiv x_0^3 + x_0 \pmod{\ell}$, then $x_0^3 + x_0$ is a square modulo ℓ. However, -1 is not a quadratic residue modulo ℓ since $\ell \equiv 3 \pmod 4$.)

c) (Extra credit) Create a program that does the following: given a prime $\ell \in \mathbb{Z}^+$ and an elliptic curve $E : y^2 = x^3 + Ax + B$ with $-16(4A^3 + 27B^2) \not\equiv 0 \pmod{\ell}$, it returns the set of point $E(\mathbb{F}_\ell)$, as well as the size $|E(\mathbb{F}_\ell)|$ (including $[0 : 1 : 0]$).

What patterns do you spot for the size of $E(\mathbb{F}_\ell)$, $E : y^2 = x^3 + x$, when $\ell \equiv 1 \pmod 4$? Based off your calculations, make a conjecture on the size.

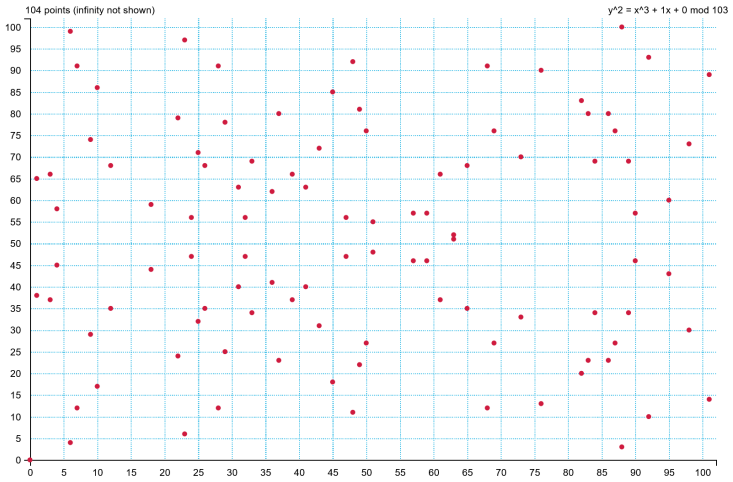


FIGURE 15. The elliptic curve $E : y^2 = x^3 + x$ modulo 103.

Exercise 5.7.5. Fix an elliptic curve over $\mathbb{Q}$ in Weierstrass form

$$E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B.$$

Then the *torsion subgroup of E over $\mathbb{Q}$,* denoted $E(\mathbb{Q})[\text{tors}]$, is the subgroup of $E(\mathbb{Q})$ of points with finite order:

$$E(\mathbb{Q})[\text{tors}] = \{P \in E(\mathbb{Q}) : NP = O \text{ for some } N \in \mathbb{Z}^+\}.$$

(Note that we include $O \in E(\mathbb{Q})[\text{tors}]$.)

This exercise explores the torsion points on E of order two.

a) Show that $P \in E$ has order two if and only if

$$P = (\alpha, 0)$$

where α is a root of $x^3 + Ax + B$.

b) As it turns out, for any elliptic curve $E_{/\mathbb{Q}}$, one has that $E(\mathbb{Q})[\text{tors}]$ is a finite abelian group. With this in mind, prove the following: assume that E has the equation $E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B$. Then if $x^3 + Ax + B$ is a reducible polynomial over $\mathbb{Q}$, then the size of $E(\mathbb{Q})[\text{tors}]$ is even.

Exercise 5.7.6. This exercise proves some basic results for elliptic curves in (short) Weierstrass form,

$$E_{/\mathbb{Q}} : y^2 = x^3 + Ax + B.$$

a) Show that E has exactly one point at infinity.

b) Show that for any point $P := (a, b) \in E(\mathbb{R})$, one has the additive inverse

$$-P = (a, -b).$$

(*Hint:* the collinearity theorem might help.)

Exercise 5.7.7. 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 solution. 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.

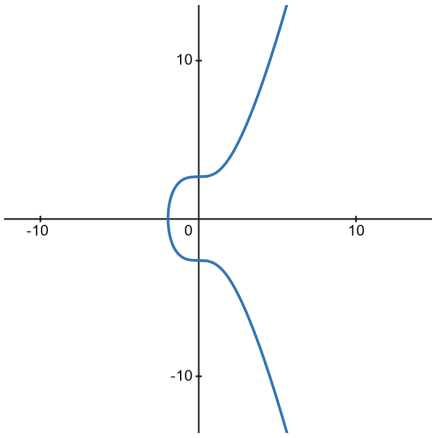


FIGURE 16. The elliptic curve $E : y^2 = x^3 + 7$.

Bonus Exercise 5.7.8. This exercise will explore the concept of the *genus* of a plane curve.

Suppose that $f(x, y) \in \mathbb{Z}[x, y]$ 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 $g = 0$, C has either zero or infinitely many rational points; for example, conics are genus zero curves. When $g = 1$, 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 curves defined by the following equations have a finite or infinite amount of 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.

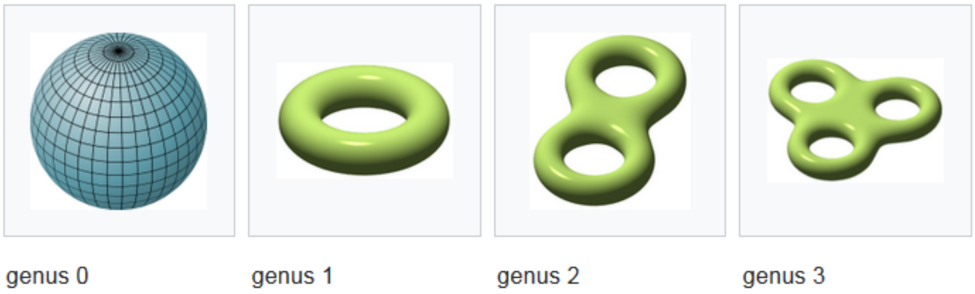


FIGURE 17. Pictures of g-holed tori in complex projective space, cf. Wikipedia.

Bonus Exercise 5.7.9. 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*

$$> \text{rad}\left(\cdot \cdot \right)^{1+} .$$

The goal of this exercise is to prove the :)-theorem. Good luck!

REFERENCES

[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).