Introduction to Number Theory

Tim Smits

June 13, 2023

Contents

Introduction

These notes are from the math 111 course I taught at UCLA in Spring 2023, which serves as an introduction to number theory intended for junior and senior mathematics majors. The prerequisite knowledge is minimal exposure to abstract algebra: familiarity with the definitions of a group, a ring, an ideal, a field, and a homomorphism is sufficient background for almost everything in these notes. The basics are contained in the appendix for those who need it. Additionally, it is assumed the reader is familiar with linear algebra in finite dimensional vector spaces. There may be many typos. Please let me know if any are found!

Number theory is one of the oldest branches of mathematics, with many fundamental results dating back to the ancient Greeks. At its core, number theory studies the *integers*. It might be hard to believe that this could be particularly hard, or even interesting. However, number theory is an *incredibly* deep, and rich area of mathematics. Here are a two reasons why studying number theory is worthwhile:

- Number theory is a *central* branch of mathematics, meaning that it heavily utilizes the three "core" branches of mathematics: algebra, analysis, and geometry. At an undergraduate level, many of the core courses are "disjoint" from each other. A first time learner of abstract algebra is not going to see any connections to real analysis, and in a first course in real analysis, students are unlikely to see the connections with geometry and topology. With number theory, however, it's much easier to see the interplay between the different branches of mathematics at an introductory level. In particular, the study of Diophantine equations bridges ideas from algebra (unique factorization), and geometry (rational points on curves).

  From a modern perspective, there are many objects that come from "algebraic", "analytic", and "geometric" ideas. One of the most important examples of this arises in the study of *elliptic curves*, where there is a geometric object (the elliptic curve E), an algebraic object (the group of rational points $E(\mathbb{Q})$), and an analytic object (an L-function attached to E). Much of modern number theory revolves around trying to unify these three different perspectives. No matter what your mathematical interests are, there are applications of it to number theory!

- As a consequence of number theory being such an old branch of mathematics, it's a subject with a very interesting history, that motivated a lot of the development of mathematics. In particular, one of the most famous problems is *Fermat's last theorem*, which says there are no non-trivial integer solutions to $x^n + y^n = z^n$ for $n \geq 3$. When $n = 2$, solutions to $x^2 + y^2 = z^2$ correspond to *Pythagorean triples*, which were

studied by the Greek mathematician Pythagoras. Fermat famously claimed that he had a proof of his theorem, which he did not write down. Fermat's last theorem went unproven for over 350 years, and it captured the interest of many historically prominent mathematicians, who made significant advancements to the field of number theory.

For those interested, here is a list of some of the mathematicians whose ideas were fundamental to the development of the topics that will be covered in these notes:

- Brahmagupta, Euclid, Pythagoras, Diophantus
- Fermat, Euler, Lagrange, Legendre, Gauss
- Dirichlet, Kummer, Dedekind, Hensel

Part I

Unique Factorization

Chapter 1

Unique factorization in $\mathbb{Z}$

1.1 Key properties of $\mathbb{Z}$

The set $\mathbb{Z}$ of integers comes with two binary operations, addition ($+$) and multiplication ($\cdot$), meaning that for any $a, b \in \mathbb{Z}$ we get well-defined integers $a + b$ and $a \cdot b$. We will take the construction of the integers and the formal definition of these operations for granted, but these operations satisfy the following properties for any $a, b, c \in \mathbb{Z}$:

1. (*Commutativity*) $a + b = b + a$ and $ab = ba$.
2. (*Associativity*) $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$.
3. (*Distributivity*) $a(b + c) = ab + ac$.
4. (*Zero*) There exists $0 \in \mathbb{Z}$ such that $a + 0 = a$.
5. (*Negatives*) There exists $-a \in \mathbb{Z}$ such that $a + (-a) = 0$.
6. (*One*) There exists $1 \in \mathbb{Z}$ such that $a \cdot 1 = a$.

More compactly, addition and multiplication are both associative and commutative, multiplication distributes over addition, and there are special elements 0 and 1 that don't "change" an integer a with respect to the relevant operation. Addition has the extra property that integers have additive inverses, which give rise to negative numbers.

Sitting inside of $\mathbb{Z}$ is a non-empty subset $P \subset \mathbb{Z}$ with the following properties:

1. For any $a, b \in P$, $a + b \in P$.
2. For any $a, b \in P$, $a \cdot b \in P$.
3. $0 \notin P$.

4. (*Trichotomy*) For all $a \in \mathbb{Z}$, exactly one of the following holds: $a \in P$, $0 \in P$, or $-a \in P$.

This subset P is, of course, the positive integers. The existence of this subset P allows us to define an ordering on $\mathbb{Z}$, by defining $a < b \iff a - b \in P$. The ordering of the integers gives rise to one of its most basic properties:

Proposition 1 (Zero product property)**.** *Let $a, b \in \mathbb{Z}$ with $ab = 0$. Then $a = 0$ or $b = 0$.*

Proof. We prove the contrapositive, that if $a \neq 0$ and $b \neq 0$ then $ab \neq 0$. First suppose that $b > 0$. Since multiplication of integers is defined via repeated addition, i.e. $a \cdot b = \underbrace{a + a + \ldots + a}_{b \text{ times}}$, then $ab \geq a$ or $ab \leq a$ depending on the sign of a. If $b < 0$, do the same thing with $(-a)(-b)$. □

The integers are very special, because not only can they be ordered, but they can be *well-ordered*.

Theorem 1.1.1 (Well-ordering principle)**.** *Let $S \subset \mathbb{N}$ be non-empty. Then S has a smallest element with respect to $<$.*

The well-ordering principle will be very important for proving key theorems that hold in the integers. One can show that the well-ordering principle is equivalent to the principle of mathematical induction, and so many of the proofs that follow could be formulated with induction instead. We leave the proof of their equivalency in the appendix.

1.2 Divisibility

Definition 1.2.1. We say for integers a, b that a **divides** b if there is an integer k such that $b = ak$, and write this as $a \mid b$.

Example 1.2.2. We have $2 \mid 10$, because $102 \cdot 5$, $(-7) \mid 49$, because $49 = (-7) \cdot (-7)$, and $5 \mid 0$ because $0 = 5 \cdot 0$. Clearly $\pm 1 \mid n$ for any integer n, however, $0 \nmid n$ for any non-zero integer n, because there is no integer k with $0 \cdot k = n$.

From the definition of divisibility, we have the following basic properties:

Proposition 2. *Let $a, b, c \in \mathbb{Z}$.*

1. *If $a \mid b$ then $a \mid b\ell$ for any $\ell \in \mathbb{Z}$.*
2. *If $a \mid b$ and $b \mid c$, then $a \mid c$.*
3. *If $a \mid b$ and $a \mid c$ then $a \mid bx + cy$ for any $x, y \in \mathbb{Z}$.*
4. *If $a \mid b$ for $b \neq 0$, then $|a| \leq |b|$.*

Proof. These follow more or less immediately from the definition and are left as an exercise. □

One of the first things one learns about the integers in grade school is that it's possible to perform *division with remainder*:

Theorem 1.2.3 (Division Algorithm)**.** *Let a and b be integers with $b > 0$. Then there exist unique integers q and r such that $a = bq + r$ with $0 \leq r < b$.*

Proof. Let $S = \{a - bq > 0 : q \in \mathbb{Z}\}$. Then S is non-empty: this is clear if $a > 0$, and if $a < 0$, then $a - (a-1)b = a(1-b) + b > 0$. By the well-ordering principle, S contains a smallest positive element, say r, which we can write as $a - bq = r$ for some q. This then gives $a = bq + r$. First we show that $0 \leq r < b$. If $r > b$, then $r = b + (r - b)$, and $(r - b) > 0$. so $a = b(q+1) + (r - b)$ with $0 < r - b < r$. This contradicts the minimality of r, so $0 \leq r < b$. This proves the existence of such q and r, so it remains to show the uniqueness of q and r.

Suppose that $a = bq + r$ and $a = bq' + r'$ with $0 \leq r < b$ and $0 \leq r' < b$. Without loss of generality, suppose that $r \leq r'$. Then from $bq + r = bq' + r'$ we get $b(q - q') = (r' - r)$. This says $(r' - r)$ is a multiple of b, and since $0 \leq r' - r < b$ this says $r' - r = 0$, i.e. $r = r'$. It's then immediate that $q = q'$ so uniqueness follows. □

Example 1.2.4. With $a = 17$ and $b = 3203$, we have $3203 = 17 \cdot 188 + 7$. This can be determined from the usual grade school algorithm of long division.

It's possible to extend the division algorithm to allow negative dividends. The proof is very similar, and we leave it as an exercise.

Theorem 1.2.5 (Extended Division Algorithm)**.** *Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Then there exist unique integers q, r such that $a = bq + r$ and $0 \leq r < |b|$.*

Proof. Exercise. □

Definition 1.2.6. The **greatest common divisior** of two integers a and b is the largest positive integer d such that:

1. $d \mid a$ and $d \mid b$
2. If $c \mid a$ and $c \mid b$, then $c \leq d$.

We write $\gcd(a, b)$ or (a, b) to denote the greatest common divisor. If $(a, b) = 1$, we say that a and b are **relatively prime**.

If d is any divisor of an integer a, then property 4 of proposition 2 says there are only finitely many possibilities for d, so it makes sense to speak of a *largest* one. If $a = 0$, then all integers are divisors of a, so $(0, a) = |a|$. Note that $(0, 0)$ is therefore undefined.

Example 1.2.7. The greatest common divisor of $a = 315$ and $b = 195$ is $d = 15$, which is easily found by factoring.

The greatest common divisor of two integers a and b has the important property that it's an integer linear combination of a and b:

Theorem 1.2.8 (Bezout's lemma)**.** *Let $a, b \in \mathbb{Z}$ and set $d = \gcd(a, b)$. Then there exist integers x and y such that $ax + by = d$.*

Proof. Let $S = \{ax + by > 0 : x, y \in \mathbb{Z}\}$, the set of all integer linear combinations of a and b that are positive. Notice that S is non-empty, because either a or $-a$ is positive and $a \in S$. By the well-ordering principle, S has a least positive element, say d', so there are integers x and y such that $ax + by = d'$. We will show that $d = d'$.

Since $d \mid a$ and $d \mid b$ by definition, $d \mid (ax + by)$ so $d \mid d'$, and therefore $d \leq d'$. To show that $d' \leq d$, we will show that d' divides every element of S. Let $s \in S$, so $s = ax' + by'$ for some $x', y' \in \mathbb{Z}$. By the division algorithm, we can write $s = d'q + r$ for unique q and r with $0 \leq r < d'$. Then $r = s - d'q = ax' + by' - (ax + by)q = a(x' - xq) + b(y' - yq)$. However, $r < d'$ and d' is the smallest positive element of S, so this forces $r = 0$. Therefore, $s = d'q$ so $d' \mid s$ for any $s \in S$. In particular, $a, b \in S$ so $d' \mid a$ and $d' \mid b$, so d' is a common divisor of a and b, so $d' \leq d$ by definition of d. □

Note that we not only proved that the greatest common divisor is an integer linear combination of a and b, but that it's the *smallest positive* integer linear combination of a and b. In particular, this means that if $ax + by = 1$ for some integers x, y, then we must have $\gcd(a, b) = 1$.

1.3 The Euclidean algorithm

In the example in the previous section, it was very easy to factor the given numbers to compute their greatest common divisior. In general, factoring is a very hard problem.

Example 1.3.1. Let $a = 1002001$ and $b = 379427895$. Then $a = 7^2 \cdot 11^2 \cdot 13^2$ and $b = 3^4 \cdot 5 \cdot 7 \cdot 11 \cdot 23^3$, so we have $\gcd(a, b) = 7 \cdot 11 = 77$.

This example would already be hard to do by hand, and if the integers are large enough, even a computer will have trouble factoring them. The following algorithm (of Euclid!) is a significantly more efficient way of doing this computation. Before we state and give the proof, we need a lemma about greatest common divisors.

Lemma 1.3.2. *Let $a, b \in \mathbb{Z}$ such that $a = bq + r$ for some integers q, r. Then $gcd(a, b) = gcd(b, r)$.*

Proof. Let $d = \gcd(a, b)$. Then $d \mid a$ and $d \mid b$ so $d \mid a - bq$, which means $d \mid r$. This says d is a common divisor of b, r, so $d \leq \gcd(b, r)$. On the other hand, set $d' = \gcd(b, r)$. We have $d' \mid b$ and $d' \mid r$ so $d' \mid bq + r$, so that $d' \mid a$, so d' is a common divisor of a, b so $d' \leq \gcd(a, b)$. Combining the two inequalities says $d' = d$. □

Theorem 1.3.3 (Euclidean algorithm)**.** *Let a, b be non-zero integers. Repeatedly carry out the division algorithm as follows:*

$$\begin{aligned} a &= bq_1 + r_1, \quad 0 \leq r_1 < |b| \\ b &= r_1 q_2 + r_2, \quad 0 \leq r_2 < r_1 \\ r_1 &= r_2 q_3 + r_3, \quad 0 \leq r_3 < r_2 \\ &\vdots \end{aligned}$$

The last non-zero remainder is the greatest common divisor of a and b.

Proof. The proof has two steps: first, we show that the algorithm eventually terminates, so that there are finitely many remainders. Then, we show that the last non-zero remainder is indeed $\gcd(a, b)$. Consider the set of remainders $\{r_n\}$. Then $\{r_n\}$ is a strictly decreasing set of positive integers, and so by the well-ordering principle, there must be some N such that $n \geq N$ means that $r_n = 0$. To see this, suppose otherwise, that all terms of the sequence are strictly positive. Then by the well-ordering principle, $\{r_n\}$ has a least positive element, say L. Therefore, there is N such that $r_N = L$. However, since $\{r_n\}$ is a strictly decreasing sequence, we have $r_{N+1} < r_N < L$, contradicting the definition of L. Therefore, there are finitely many non-zero remainders output by the algorithm, call them $r_1, \ldots, r_N$. Repeatedly applying the above lemma, we find $\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \ldots = \gcd(r_{N-1}, r_N) = \gcd(r_N, 0) = r_N$. □

Example 1.3.4. Let $a = 75$ and $b = 45$. We have

$$\begin{aligned} 75 &= 45 \cdot 1 + 30 \\ 45 &= 30 \cdot 1 + 15 \\ 30 &= 15 \cdot 2 + 0 \end{aligned}$$

So that $(75, 45) = 15$.

Example 1.3.5. Let $a = 517$ and $b = 89$. We have

$$\begin{aligned} 517 &= 89 \cdot 5 + 72 \\ 89 &= 72 \cdot 1 + 17 \\ 72 &= 17 \cdot 4 + 4 \\ 17 &= 4 \cdot 4 + 1 \\ 4 &= 4 \cdot 1 + 0 \end{aligned}$$

So that $(517, 89) = 1$.

Example 1.3.6. Let $a = 379427895$ and $b = 1002001$. We have

$$\begin{aligned} 379427895 &= 1002001 \cdot 378 + 671517 \\ 1002001 &= 671517 \cdot 1 + 330484 \\ 671517 &= 330484 \cdot 2 + 10549 \\ 330484 &= 10549 \cdot 31 + 3465 \\ 10549 &= 3465 \cdot 3 + 154 \\ 3465 &= 154 \cdot 22 + 77 \\ 154 &= 77 \cdot 2 + 0 \end{aligned}$$

So that $(379427895, 1002001) = 77$.

From the above example, even though the numbers were moderately large, the Euclidean algorithm finished rather quickly. In general, the algorithm is very efficient. It runs in $O(\log(\min(a, b))$ time! The Euclidean algorithm is not only useful for computing the greatest common divisor; we can use it to tell us *what* integer linear combination of a and b to take to get the greatest common divisor!

Example 1.3.7. Let $a = 517$ and $b = 89$. We have

$$\begin{aligned} 517 &= 89 \cdot 5 + 72 \\ 89 &= 72 \cdot 1 + 17 \\ 72 &= 17 \cdot 4 + 4 \\ 17 &= 4 \cdot 4 + 1 \end{aligned}$$

Which tells us that $(517, 89) = 1$. We now back substitute to solve for 1 in terms of the previous lines. We have

$$\begin{aligned} 1 &= 17 - 4 \cdot 4 \\ &= 17 - 4 \cdot (72 - 17 \cdot 4) = 17 \cdot 17 - 4 \cdot 72 \\ &= 17 \cdot (89 - 72 \cdot 1) - 4 \cdot 72 = 89 \cdot 17 - 72 \cdot 21 \\ &= 89 \cdot 17 - (517 - 89 \cdot 5) \cdot 21 = 89 \cdot 122 - 517 \cdot 21 \end{aligned}$$

This says the integers we are looking for in Bezout's lemma are $x = -21$ and $y = 122$.

In particular, the Euclidean Algorithm provides an *alternate* proof of Bezout's lemma that doesn't use the well-ordering principle: run the Euclidean algorithm on the pair (a, b) and then perform back substitution to construct x, y such that $ax + by = \gcd(a, b)$.

1.4 Linear Diophantine equations

A *Diophantine equation* is an equation, typically a polynomial equation, with integer coefficients. Historically, Diophantine equations were studied by the Greek mathematician Diophantus. Many problems in number theory can be translated into some question about finding integer solutions to Diophantine equations, e.g. Pythagorean triples correspond to integer solutions to the Diophantine equation $x^2 + y^2 = z^2$.

We begin with some useful properties of the greatest common divisor that we'll frequently use.

Proposition 3. *Let $a, b, c \in \mathbb{Z}$.*

(a) $gcd(ka, kb) = k \cdot gcd(a, b)$ for any integer $k > 0$.

(b) Let $g = gcd(a, b)$. Then $gcd(\frac{a}{g}, \frac{b}{g}) = 1$.

(c) If $gcd(a, c) = 1$ and $gcd(b, c) = 1$ then $gcd(ab, c) = 1$.

Proof.

(a) We have $\gcd(a,b) = \min\{ax+by > 0 : x,y \in \mathbb{Z}\}$, the smallest positive integer linear combination of a and b. From this, it's then quite clear that $k \cdot \min\{ax+by > 0 : x,y \in \mathbb{Z}\} = \min\{(ka)x+(kb)y > 0 : x,y \in \mathbb{Z}\} = \gcd(ka,kb)$.

(b) Let $g = \gcd(a,b)$. Then $g = \gcd(\frac{a}{g} \cdot g, \frac{b}{g} \cdot g) = g \cdot \gcd(\frac{a}{g}, \frac{b}{g})$ by (a). Since $g > 0$, this means $\gcd(\frac{a}{g}, \frac{b}{g}) = 1$.

(c) By Bezout's lemma, there exist integers x, y such that $ax+cy = 1$ and x', y' such that $bx'+cy' = 1$. Multiplying these two equations together, $(ax+cy)(bx'+cy') = 1$, so $ab(xx') + c(axy' + bx'y + cyy') = 1$. This says 1 is an integer linear combination of ab and c, so that $\gcd(ab,c) = 1$.

□

Proposition 4. *Let $a,b,c \in \mathbb{Z}$ with $a \mid bc$ and $\gcd(a,c) = 1$. Then $a \mid b$.*

Proof. Since $\gcd(a,c) = 1$, by Bezout's lemma there are integers x, y such that $ax+cy = 1$. Multiplying by b says $abx+bcy = b$. Since $a \mid bc$, then $a \mid abx+bcy$, i.e. $a \mid b$. □

We will now give a complete classification of integer solutions to linear Diophantine equations.

Theorem 1.4.1. *Consider the equation $ax+by = c$ for $a,b,c \in \mathbb{Z}$.*

(a) A solution exists in $\mathbb{Z}^2$ if and only if $\gcd(a,b) \mid c$.

(b) If $(x_0,y_0) \in \mathbb{Z}^2$ is one solution, then all solutions are of the form $(x_0+b'k, y_0-a'k)$ for some $k \in \mathbb{Z}$, where $a' = \frac{a}{\gcd(a,b)}$ and $b' = \frac{b}{\gcd(a,b)}$.

Proof.

(a) Suppose that $(x,y) \in \mathbb{Z}^2$ is a solution to the equation. Then $\gcd(a,b)$ divides both a and b, so $\gcd(a,b) \mid ax+by$, which says $\gcd(a,b) \mid c$. Conversely, suppose that $\gcd(a,b) \mid c$. Write $c = \gcd(a,b)k$ for some integer k. By Bezout's lemma, there exist integers x, y such that $ax+by = \gcd(a,b)$. Multiplying by k, we have $a(kx)+b(ky) = c$, so that (kx,ky) is a solution.

(b) Suppose that (x,y) is a second solution to the equation, so that $ax_0+by_0 = c$ and $ax+by = c$. Equating, we find $a(x-x_0) = b(y_0-y)$, so dividing through by $\gcd(a,b)$ we get $a'(x-x_0) = b'(y_0-y)$. Since $a' \mid b'(y_0-y)$ and $\gcd(a',b') = 1$, this means $a' \mid y_0-y$ so we can write $y_0-y = ka$ for some integer k. Similarly, we can write $x-x_0 = b'\ell$ for some integer ℓ. Substituting back in, this means that $a'b'\ell = a'b'k$, so that $\ell = k$. This says $(x,y) = (x_0+b'k, y_0-a'k)$ for some integer k, so any solution must be of this form. Finally, it's clear by a simple substitution that any pair $(x_0+b'k, y_0-a'k)$ *is* a solution to $ax+by = c$, so that these constitute all possible solutions.

□

Example 1.4.2. Suppose we want to find all integer solutions to the equation $37x+47y = 22$. First, we run the Euclidean algorithm:

$$\begin{aligned} 47 &= 37 \cdot 1 + 10 \\ 37 &= 10 \cdot 3 + 7 \\ 10 &= 7 \cdot 1 + 3 \\ 7 &= 3 \cdot 2 + 1 \end{aligned}$$

So that $\gcd(47, 37) = 1$. Performing back substitution, we find that $(14, -11)$ is one solution to the equation $37x + 47y = 1$. Multiplying by 22, we have $(308, -242)$ is a solution to the equation $37x + 47y = 22$. The above theorem then says that the solution set to this equation is $\{(308 + 47k, -242 - 37k) : k \in \mathbb{Z}\}$.

1.5 Unique factorization in $\mathbb{Z}$

Definition 1.5.1. An integer $p > 1$ is called **prime** if the only positive divisors of p are 1 and p.

The sequence of prime numbers starts off as $2, 3, 5, 7, 11, \ldots$. Primes are the "building blocks" of the integers, in the sense that all integers are constructed from primes. An alternate characterization of primes is the following:

Proposition 5 (Euclid's lemma)**.** *Let $p > 1$. Then p is prime if and only if for any integers a, b, $p \mid ab \implies p \mid a$ or $p \mid b$.*

Proof. First, suppose that p is prime. Let $p \mid ab$, and suppose that $p \nmid a$. Then $(a, p) = 1$ by definition of p being prime, so by Bezout's lemma, there are integers x, y with $px + ay = 1$. Multiplying by b says $pbx + aby = b$, and since $p \mid ab$, we have $p \mid pbx + aby$, so $p \mid b$.

Conversely, suppose that $p > 1$ is an integer with the property that $p \mid ab \implies p \mid a$ or $p \mid b$. Let d be a positive divisor of p. Then we can write $p = dk$ for some positive integer k. Certainly, $p \mid p$, so $p \mid dk$ means that $p \mid d$ or $p \mid k$. However, since $1 \leq d, k \leq p$, the only way this is possible is if $d = 1$ or $d = p$, so that the only positive divisors of p are 1 and p. □

This characterization of primes will be the key to proving unique factorization.

Theorem 1.5.2 (Fundamental Theorem of Arithmetic)**.** *Let $n > 1$ be an integer. Then there exist unique primes $p_1, \ldots, p_k$ and unique positive integers $e_1, \ldots, e_k$ such that $n = p_1^{e_1} \cdots p_k^{e_k}$. That is, every integer has a unique factorization (up to order of factors) into a product of primes.*

Proof. There are two parts to the proof. First, we show that we can write every integer $n > 1$ as *some* product of prime numbers, and then we will show that such a choice of primes are *unique*. Both of these statements will be proven using strong induction.

Existence:

Note that $n = 2$ is a prime. Now suppose for some k that the integers $2, 3, \ldots, k$ can

be written as a product of primes. Consider the integer $k+1$: if it is prime, we are done. Otherwise, $k+1$ is not prime, so by definition it has a non-trivial positive divisor. Write $k+1 = ab$ for some integers a and b. Necessarily, $1 < a, b < k+1$, so in particular a and b are integers between 2 and k. By induction hypothesis, both a and b can be written as a product of primes, and therefore $k+1$ is a product of primes as well. By induction, we then see that every integer $n > 1$ is a product of primes.

Uniqueness:

Note that 2 is a prime so it's a product of primes in a unique way. Now suppose that for some k, we have that $2, 3, \ldots, k$ all have a factorization using a unique set of primes. If $k+1$ is prime, again we are done. Otherwise, suppose that we can write $k+1 = p_1 \cdots p_m = q_1 \cdots q_\ell$ for some primes p_i and q_j. Then $p_1 \mid q_1 \cdots q_\ell$, so by inductively applying Euclid's lemma, one finds $p_1 \mid q_j$ for some j, and since q_j are prime, this says $p_1 = q_j$ for some j. By reordering the factors as necessary, assume $p_1 = q_1$. Cancelling p_1 from both sides, we have $a = p_2 \cdots p_m = q_2 \cdots q_\ell$. However, $1 < a < k$ so by assumption, a has unique factorization, i.e. $m = \ell$ and $p_i = q_i$ for all $2 \leq i \leq m$ (after reordering if necessary). Since $k+1 = ap_1$, and $p_1 = q_1$, this shows $k+1$ has unique factorization as desired. By induction, every $n > 1$ has unique factorization. Collecting terms of the same prime together shows that n is of the form listed in the statement and the uniqueness of the exponents is immediate. □

Another way of thinking about the statement of unique factorization is the following:

Corollary 1.5.3. *For any integer $n > 1$, we can write $n = \prod_p p^{e_p}$ with $e_p \geq 0$, and $e_p > 0$ for finitely many primes p.*

This point of view allows one to view an integer as an infinite tuple $(e_2, e_3, \ldots)$ where only finitely many e_p are non-zero, the point being that the data of an integer is given by the data of the exponents of each prime in $\mathbb{Z}$.

We now explore some consequences of the Fundamental Theorem of Arithmetic.

Definition 1.5.4. The **least common multiple** of two integers a, b, denoted $\text{lcm}(a, b)$ is the unique positive integer ℓ with the following properties:

1. $a \mid \ell$ and $b \mid \ell$.
2. For any positive m such that $a \mid m$ and $b \mid m$, we have $\ell \leq m$.

For any two integers a and b, we can always factor them into a *common set* of primes by allowing exponents to be 0. This leads to another computation of the greatest common divisor, and a computation of the least common multiple.

Proposition 6. *Let $a = p_1^{e_1} \cdots p_k^{e_k}$ and $b = p_1^{f_1} \cdots p_k^{f_k}$ be prime factorizations of a and b into a common set of primes. Then $gcd(a, b) = p_1^{\min\{e_1, f_1\}} \cdots p_k^{\min\{e_k, f_k\}}$ and $lcm(a, b) = p_1^{\max\{e_1, f_1\}} \cdots p_k^{\max\{e_k, f_k\}}$.*

Proof. We prove the first statement, and leave the second as an exercise as it is very similar. Let $d = \gcd(a, b)$. Then since d divides both a and b, we can write $d = p_1^{t_1} \cdots p_k^{t_k}$ for some non-negative integers t_i where $t_i \leq e_i$ and $t_i \leq f_i$. Therefore, $t_i \leq \min\{e_i, f_i\}$ for all i. It's clear that setting $t_i = \min\{e_i, f_i\}$ for all i gives us a divisor of a and b, which is therefore the greatest common divisor because the size of each exponent has been maximized. □

Corollary 1.5.5. *Let $a, b > 0$ be integers. Then $ab = gcd(a, b) \cdot lcm(a, b)$.*

Proof. The prime factorization of ab is given by multiplying the prime factorizations of a and b together. Writing $a = p_1^{e_1} \cdots p_k^{e_k}$ and $b = p_1^{f_1} \cdots p_k^{f_k}$, the exponent of p_i in ab is given by $e_i + f_i = \min\{e_i, f_i\} + \max\{e_i, f_i\}$ from which the result follows. □

The above relation gives an efficient way to compute the least common multiple of two integers! There is no "Bezout-like" relation for the least common multiple like there is for the greatest common divisor, but one has $\text{lcm}(a, b) = \frac{ab}{\gcd(a,b)}$. The Euclidean Algorithm can efficiently compute $\gcd(a, b)$, and therefore can be used to efficiently compute $\text{lcm}(a, b)$.

Proposition 7. *For any integer $n > 1$ there exist unique integers a, b such that $n = ab^2$ with a squarefree (not divisible by the square of any prime).*

Proof. Write $n = p_1^{e_1} \cdots p_k^{e_k}$ as a product of primes. For each exponent, write $e_i = 2f_i + r_i$ with $r_i = 0$ or $r_i = 1$ by the division algorithm. Then $n = \prod_{i=1}^k p_i^{r_i} \cdot (\prod_{i=1}^k p_i^{f_i})^2 = ab^2$. That a and b are the unique integers with this property are clear. □

As a final application of unique factorization, we give another proof of the next result:

Proposition 8. *Let $a, b, c > 0$ such that $gcd(a, c) = 1$ and $gcd(b, c) = 1$. Then $gcd(ab, c) = 1$.*

Proof. Let The prime factorization of ab comes from multiplying the prime factorizations of a and b together. Let p be a prime dividing c. Then by assumption, $p \nmid a$ and $p \nmid b$, so $p \nmid ab$. Since any common divisor $d > 1$ of ab and c must be divisible by some prime dividing c, this proves $\gcd(ab, c) = 1$. □

1.6 Primes

Since primes build the integers, many questions in number theory are about trying to understand the set of prime numbers. The most basic question one can ask, is how many prime numbers are there? This was first answered by Euclid, around 300 BC.

Theorem 1.6.1 (Euclid)**.** *There are infinitely many primes.*

Proof. Suppose we start with a finite list of primes $p_1, \ldots, p_n$. Consider $N = p_1, \cdots p_n + 1$. We know that N must have a prime divisor, however N is not divisible by any of the primes p_i because it leaves remainder of 1 upon division. Therefore, there is some other prime p_{n+1} such that $p_{n+1} \mid N$. Therefore, given any finite set of primes we can produce a new prime, so there must be infinitely many primes. □

While satisfying, there is a much better argument of this fact that was given by Euler.

Theorem 1.6.2 (Euler)**.** *There are infinitely many primes.*

Proof. Suppose there were finitely many primes, $p_1, \ldots, p_N$. The we get the following product expansion

$$\prod_{i=1}^{N} \frac{1}{1-\frac{1}{p_i}} = \prod_{i=1}^{N} (1 + \frac{1}{p_i} + \ldots)$$

by recognizing each term $\frac{1}{1-\frac{1}{p_i}}$ as the sum of the geometric series $\sum_{k=0}^{\infty} \frac{1}{p_i^k}$. By the Fundamental Theorem of Arithmetic, we can write any integer $n > 1$ as $n = p_1^{e_1} \cdots p_N^{e_N}$ where $e_i \geq 0$. We can recover any such choice of integer as a denominator of a term in the expanded product by picking the appropriate term $\frac{1}{p_i^{e_i}}$ from each infinite series in the product. Therefore, for any $n > 1$ we see that $\frac{1}{n}$ is a term in the product. Therefore, we must have

$$\sum_{n=1}^{\infty} \frac{1}{n} < \prod_{i=1}^{N} \frac{1}{1-\frac{1}{p_i}}$$

The left hand side is the harmonic series, which from calculus diverges to ∞, which results in a contradiction. Therefore, there must be infinitely many primes. □

Euler's argument is much better than Euclid's argument, because it actually provides some additional information about how "large" the set of primes are.

Theorem 1.6.3 (Euler)**.** *The infinite series $\sum_p \frac{1}{p}$ diverges, where the sum is taken over all primes p.*

Proof. The argument is similar to the argument above, but here we need to be a bit more careful with the analysis. For any integer N, similar to before, we can write

$$\prod_{p \leq N} \frac{1}{1-\frac{1}{p}} = \prod_{p \leq N} (1 + \frac{1}{p} + \ldots)$$

by expanding out the terms in the product as infinite series. Again, the same logic tells us that we must have

$$\sum_{n=1}^{N} \frac{1}{n} \leq \prod_{p \leq N} \frac{1}{1-\frac{1}{p}}.$$

Taking logarithms of both sides,

$$\log(\sum_{n=1}^{N} \frac{1}{n}) \leq \sum_{p \leq N} -\log(1 - \frac{1}{p}).$$

Using calculus, one can show that $-\log(1-\frac{1}{p}) \leq \frac{1}{p} + \frac{1}{p^2}$, so we find that

$$\log(\sum_{n=1}^{N} \frac{1}{n}) \leq \sum_{p \leq N} \frac{1}{p} + \sum_{p \leq N} \frac{1}{p^2}.$$

Taking $N \to \infty$, the second term on the right converges by a comparison with $\sum_{n=1}^{\infty} \frac{1}{n^2}$, and the term on the left diverges to ∞ because the harmonic series diverges. This then proves that $\sum_p \frac{1}{p}$ diverges. □

Euler's result tells us that not only are there infinitely many primes, but the set of primes is "large enough" for $\sum_p \frac{1}{p}$ to diverges. This gives us some information that we didn't have before! Note that not all infinite sets satisfy the property that $\sum_{x \in S} \frac{1}{x} = \infty$, e.g. taking S to be the set of square shows this is false. This really is saying something interesting about the primes!

Using the integral test, one can show that $\sum_{n=1}^{N} \frac{1}{n} \approx \log(N)$, so that the partial sums of the harmonic series grow like logarithms. Therefore, the harmonic series diverges slowly. It's a theorem of Merten that $\sum_{p \leq N} \frac{1}{p} \approx \log(\log(N))$, which diverges even *slower*! Using this approximation, we would need $N > 528491311$ for $\sum_{p \leq N} \frac{1}{p} \approx \log(\log(N)) > 3$! That this sum diverges is *very* much not obvious, and if one did not see a proof it would be very easy to convince themselves that it is bounded!

Definition 1.6.4. The **prime counting function** $\pi(x)$ is defined by $\pi(x) = \#\{p \leq x : p \text{ is prime}\}$.

Example 1.6.5. $\pi(5) = 3$ and $\pi(10.5) = 4$.

The benefit of consider $\pi(x)$ as a function of *real* numbers is so that calculus can be used to analyze it. The previous results tell us that $\lim_{x \to \infty} \pi(x) = \infty$. One of the most important problems in all of number theory is to understand how $\pi(x)$ grows. The fundamental result about the growth rate of primes is the following:

Theorem 1.6.6 (Prime Number Theorem)**.** $\lim_{x \to \infty} \frac{\pi(x)}{x/\log(x)} = 1$

In other words, the Prime Number Theorem says that as $x \to \infty$, $\pi(x) \approx \frac{x}{\log(x)}$. There is a much better approximation to $\pi(x)$, first discovered by Gauss.

Definition 1.6.7. The **logarithmic integral** $\mathrm{Li}(x)$ is defined by $\mathrm{Li}(x) = \int_2^x \frac{1}{\log(t)}\, dt$.

It's a standard calculus exercise to show that $\lim_{x \to \infty} \frac{\mathrm{Li}(x)}{x/\log(x)} = 1$, so if one believes the Prime Number Theorem, then one also has $\pi(x) \approx \mathrm{Li}(x)$ as $x \to \infty$. Since $\frac{\pi(x)}{x} \approx \frac{1}{\log(x)}$, and $\frac{\pi(x)}{x}$ is the "probability" that a randomly chosen number in the interval $[2, x]$ is prime, informally we see that the "density" primes is given by $\frac{1}{\log(x)}$. Integrating this "density" function on the interval $[2, x]$ should give the number of primes in the interval $[2, x]$. The Prime Number Theorem says that this is actually a pretty reasonable approximation, so that primes roughly "behave randomly".

By repeatedly integrating by parts, one may show that

$$\mathrm{Li}(x) \approx \frac{x}{\log(x)} + \frac{x}{\log(x)^2} + \frac{2x}{\log(x)^3} + \dots$$

which shows that $\mathrm{Li}(x)$ is indeed a "better" approximation than $\frac{x}{\log(x)}$, and that other easy approximations may be found by just taking a few terms of this expansion. How accurate of an approximation is $\mathrm{Li}(x)$ exactly? Here's a table of $\pi(x)$ and $\mathrm{Li}(x)$ for various values of x.

x	$\pi(x)$	Li(x)
10^4	1229	≈ 1229
10^5	9592	≈ 9571
10^6	78498	≈ 78380
10^{10}	455052511	≈ 454793911

As can be seen, it's generally quite close! Note that although all the values in this table are *underestimates*, $\pi(x) - \text{Li}(x)$ changes sign infinitely often!

How can one quantify how "good" of an approximation $\text{Li}(x)$ is to $\pi(x)$? This is the Riemann Hypothesis!

Conjecture 1.6.8 (Riemann Hypothesis)**.** *There exists a real number $C > 0$ such that for x sufficiently large,*

$$|\pi(x) - Li(x)| \leq C\sqrt{x}\log(x)$$

In other words, the Riemann Hypothesis conjectures that the "error term" $\pi(x) - \text{Li}(x)$ grows roughly like $\sqrt{x}\log(x)$ (up to constant factor and sign), which gives important information about the distribution of prime numbers. The "classical" statement of the Riemann Hypothesis is the following:

Conjecture 1.6.9 (Riemann Hypothesis)**.** *For $s \in \mathbb{C}$ with $Re(s) > 1$, let $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$. Then any zero of $\zeta(s)$ outside of $-2, -4, -6, \cdots$ must have a real part of $\frac{1}{2}$.*

Although not obvious, these two formulations are indeed equivalent. That there is some connection between the Riemann Zeta function and the distribution of primes can be seen from Euler's proof that there are infinitely many of them. One may factor

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

as Euler did, and note that taking $s \to 1$ results in the proof that there are infinitely many primes!

In general, most results about prime numbers are *very* hard to prove. Another famous open problem is the *Twin Prime Conjecture*, which states the following:

Conjecture 1.6.10 (Twin Prime Conjecture)**.** *Let p_n denote the n-th prime number. There are infinitely many n such that $p_{n+1} - p_n = 2$.*

The best known result, proven in 2014, is that there are infinitely many values of n with $p_{n+1} - p_n < 246$.

Another famous question about prime numbers is the distribution of primes in *arithmetic progressions*. The following is a famous result of Dirichlet:

Theorem 1.6.11 (Dirichlet)**.** *Let $a, b > 0$ with $gcd(a, b) = 1$. There are infinitely many primes of the form $ak + b$.*

The idea of the proof is a modified version of Euler's argument. Dirichlet shows that with $S = \{p : p = ak + b, p \text{ prime}\}$, one has $\sum_{p \in S} \frac{1}{p} = \infty$. However, the argument is significantly harder. Simple cases of Dirichlet's Theorem can be proven by hand, but there is no "universal" technique for proving there *are* infinitely many primes of form $ak + b$ for fixed values of a, b. To show there are infinitely many primes of the form $23 + 31k$, for example, there is likely no known method outside of just proving Dirichlet's Theorem.

1.7 Exercises

1. Apply the extended division algorithm to each pair of numbers a, b. That is, for each one, find integers q and r such that $a = b \cdot q + r$ with $0 \leq r < |b|$.

   (a) $a = 47$, $b = -13$

   (b) $a = 956$, $b = -27$

   (c) $a = 29657452$, $b = -4382$

2. For each of the pairs of integers (a, b) below, do the following:

   (i) Run the Euclidean algorithm to compute $\gcd(a, b)$.

   (ii) Use back substitution to find integers x, y such that $ax + by = \gcd(a, b)$.

   (a) $(504, 94)$

   (b) $(-1260, 816)$

3. Compute the gcd and lcm of $2^2 3^3 5^5 7^7 11^{11}$ and $2^7 3^5 5^3 7^2 13$.

4. Let a, b be integers with $\gcd(a, 25) = 5$ and $\gcd(b, 125) = 25$. Compute $\gcd(ab, 625)$, $\gcd(a + b, 625)$ and $\text{lcm}(ab, 625)$.

5. Use the Euclidean algorithm to compute $\text{lcm}(3134376, 17599768)$.

6. (a) Find all integer solutions to $2260x + 816y = 6968$.

   (b) Find all *positive* integer solutions to $54x + 21y = 906$.

   (c) Find all integer solutions to $2x + 3y + 5z = 1$.

7. Prove Proposition 1.2.2.

8. Prove Theorem 1.2.5.

9. Suppose that $n > 1$ is an odd integer. Prove that $8 \mid n^2 - 1$.

10. Show that if $n > 1$ is odd, then $24 \mid n^3 - n$.

11. Show that the sum of three consecutive odd integers is divisible by 9.

12. Find all rectangles with integer sides with equal area and perimeter.

13. (a) For any integers $x > 1$ and $m, n \geq 0$, prove that $(x^n - 1) \mid (x^m - 1)$ if and only if $n \mid m$.

    (b) For any integer x, prove that if n is odd, $(x + 1) \mid (x^n + 1)$.

14. In this problem, you will give a proof that $\sqrt{2}$ is irrational using the Euclidean algorithm. Suppose that $\sqrt{2}$ was rational, so it can be written as $\sqrt{2} = \frac{a}{b}$ for some positive integers a, b with $b \neq 0$.

    (a) Show that $a = b \cdot 1 + (a - b)$ with $0 \leq a - b < b$, so that this is the first step in the Euclidean algorithm on the pair (a, b) with $q_1 = 1$ and $r_1 = a - b$.

    (b) Write down the next step in the Euclidean algorithm by performing the division algorithm on the pair $(b, a - b)$. What is q_2? What is the ratio r_1/r_2? (Your answers should be *numbers*, not involving the letters a, b).

    (c) Prove that $q_n = q_2$ and $\frac{r_{n-1}}{r_n} = \frac{r_1}{r_2}$ for all $n \geq 2$. (*Hint: prove these both simultaneously via induction.*)

    (d) Explain why the truth of the statement in (c) yields a contradiction, therefore proving that $\sqrt{2}$ must not be rational.

15. The *Fibonacci sequence* is defined by

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

    Prove by induction that for all $n \geq 1$, the number of steps required for the Euclidean algorithm on the pair (F_{n+1}, F_n) to terminate is exactly n.

16. Prove that if $\gcd(a, b) = d$ then $\gcd(a^2, b^2) = d^2$.

17. Prove that if $\gcd(a, b) = 1$, then $\text{lcm}(a, b) = ab$.

18. Prove that $\gcd(a, bc) \mid \gcd(a, b)\gcd(a, c)$.

19. Prove that if $\gcd(a, b) = 1$, then $\gcd(a + b, ab) = 1$.

20. Prove that if $\gcd(a, b) = 1$ and $ab = c^2$, then a and b are both squares.

21. Suppose that $a, b \in \mathbb{Z}$ with $a \mid c$ and $b \mid c$. Prove that $\text{lcm}(a, b) \mid c$.

22. Let a, b be integers and let $n \geq 1$. Prove that $a^n \mid b^n$ if and only if $a \mid b$.

23. Prove that for any integer $n \geq 1$ there exist unique integers r, m with m odd such that $n = 2^r m$.

24. Find an integer n such that $\frac{n}{2}$ is a square, $\frac{n}{3}$ is a cube, and $\frac{n}{5}$ is a fifth power.

25. Let n be a positive integer such that the exponent of each prime in its factorization is at least two. For example, $1944 = 2^3 \cdot 3^5$. Prove that there are integers a, b such that $n = a^2 b^3$.

26. (a) Let $a, b \geq 1$ be integers. Prove that if $(a^2 - b^2) \mid (a^2 + b^2)$, then $(a^2 - b^2) \mid 2\gcd(a, b)^2$.

    (b) Prove there are no integers $a, b \geq 1$ such that $(a^2 - b^2) \mid (a^2 + b^2)$.

27. Define $v_p(n)$ to be the exponent of p in the prime factorization of n. For example, $v_3(45) = 2$ because $45 = 3^2 \cdot 5$. Prove that $v_p(n!) = \lfloor \frac{n}{p} \rfloor + \lfloor \frac{n}{p^2} \rfloor + \ldots$, where $\lfloor x \rfloor$ denotes the greatest integer that is at most x.

28. We define the function $\tau(n)$ to be the number of positive divisors of n. For example, $\tau(12) = 6$ because it has positive divisors $1, 2, 3, 4, 6, 12$.

    (a) Let m, n be integers with $\gcd(m, n) = 1$. Prove that if $d \mid mn$, then $d = d_1 d_2$ for unique d_1, d_2 with $d_1 \mid n$ and $d_2 \mid m$. Explain why this means $\tau(mn) = \tau(m)\tau(n)$.

    (b) Write down a formula for $\tau(n)$ in terms of the prime factorization of n. (*Hint: start by computing $\tau(p^e)$ for p prime and $e \geq 1$.*)

29. Compute the number of divisors of 20!.

30. How many zeroes does 169! end in?

31. Find all integers $n < 100$ such that $\tau(n) = 12$.

32. Let $p \geq 5$ be a prime. Show that $p^2 + 2$ is composite (*Hint: show that $p^2 + 2$ is divisible by 3*).

33. Prove that if $n > 1$ is composite, then n has a prime factor p with $p \leq \sqrt{n}$.

34. (a) Prove that if $2^n - 1$ is prime, then n is prime. Similarly, prove that if $2^n + 1$ is prime, then n is a power of 2.

    (b) Show that the converse of each statement is false (you may use WolframAlpha for computations).

35. Prove that there are infinitely many primes of the form $4k + 3$.

36. (Sieve of Eratosthenes) Write down all natural numbers from 1 to 100, perhaps on a 10×10 array. Circle the number 2, the smallest prime. Cross off all numbers divisible by 2. Circle 3, the next number that is not crossed out. Cross off all larger numbers that are divisible by 3. Continue to circle the smallest number that is not crossed out and cross out all its multiples. Repeat. Why are the circled numbers all the primes less than 100?

37. Prove that $\lim_{x \to \infty} \frac{\mathrm{Li}(x)}{x/\log(x)} = 1$.

Chapter 2

Unique factorization in Euclidean domains

2.1 Unique factorization in $F[x]$

The integers are not just the only object of interest to number theorists. There is a very close link between numbers and polynomials. For example, instead of thinking of the number $\frac{1}{2} \in \mathbb{Q}$, one may instead think of it as the (unique) root of the polynomial $2x - 1 \in \mathbb{Z}[x]$. Throughout these notes, **we assume that all rings are commutative.**

Definition 2.1.1. Let R be a ring and let $p(x) = a_n x^n + \ldots + a_0 \in R[x]$ be a polynomial. The **degree** of $p(x)$ is the largest exponent out of the non-zero terms in the sum. We call $p(x)$ **monic** if the leading coefficient (coefficient of the highest exponent term) is 1.

Definition 2.1.2. Let $f(x), g(x) \in R[x]$. We say that $f(x)$ **divides** $g(x)$ and write $f(x) \mid g(x)$ if there is a polynomial $h(x) \in R[x]$ such that $g(x) = f(x)h(x)$.

For our purposes, we're going to mainly focus on when $R = F$ is a *field*, as the polynomial rings $F[x]$ will turn out to be those that are most similar. This is not to say that other polynomial rings are not of interest! In particular, $\mathbb{Z}[x]$ is rather important.

Proposition 9. *Suppose that $f(x), g(x) \in R[x]$ for $R = \mathbb{Z}$ or $R = F$ a field. Then $deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x))$.*

Proof. Suppose that the leading terms of $f(x)$ and $g(x)$ are $a_n x^n$ and $b_m x^m$ respectively. Then the leading term of $f(x)g(x)$ is $a_n b_m x^{n+m}$. Since $a_n b_m \in R$, we must have $a_n b_m \neq 0$ because $a_n, b_m \neq 0$ and the product of two non-zero elements in R cannot be zero. □

Definition 2.1.3. Let $f(x) \in F[x]$. We say that $f(x)$ is **irreducible** if $f(x) = g(x)h(x)$ for some $g(x), h(x) \in F[x]$ means either $g(x)$ is constant or $h(x)$ is constant. If $f(x)$ is not irreducible, we call it **reducible**.

We have the following analogies between $\mathbb{Z}$ and $F[x]$:

$\mathbb{Z}$	$F[x]$
± 1	non-zero constant polynomials
$\lvert n \rvert$	$\deg(f(x))$
positive	monic
prime	irreducible

In $\mathbb{Z}$, we got unique factorization by following the chain of reasoning:

Division Algorithm $\implies$ Euclidean Algorithm $\implies$ Bezout's lemma $\implies$ Euclid's lemma $\implies$ Unique factorization.

In $F[x]$, we're going to prove analogous versions of all of these results and follow the same chain of reasoning to come to the same conclusion!

Theorem 2.1.4 (Division Algorithm)**.** *Let $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. There exist unique $q(x), r(x) \in F[x]$ such that $f(x) = g(x)q(x) + r(x)$ with $r(x) = 0$ or $\deg(r(x)) < \deg(g(x))$.*

Proof. If $g(x) \mid f(x)$ this is obvious (with $r(x) = 0$), so assume it doesn't. Let $S = \{f(x) - g(x)q(x) : q(x) \in F[x]\}$. Note that $f(x) \in S$, so that S is non-empty. By the well-ordering principle, choose $r(x) \in S$ of minimal non-negative degree. This means that there is $q(x) \in F[x]$ such that $f(x) - g(x)q(x) = r(x)$, i.e. $f(x) = g(x)q(x) + r(x)$. First, we show the remainder bound. If $\deg(r(x)) \geq \deg(g(x))$, then write $r(x) = r_d x^d + \ldots$ and $g(x) = g_k x^k + \ldots$, where $r_d x^d$ and $g_k x^k$ are the leading terms of $r(x)$ and $g(x)$ respectively. Then $r(x) - g(x) \cdot (\frac{r_k}{g_k} x^{d-k}) = f(x) - g(x)(q(x) + \frac{r_d}{g_k} x^{d-k}) \in S$, however this polynomial has smaller degree than $r(x)$, which was a polynomial of minimal degree among the elements of S. This is a contradiction, so this means $\deg(r(x)) < \deg(g(x))$.

For the uniqueness statement, if two pairs $(q_1(x), r_1(x))$ and $(q_2(x), r_2(x))$ satisfy the conditions of the theorem, this means that $g(x)(q_1(x) - q_2(x)) = r_2(x) - r_1(x)$. Taking degrees, we must have $\deg(g(x)) + \deg(q_1(x) - q_2(x)) = \deg(r_2(x) - r_1(x))$. If both $q_1(x) - q_2(x)$ and $r_2(x) - r_1(x)$ are non-zero, then we would have $\deg(g(x)) + \deg(q_1(x) - q_2(x)) < \deg(g(x))$, which is impossible. Therefore, at least one of these polynomials is zero, and once one of them is it's clear the other must be as well. Therefore, $q_1(x) = q_2(x)$ and $r_1(x) = r_2(x)$ as desired. □

Example 2.1.5. Let $f(x) = x^4 - 4x^3 + 3x^2 + 2x - 1$ and $g(x) = x^2 - 2x + 1$ in $\mathbb{Q}[x]$. Then we have $f(x) = (x^2 - 2x - 2)g(x) + 1$. Once again, this process is equivalent to the process of polynomial long division that one learns in middle school mathematics.

In the proof of the division algorithm, we needed to divide by the leading coefficient of $g(x)$. Therefore, we really did use that F was a field, and so for a general ring $R[x]$, there's no hope for a general division algorithm. However, in any ring $R[x]$, we can always divide by *monic* polynomials, as the same proof goes through!

Definition 2.1.6. Let $f(x), g(x) \in F[x]$. The **greatest common divisor** of $f(x)$ and $g(x)$, $\gcd(f(x), g(x))$ is the *monic* polynomial $d(x)$ that satisfies:

1. $d(x) \mid f(x)$ and $d(x) \mid g(x)$.
2. For any monic $h(x)$ such that $h(x) \mid f(x)$ and $h(x) \mid g(x)$ we have $\deg(h(x)) < \deg(g(x))$.

In $\mathbb{Z}[x]$, we cannot always rescale a polynomial to be monic like we can in $F[x]$, so if we want to define a greatest common divisor in this ring we need to alter the definition slightly. The second condition can be instead be replaced with "$h(x) \mid f(x)$ and $h(x) \mid g(x) \implies h(x) \mid d(x)$". If we then require the leading coefficient of $d(x)$ to be *positive*, we can then pick out "a" greatest common divisor. Note that in our definition for $F[x]$, we require the greatest common divisor to be monic, otherwise there are *many* polynomials that satisfy the above properties (which all just differ by some constant multiple).

Just as before, we can repeatedly run the division algorithm over and over again to compute the greatest common divisor of two polynomials $f(x)$ and $g(x)$ in $F[x]$, giving us a Euclidean Algorithm. Also, just as before, once we have a Euclidean Algorithm, it can be run backwards to write $\gcd(f(x), g(x))$ as an $F[x]$-linear combination of $f(x)$ and $g(x)$, giving us a Bezout lemma. We'll provide precise statements of these results for completeness sake, but the proofs are left as exercises for the reader.

Theorem 2.1.7 (Euclidean Algorithm)**.** *Let $f(x), g(x)$ be non-zero polynomials. Repeatedly carry out the division algorithm as follows:*

$$\begin{aligned}
f(x) &= g(x)q_1(x) + r_1(x), \quad \deg(r_1(x)) = 0 \text{ or } \leq \deg(r_1(x)) < \deg(g(x)) \\
g(x) &= r_1(x)q_2(x) + r_2(x), \quad \deg(r_2(x)) = 0 \text{ or } \leq \deg(r_2(x)) < \deg(r_1(x)) \\
r_1(x) &= r_2(x)q_3(x) + r_3(x), \quad \deg(r_3(x)) = 0 \text{ or } \leq \deg(r_3(x)) < \deg(r_2(x)) \\
&\vdots
\end{aligned}$$

The last non-zero remainder is the greatest common divisor of $f(x)$ and $g(x)$, up to constant multiple.

Theorem 2.1.8 (Bezout's lemma)**.** *Let $f(x), g(x) \in F[x]$. Then there are $p(x), q(x) \in F[x]$ such that $f(x)p(x) + g(x)q(x) = \gcd(f(x), g(x))$.*

Example 2.1.9. Let $f(x) = x^3 + x^2 + 1$ and $g(x) = x^2 + x + 1$ in $\mathbb{Q}[x]$. First, let's compute $\gcd(f(x), g(x))$. Running the Euclidean Algorithm produces the following sequence of steps:

$$\begin{aligned}
f(x) &= x \cdot g(x) + (1 - x) \\
g(x) &= (1 - x)(-x - 2) + 3 \\
-x - 2 &= 3 \cdot (\frac{1}{3}(-x - 2)) + 0
\end{aligned}$$

The last non-zero remainder is 3, which when rescaled to be monic, is 1. This says that $\gcd(f(x), g(x)) = 1$, so that $f(x)$ and $g(x)$ are relatively prime. To find $p(x), q(x)$ such that $f(x)p(x) + g(x)q(x) = 1$, first, we perform back substitution. Doing so, we find that $(1 - x^2 - 2x)g(x) + (x + 2)f(x) = 3$, and then rescaling gives $\frac{1-x^2-2x}{3}g(x) + \frac{x+2}{3}f(x) = 1$.

Now that we have a Euclidean Algorithm and a Bezout lemma, we need a Euclid lemma before we can get unique factorization.

Theorem 2.1.10 (Euclid's lemma)**.** *Suppose that $f(x) \in F[x]$ is irreducible and $f(x) \mid g(x)h(x)$ in $F[x]$. Then $f(x) \mid g(x)$ or $f(x) \mid h(x)$.*

Proof. The proof is the same as in $\mathbb{Z}$: suppose that $f(x) \nmid g(x)$. Then since $f(x)$ is irreducible, we have $\gcd(f(x), g(x)) = 1$, so by Bezout's lemma there are $p(x), q(x)$ such that $f(x)p(x) + g(x)q(x) = 1$. Multiplying by $h(x)$, we have $f(x)h(x)p(x) + g(x)h(x)q(x) = h(x)$. Since $f(x)$ divides both terms in the sum on the left, we must have $f(x) \mid h(x)$ as desired. □

Theorem 2.1.11 (Unique factorization)**.** *Let $f(x) \in F[x]$ be a non-constant monic polynomial. There exist unique monic irreducibles $\pi_1(x), \ldots, \pi_k(x)$ and integers $e_1, \ldots, e_k > 0$ such that $f(x) = \pi_1(x)^{e_1} \cdots \pi_k(x)^{e_k}$.*

Proof. As before, the proof has two parts. First, we show that every non-constant polynomial in $F[x]$ is a product of monic irreducibles, and then we show this happens for a unique set of monic irreducibles.

Existence:

We proceed by induction on $\deg(f(x))$. If $\deg(f(x)) = 1$, then we're done, as any degree 1 polynomial is irreducible. Suppose we know that any monic polynomial of degree $\leq k$ can be written as a product of monic irreducibles for some k. Let $f(x)$ be a monic degree $k+1$ polynomial. If $f(x)$ is irreducible, we're done, otherwise $f(x)$ is reducible, so we can write $f(x) = g(x)h(x)$ for some $g(x), h(x) \in F[x]$ with $1 \leq \deg(g(x)), \deg(h(x)) \leq k$. Without loss of generality, since $f(x)$ is monic and we're working over a *field*, we can rescale $g(x)$ and $h(x)$ to both be monic. By assumption, both $g(x)$ and $h(x)$ factor into products of monic irreducibles, and therefore $f(x)$ does as well. By induction, the result is true for any monic polynomial of degree at least 1.

Uniqueness:

If $\deg(f(x)) = 1$, we're again done, as $f(x)$ is irreducible. Now, suppose that we know the factorization is unique for all monic $f(x)$ of degree $\leq k$ for some k. Let $f(x)$ be a monic polynomial of degree $k+1$. Once more, if $f(x)$ is irreducible, we're done. Otherwise, $f(x)$ is reducible, so it can be written as a product of monic irreducibles. Let $f(x) = \pi_1(x) \cdots \pi_k(x) = q_1(x) \cdots q_\ell(x)$ be a two factorization of $f(x)$ into monic irreducibles. Since $\pi_1(x)$ divides the left hand side, it must divide $q_i(x)$ for some i by Euclid's lemma. Without loss of generality, suppose that $\pi_1(x) \mid q_1(x)$. Since both $\pi_1(x)$ and $q_1(x)$ are *monic* and irreducible, this means that $\pi_1(x) = q_1(x)$. Therefore, we must have $\pi_2(x) \cdots \pi_k(x) = q_2(x) \cdots q_\ell(x)$. This is a polynomial of degree $\leq k$, so by induction hypothesis, it has a unique factorization into a product of monic irreducibles. This forces $k = \ell$ and that $\{\pi_i(x)\} = \{q_j(x)\}$ for $i \geq 2$. Since $\pi_1(x) = q_1(x)$, this means we actually have $\{\pi_i(x)\} = \{q_j(x)\}$ for all i, so that the two sets of monic irreducible factors of $f(x)$ are actually the same. Therefore by induction, any monic non-constant polynomial has a unique factorization into monic irreducibles. Collecting powers of the same irreducible factors gives the form stated in the theorem. □

Note that if we didn't required that the irreducible factors be monic, then the factorization would only be unique up to a constant-multiple. The key point is that two *monic* irreducible polynomials in $F[x]$ are equal if they divide each other.

2.2 Unique factorization in $\mathbb{Z}[i]$

Definition 2.2.1. The **Gaussian integers** $\mathbb{Z}[i]$ are defined by $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$.

We have that $\mathbb{Z}[i] \subseteq \mathbb{C}$ is a subring, so we can and multiply Gaussian integers the same way that we can add and multiply complex numbers.

Definition 2.2.2. For $\alpha = a + bi \in \mathbb{Z}[i]$, the **norm** of α is defined by $N(\alpha) = \alpha\bar{\alpha} = a^2 + b^2$.

In the Gaussian integers, the norm is the notion of size, analogous to the degree of a polynomial or the absolute value of an integer.

Proposition 10. *Let $\alpha, \beta \in \mathbb{Z}[i]$. Then $N(\alpha\beta) = N(\alpha)N(\beta)$.*

Proof. Let $\alpha = a + bi$ and $\beta = c + di$, so that $\alpha\beta = (ac - bd) + (ad + bc)i$. Then we have $N(\alpha)N(\beta) = (a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (bc + ad)^2 = N(\alpha\beta)$. □

Definition 2.2.3. We call $\alpha \in \mathbb{Z}[i]$ a **unit** if there is $\beta \in \mathbb{Z}[i]$ such that $\alpha\beta = 1$.

Proposition 11. *$\alpha \in \mathbb{Z}[i]$ is a unit if and only if $\alpha = \pm 1, \pm i$.*

Proof. We have $\alpha\beta = 1$ in $\mathbb{Z}[i]$ if and only if $N(\alpha)N(\beta) = 1$ in $\mathbb{Z}$. Since $N(\alpha) \geq 0$, this means $N(\alpha) = 1$. Writing $\alpha = a + bi$, this means $a^2 + b^2 = 1$. The integer solutions to this equation are easily seen to be $(\pm 1, 0)$ and $(0, \pm 1)$, corresponding to the four numbers ± 1 and $\pm i$. Each of these are obviously units, so there are exactly four. □

In $\mathbb{Z}$ the units are ± 1, and in $F[x]$ the units are non-zero constant polynomials. In both of these settings, we have a "canonical" choice of unit: in $\mathbb{Z}$, we can pick out 1 because its positive, and in $F[x]$ we can always rescale to make polynomials monic. However, in $\mathbb{Z}[i]$, we do not have such a canonical choice of unit. This is going to result in a lot of statements in $\mathbb{Z}[i]$ only being unique up to unit multiple.

Definition 2.2.4. For $\alpha, \beta \in \mathbb{Z}[i]$, we say that α **divides** β and write $\alpha \mid \beta$ if there is $\gamma \in \mathbb{Z}[i]$ such that $\beta = \alpha\gamma$.

Example 2.2.5. $14 - 3i = (4 + 5i)(1 - 2i)$, so $4 + 5i \mid 14 - 3i$. However, $4 + 5i \nmid 14 + 3i$. In $\mathbb{C}$, we have $\frac{14+3i}{4+5i} = \frac{71}{41} - \frac{58}{41}i \notin \mathbb{Z}[i]$.

Often times, its useful to pass from $\mathbb{Z}[i]$ to $\mathbb{Z}$ by taking norms. One such use is for checking for potential divisibility:

Proposition 12. *Let $\alpha, \beta \in \mathbb{Z}[i]$. If $\alpha \mid \beta$ then $N(\alpha) \mid N(\beta)$ in $\mathbb{Z}$.*

Proof. Suppose that $\alpha \mid \beta$. Then $\beta = \alpha\gamma$ for some $\gamma \in \mathbb{Z}[i]$, taking norms results in $N(\beta) = N(\alpha)N(\beta)$, which says that $N(\alpha) \mid N(\beta)$ in $\mathbb{Z}$. □

Next, we state and prove the division algorithm in $\mathbb{Z}[i]$.

Theorem 2.2.6 (Division Algorithm)**.** *Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$. Then there exist $\gamma, \rho \in \mathbb{Z}[i]$ such that $\alpha = \beta\gamma + \rho$ with $0 \leq N(\rho) \leq \frac{1}{2}N(\beta)$.*

Proof. Let $\alpha = a + bi$ and $\beta = c + di$. In $\mathbb{C}$, we have $\frac{\alpha}{\beta} = \frac{\alpha\bar{\beta}}{N(\beta)} = \frac{ac+bd}{c^2+d^2} + \frac{ad-bc}{c^2+d^2}i = r + si$ for $r = \frac{ac+bd}{c^2+d^2}$ and $s = \frac{ad-bc}{c^2+d^2}$. Choose $p, q \in \mathbb{Z}$ such that $|r - p| \leq \frac{1}{2}$ and $|s - q| \leq \frac{1}{2}$ (which may be achieved by rounding r, s to the nearest integer). Set $\gamma = p + qi$, and let $\theta = (r - p) + (s - q)i = \frac{\alpha}{\beta} - \gamma$. This says that $\alpha = \beta\theta + \beta\gamma$, so that $\beta\theta \in \mathbb{Z}[i]$. We then have $N(\beta\theta) = N(\beta)N(\theta) \leq N(\beta)((\frac{1}{2})^2 + (\frac{1}{2})^2) = \frac{1}{2}N(\beta)$. Therefore, we're done by taking $\rho = \beta\theta$. □

Example 2.2.7. With $\alpha = 11 + 10i$ and $\beta = 4 + i$, we have $\frac{\alpha}{\beta} = \frac{54}{17} + \frac{29}{17}i \in \mathbb{C}$. We take $\gamma = 3+2i$, and $\rho = \alpha - \beta\gamma = 1-i$. Then the division algorithm says that $\alpha = \beta(3+2i)+(1-i)$.

Unlike in both $\mathbb{Z}$ and $F[x]$, we do not have uniqueness of γ and ρ in the division algorithm for $\mathbb{Z}[i]$. For example,

$$1 + 8i = (2 - 4i)(i - 1) - (1 + 2i)$$
$$1 + 8i = (2 - 4i)(-2 + i) + (1 - 2i)$$

This is related to not being able to pick out a canonical choice of unit in $\mathbb{Z}[i]$. If one dropped the requirement that the remainder in $\mathbb{Z}$ is positive, uniqueness would also be lost. For example,

$$9 = 4 \cdot 2 + 1$$
$$9 = 4 \cdot 3 + -3$$

Once again, the division algorithm in $\mathbb{Z}[i]$ leads to a Euclidean algorithm in $\mathbb{Z}[i]$. This give us a notion of greatest common divisor, and by running the Euclidean algorithm backwards, a Bezout lemma. However, due to the unit issue in $\mathbb{Z}[i]$, we have to be a bit careful. It's not possible to pick out "a" greatest common divisor like in $\mathbb{Z}$ and $F[x]$. There is some inherent ambiguity in the definition!

Definition 2.2.8. Let $\alpha, \beta \in \mathbb{Z}[i]$. We call $\delta \in \mathbb{Z}[i]$ a **greatest common divisor** of α, β if:

1. $\delta \mid \alpha$ and $\delta \mid \beta$.
2. For any other $\gamma \in \mathbb{Z}[i]$ with $\gamma \mid \alpha$ and $\gamma \mid \beta$ we have $\gamma \mid \delta$.

In particular, if δ is a greatest common divisor of α, β then so are $\pm\delta, \pm i\delta$. When we write $\gcd(\alpha, \beta)$, we mean a choice of greatest common divisor. If $\gcd(\alpha, \beta)$ is a unit, then we say that α, β are relatively prime.

Theorem 2.2.9. *Let $\alpha, \beta \in \mathbb{Z}[i]$. Then there exist $\gamma, \rho \in \mathbb{Z}[i]$ such that $\alpha\gamma + \beta\rho = \gcd(\alpha, \beta)$ for any choice $\gcd(\alpha, \beta)$.*

Proof. The proof is as usual. The Euclidean algorithm terminates in *a* greatest common divisor δ of α and β, so running it backwards produces γ, ρ such that $\alpha\gamma + \beta\rho = \delta$. We just need to check that any two greatest common divisors differ by a unit multiple, so that we can obtain any of them from one equation. However, this is easy, as if δ, δ' are two greatest

common divisors of α, β, then $\delta' \mid \delta$ and $\delta \mid \delta'$. This means $N(\delta') \mid N(\delta)$ and $N(\delta) \mid N(\delta')$, so that $N(\delta) = N(\delta')$. Then from $\delta' \mid \delta$, we see that $\delta = \varepsilon\delta'$ for some $\varepsilon \in \mathbb{Z}[i]$, so taking norms says $N(\varepsilon) = 1$, which means that ε is a unit as desired. □

Example 2.2.10. Let $\alpha = 32 + 9i$ and $\beta = 4 + 11i$. Running the Euclidean algorithm,

$$\begin{aligned} 32 + 9i &= (4 + 11i)(2 - 2i) + (2 - 5i) \\ 4 + 11i &= (2 - 5i)(-2 + i) + (3 - i) \\ 2 - 5i &= (3 - i)(1 - i) - i \\ 3 - i &= (-i)(1 + 3i) + 0 \end{aligned}$$

The last non-zero remainder is a unit, so α and β are relatively prime. If we back substitute, we can solve the Bezout equation. Doing so, we find that $\alpha(3i) + \beta(-7 - 5i) = -i$.

In order to talk about unique factorization, we need to decide what it means for $\alpha \in \mathbb{Z}[i]$ to be prime.

Definition 2.2.11. Let $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) > 1$. We call α **prime** if α has no non-trivial divisors in $\mathbb{Z}[i]$, i.e. if $\alpha = \beta\gamma$ then either β or γ is a unit in $\mathbb{Z}[i]$.

One can sometimes easily detect if a Gaussian integer is prime or not by looking at its norm:

Proposition 13. *Suppose that $\alpha \in \mathbb{Z}[i]$ has $N(\alpha)$ prime in $\mathbb{Z}$. Then α is prime in $\mathbb{Z}[i]$.*

Proof. Suppose that $\alpha = \beta\gamma$ for some $\beta, \gamma \in \mathbb{Z}[i]$. Taking norms, this means $p = N(\alpha) = N(\beta)N(\gamma)$ for some prime $p \in \mathbb{Z}$. This means either $N(\beta)$ or $N(\gamma)$ equals 1 since p is prime, and therefore that either β or γ is a unit in $\mathbb{Z}[i]$. □

Be careful! The converse of this statement is not true. For example, 3 is prime in $\mathbb{Z}[i]$, but its norm $N(3) = 9$ is not prime in $\mathbb{Z}$. To see that 3 is prime, suppose that $3 = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$. Taking norms, we would have $9 = N(\alpha)N(\beta)$. If the factorization is non-trivial, then $N(\alpha) = 3$. Writing $\alpha = a + bi$, we would need $a^2 + b^2 = 3$. However, one can easily see that this has no integer solutions! Therefore, it's not possible to have a non-trivial factorization.

Once again, before we can get unique factorization, we need Euclid's lemma.

Proposition 14 (Euclid's lemma)**.** *Suppose that $\pi \in \mathbb{Z}[i]$ is prime and $\pi \mid \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$. Then $\pi \mid \alpha$ or $\pi \mid \beta$.*

Proof. Same as usual. If $\pi \nmid \alpha$, then since π is prime $\gcd(\pi, \alpha)$ is a unit. Therefore by Bezout, we can find γ, ρ such that $\pi\gamma + \alpha\rho = 1$. Multiplying through by β shows that $\pi \mid \beta$. □

We're now ready to state and prove unique factorization in $\mathbb{Z}[i]$.

Theorem 2.2.12. *Let $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) > 1$. Then there exist $\pi_1, \ldots, \pi_k \in \mathbb{Z}[i]$ prime such that $\alpha = \pi_1 \cdots \pi_k$. This is unique, in the following sense: if $\pi_1', \ldots, \pi_\ell'$ is another prime factorization, then $k = \ell$ and there is some reordering so that π_i and π_i' differ by a unit.*

Proof. We follow the same proof structure as we have done for the other two locations. First we prove every such α can be written as a product of primes, and then we show the notion of uniqueness. We do this by induction on the norm.
Existence:
If $N(\alpha) = 2$, we're done since α is prime. Suppose that any α with $N(\alpha) \leq k$ for some k can be written as a product of primes, and let $\alpha \in \mathbb{Z}[i]$ have $N(\alpha) = k + 1$. If α is prime, we're done. Otherwise, write $\alpha = \beta\gamma$ for some β, γ with $1 \leq N(\beta), N(\gamma) \leq k$. By assumption, both β and γ can then be written as a product of primes, and therefore α can as well. By induction, any such α is a product of primes.

Uniqueness:
Again, if $N(\alpha) = 2$ we're done. Assume we know that for any α with $N(\alpha) \leq k$ for some k, that the prime factorization is unique in the sense as stated in the theorem. Let $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) = k+1$. If α is prime, again we're done, otherwise α can be written as a product of primes. Write $\alpha = \pi_1 \cdots \pi_k = \pi_1' \cdots \pi_\ell'$ as two prime factorizations. Since π_1 is prime and π_1 divides the left hand side, by Euclid's lemma we must have $\pi_1 \mid \pi_i'$ for some i. Without loss of generality, assume that $\pi_1 \mid \pi_1'$. This means that $\pi_1' = \pi_1\gamma$ for some γ, and since π_1' is *prime*, this means that γ must be a unit. Canceling π_1 from both sides, this says $\pi_2 \cdots \pi_k = \gamma\pi_2' \cdots \pi_\ell'$. Since $\gamma\pi_2$ is still a prime, by induction hypothesis, we must have $k = \ell$. and after reordering, π_i must differ from π_i' by a unit (here, by abuse of notation, we are taking $\pi_2' = \gamma\pi_2'$ in the original sense of π_2'). The point, then, is that a unit multiple of $\gamma\pi_2'$ is still just a unit multiple of π_2', and so π_i differs by a unit multiple of π_i' for *all* $i \geq 1$. Therefore by induction, we're done. □

You're probably used to factoring both integers and polynomials throughout your life, but it's unlikely you have experience with factoring Gaussian integers. How can this be done? Take norms!

Example 2.2.13. Let $\alpha = 3 + 4i$. Then $N(\alpha) = 25$, so any non-trivial divisor of α must have norm 5. Let $\beta = a+bi$, so we want to solve $N(\beta) = a^2+b^2 = 5$. One may check that the only solutions to this are $(\pm 2, \pm 1)$ and $(\pm 1, \pm 2)$, which correspond to the Gaussian integers $1 + 2i, 1 - 2i$, and all their unit multiples. Both of these are prime because their norm is 5, so we can just check for divisibility by hand. We find that $\frac{3+4i}{1-2i} = \frac{-5+10i}{5} = -1 + 2i$, so that $3 + 4i = (1 - 2i)(-1 + 2i) = -(1 - 2i)^2$ is a prime factorization of $3 + 4i$.

2.3 Euclidean domains

In the above two sections, we've seen two other settings that have unique factorization, just like in the integers. In both of these places, we followed the exact same chain of reasoning to reach this conclusion. In this section, we'll use the language of rings to abstract away the properties needed to follow this chain of reasoning.

Definition 2.3.1. An **integral domain** is a ring R such that for any $a, b \in R$ with $a, b \neq 0$, $ab \neq 0$.

Definition 2.3.2. A **Euclidean domain** is an integral domain R with a function $N : R\backslash \to \mathbb{N}$ with $N(0) = 0$ such that for any $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ such that $a = bq + r$ with $r = 0$ or $N(r) < N(b)$. The function N is called the **norm**.

This definition says that a Euclidean domain is an integral domain that has a division algorithm.

Example 2.3.3. The ring $\mathbb{Z}$ is a Euclidean domain with $N(n) = |n|$. For F a field, $F[x]$ is a Euclidean domain with norm given by $N(p(x)) = \deg(p(x))$. $\mathbb{Z}[i]$ is a Euclidean domain with norm $N(\alpha) = \alpha\bar{\alpha}$.

If R is a Euclidean domain, then because R has a division algorithm, R has a Euclidean algorithm. In $\mathbb{Z}[i]$, we saw that the greatest common divisor is only defined *up to unit multiple*. There, it was easy to find all the units, because the norm was multiplicative. This need not be true in general, and a general Euclidean domain could have many units. For this reason, it will be more beneficial to state results in terms of *ideals* of R: two principal ideals (d) and (d') are equal if and only if $d = ud'$ for some $u \in R^\times$, so the *ideal* generated by an element does not care about multiplication by a unit.

Proposition 15. *Let R be a Euclidean domain. Then every ideal of R is principal.*

Proof. Let $I \neq (0)$ be an ideal of R. By the well-ordering principle, there exists $d \in I$ such that d has minimal among all elements of I. Then clearly $(d) \subseteq I$. For any $a \in I$, since R is Euclidean, we may write $a = dq + r$ with $r = 0$ or $N(r) < N(d)$. As $r = a - dq \in I$, by definition of d we must have $r = 0$, so $a \in (d)$ and therefore $I = (d)$. □

The next corollary follows immediately:

Corollary 2.3.4. *(Bezout's lemma) Let R be a Euclidean domain. For any $a, b \in R$ there exists $d \in R$ such that $(a, b) = (d)$.*

This means that there exist $x, y \in R$ such that $ax + by = d$, so this indeed the same as the form of Bezout's lemma we are used to. We now give a definition of the greatest common divisor of two elements:

Definition 2.3.5. Let R be a Euclidean domain. Let $a, b \in R$. We say that a **divides** b and write $a \mid b$ if there is $x \in R$ such that $ax = b$. We call $d \neq 0$ a **greatest common divisor** of a and b if d is a generator of the ideal (a, b).

Proposition 16. *Let R be a Euclidean domain and let $a, b \in R$. Then d is a greatest common divisor of a and b if and only if d is a common divisor of a and b and for any other common divisor d' of a and b, we have $d' \mid d$.*

Proof. First, suppose that d is a greatest common divisor of a and b, so that $(a, b) = (d)$. Suppose that $d' \mid a$ and $d' \mid b$. Then we may write $a = d'r$ and $b = d's$ for some $r, s \in R$. As $(a, b) = d$, there are $x, y \in R$ with $ax + by = d$, so $d'(rx + sy) = d$, which says $d' \mid d$.

Conversely, suppose that d has the stated property. We need to show that $(a, b) = (d)$. We know that $(a, b) = (d')$ for some $d' \in R$, so there are $x, y \in R$ with $ax + by = d'$. As $d \mid a$

and $d \mid b$ this means $d \mid d'$. On the other had, as $a, b \in (d')$ this means that $d' \mid a$ and $d' \mid b$, so $d' \mid d$. This means there are $r, s \in R$ with $d' = rd$ and $d = sd'$, so $(1 - rs)d = 0$. As R is a domain, this means $rs = 1$ so that d and d' differ by a unit multiple, so $(a, b) = (d') = (d)$ as desired. □

We will use the notation $\gcd(a, b)$ to mean a choice of greatest common divisor as in the previous section. Just like in $\mathbb{Z}[i]$, we can use the same proof as in $\mathbb{Z}$ to show that for any $a, b \in R$ not both zero, the last non-zero remainder upon running the Euclidean algorithm is a greatest common divisor of a and b.

Next, we need a notion of what it means for an element in R to be prime.

Definition 2.3.6. Let R be an integral domain. We call an element $p \in R$ **prime** if p is not a unit and for any $a, b \in R$, if $p \mid ab$ then $p \mid a$ or $p \mid b$.

Proposition 17. *Let R be a Euclidean domain. Then $p \in R$ is prime if and only if $p = ab$ for $a, b \in R$ means either a or b is a unit.*

Proof. First, suppose that p is prime in R, and $p = ab$. Then $p \mid ab$ clearly, so $p \mid a$ or $p \mid b$ by definition. Without loss of generality, assume that $p \mid a$, so $a = pr$ for some $r \in R$. Multiplying by b shows that $br = 1$, so that b is a unit.

Conversely, suppose that $p \in R$ has no non-trivial divisors, and suppose that $p \mid ab$. If $p \nmid a$, the claim is that $(a, p) = (1)$, so that there are $x, y \in R$ with $ax + py = 1$, and then the usual argument of multiplying by b shows that $p \mid b$, so that p is prime. Let d be a greatest common divisor of a and p, so $(a, p) = (d)$. This means that $p = dr$ for some $r \in R$, and therefore either d or r is a unit. Since $a \in (d)$, there is $s \in R$ with $a = ds$. If d is not a unit, then this means r is a unit, and so $a = p(r^{-1}s)$, which is a contradiction, as $p \nmid a$. Therefore, d is a unit, and we're done. □

Note that in the above argument, the first half did not use anywhere that R was a Euclidean domain. This means that in *any* integral domain, if p is prime then p has no non-trivial divisors. However, the backwards direction used that R was Euclidean in a non-trivial way. If R is a domain where every ideal of R is principal, we call R a *principal ideal domain* (PID for short). We then have a well-defined notion of a greatest common divisor in a PID by the same definition, and the same argument shows that the above proposition holds in any PID. For an example of a ring where these notions are *not* the same, see the exercises.

We are now ready to discuss unique factorization in Euclidean domains. We'd like to give a proof in the spirit that was given for $\mathbb{Z}, F[x], \mathbb{Z}[i]$. All of those proofs proceeded by induction on the norm of an element, which all relied on the key fact that in these rings, if $a \mid b$, then $N(a) \leq N(b)$. Our definition of a Euclidean domain does not guarantee that a norm function has this property. The standard proof in many abstract algebra books is to use the property that a PID (and therefore, a Euclidean domain) is *Noetherian*, which means that any ascending chain of ideals $I_1 \subseteq I_2 \subseteq \ldots$ eventually stabilizes, i.e. there is

some n such that $I_n = I_k$ for all $k \geq n$.

However, it turns out that if R is a Euclidean domain, we can always find a *different* norm for R that *does* have the desired property.

Lemma 2.3.7. *Let R be a Euclidean domain with norm function N. There there exists another norm N' on R such that for all non-zero $a, b \in R$, $N'(a) \leq N'(ab)$ and R is Euclidean with respect to N'.*

Proof. Define $N' : R \to \mathbb{N}$ by $N'(0)$ and $N'(a) = \min_{b \neq 0} N(ab)$, i.e. N' is the smallest value of N among all non-zero multiples of a. First, we show that $N'(a) \leq N'(ab)$ for non-zero $a, b \in R$. By definition, we can write $N'(ab) = N(abc)$ for some $c \neq 0 \in R$. Then as abc is a non-zero multiple of a, $N'(a) \leq N(abc) = N'(ab)$. Next, we show that R admits a division algorithm with respect to N'. Let $a, b \in R$ with $b \neq 0$. Write $N'(b) = N(bc)$ for some $c \neq 0 \in R$. As R has a division algorithm with respect to N, there are $q, r \in R$ with $a = (bc)q + r$ and $r = 0$ or $N(r) < N(bc)$. Set $q' = qc$, so $a = bq' + r$ with $r = 0$ or $N(r) < N(bc)$. As $N(bc) = N'(b)$, and $N'(r) \leq N(r \cdot 1) = N(r)$, we have $N'(r) < N'(b)$, so R is Euclidean with respect to N'. □

Theorem 2.3.8. *Let R be a Euclidean domain. For any non-zero $r \in R$ that is not a unit, there exist $p_1, \ldots, p_k \in R$ prime such that $r = p_1 \cdots p_k$. This factorization is unique in the following sense: if $r = q_1 \cdots q_\ell$ is another prime factorization of r, then $k = \ell$ and $p_i = u_i q_i$ for some units u_i.*

Proof. By the above, we may assume that the norm on R satisfies the inequality in the statement of the lemma. The proof then proceeds as usual, by induction on the norm of a. We leave the details as an exercise. □

2.4 Exercises

1. Verify that the same proof of the division algorithm works in an arbitrary polynomial ring $R[x]$ when $g(x)$ is monic.

2. Run the Euclidean algorithm backwards in $\mathbb{Q}[x]$ to find $p(x), q(x) \in \mathbb{Q}[x]$ such that $(4x^3 - 2x^2 - 3x + 1)p(x) + (2x^2 - x - 2)q(x) = 1$.

3. Show that $\mathbb{Z}[x]$ does not have a "Bezout lemma" by showing that $\gcd(2, x) = 1$, but there is no way to write 1 as a $\mathbb{Z}[x]$-linear combination of 2 and x.

4. Let F be a field. Prove there are infinitely many monic irreducible polynomials in $F[x]$. Does your proof still work in $\mathbb{Z}[x]$?

5. Let $p(x), q(x), r(x) \in \mathbb{R}[x]$ be non-constant. Prove that if $\gcd(p(x), q(x)) = 1$ and $r(x) \mid p(x) + q(x)$, then there exist polynomials $a(x), b(x) \in \mathbb{R}[x]$ such that $a(x)p(x) + (q(x)r(x))b(x) = 1$.

6. Use norms to find prime factorizations of $3 + 5i$ and $1 + 18i$ in $\mathbb{Z}[i]$.

7. Run the Euclidean algorithm backwards in $\mathbb{Z}[i]$ to find $\alpha, \beta \in \mathbb{Z}[i]$ such that $(8-i)\alpha + (9+2i)\beta = 2+i$.

8. Suppose that $\alpha, \beta, \gamma \in \mathbb{Z}[i]$ with $\alpha^7 \mid \beta^8$ and $\beta^5 \mid \gamma^4$. Prove that $\alpha \mid \gamma$.

9. Define $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. This forms a ring, with the operations of addition and multiplication in $\mathbb{Z}[\sqrt{-5}]$ analogous to how addition and multiplication in $\mathbb{Z}[i]$ work. For $\alpha \in \mathbb{Z}[\sqrt{-5}]$, the *norm* of α is defined by $N(\alpha) = (a + b\sqrt{-5})(a - b\sqrt{-5}) = a^2 + 5b^2$. The same argument that shows the norm in $\mathbb{Z}[i]$ is multiplicative also shows the norm in $\mathbb{Z}[\sqrt{-5}]$ is multiplicative.

   (a) A number $\alpha \in \mathbb{Z}[\sqrt{-5}]$ is called *irreducible* if $\alpha = \beta\gamma$ in $\mathbb{Z}[\sqrt{-5}]$ means either $\beta = \pm 1$ or $\gamma = \pm 1$. Show that 3 is irreducible in $\mathbb{Z}[\sqrt{-5}]$.

   (b) (If you've seen ring theory) A number $\alpha \in \mathbb{Z}[\sqrt{-5}]$ is called *prime* if $\alpha \mid \beta\gamma$ in $\mathbb{Z}[\sqrt{-5}]$ means $\alpha \mid \beta$ or $\alpha \mid \gamma$. Show that $3 + 2\sqrt{-5}$ is prime in $\mathbb{Z}[\sqrt{-5}]$.

   (c) Show that 3 is *not* prime in $\mathbb{Z}[\sqrt{-5}]$ by showing that $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$, but 3 divides neither $1 \pm \sqrt{-5}$. Deduce that $\mathbb{Z}[\sqrt{-5}]$ does not have unique factorization into irreducibles.

10. A *Pythagorean triple* is a tuple (x, y, z) of positive integers such that $x^2 + y^2 = z^2$. A Pythagorean triple (x, y, z) is called *primitive* if $\gcd(x, y, z) = 1$. You'll use the unique factorization of $\mathbb{Z}[i]$ to classify all primitive Pythagorean triples.

    (a) Show that if (x, y, z) is primitive, then z is odd and exactly one of x, y is even.

    (b) Let $\alpha, \beta, \gamma \in \mathbb{Z}[i]$ such that α, β are relatively prime and $\alpha\beta = \gamma^2$. Prove that both α and β are perfect squares in $\mathbb{Z}[i]$.

    (c) Suppose $\alpha \in \mathbb{Z}[i]$. Show that $(1+i) \mid \alpha$ if and only if $N(\alpha)$ is even.

    (d) Let $\alpha = x + yi$ with x odd, y even, and $\gcd(x, y) = 1$. Prove that α and $\bar{\alpha}$ are relatively prime. *(Hint: if δ is a gcd, show that $\delta \mid 2 = -i(1+i)^2$. Then, use part (c))*

    (e) Deduce that (x, y, z) is a primitive Pythagorean triple if and only if $x = m^2 - n^2, y = 2mn, z = m^2 + n^2$ for some integers m, n.

11. Prove theorem 2.3.8.

Part II

Modular Arithmetic

Chapter 3

Arithmetic in quotient rings

3.1 The quotient ring and $\mathbb{Z}/n\mathbb{Z}$

Let R be a commutative ring and let I be an ideal of R. We start with a general ring theoretic construction, before specializing to our particular cases of interest.

Define an equivalence relation on R by $a \sim b$ if and only if $a - b \in I$. One may check that this is indeed an equivalence relation. For any $a \in R$, we will interchangeably write any of the following to mean the equivalence class a: $[a]$,$a + I$, $\bar{a}$, $a \bmod I$.

Definition 3.1.1. The **quotient ring** R/I is, as a set, $R/I = \{[a] : a \in R\}$ equipped operations $+$ and $\cdot$ defined by $[a] + [b] := [a + b]$ and $[a] \cdot [b] := [ab]$. The additive and multiplicative identity elements with respect to these operations are $[0]$ and $[1]$.

We leave it to the reader to check that these operations are well-defined (i.e. do not depend on a choice of representative of the equivalence class) and satisfy the ring axioms. In terms of notation, we usually write $a \equiv b \bmod I$ to mean $[a] = [b]$ in R/I, or when $I = (r)$ is principal, we write $a \equiv b \bmod r$.

There is a natural ring homomorphism $\pi : R \to R/I$ given by $\pi(r) = r \bmod I$ called the **reduction map**. The process of passing from a ring R to a quotient ring R/I via the reduction map is a powerful technique for understanding R.

In $\mathbb{Z}$, all ideals look like (n) for some integer $n > 0$. The quotient ring $\mathbb{Z}/(n)$ is called the **integers modulo n**, and is often instead written $\mathbb{Z}/n\mathbb{Z}$. By the division algorithm, any integer $a > 0$ can be written as $a = nq + r$ with $0 \leq r < n$. This means that $a \equiv r \bmod n$, and so any $[a] \in \mathbb{Z}/n\mathbb{Z}$ looks like $[r]$ for some $0 \leq r \leq n - 1$. Since all these equivalence classes are *distinct*, this means that $\mathbb{Z}/n\mathbb{Z} = \{[r] : 0 \leq r \leq n - 1\}$, so that $\mathbb{Z}/n\mathbb{Z}$ is a finite ring of size n.

Example 3.1.2. In $\mathbb{Z}/10\mathbb{Z}$, we have $5 + 5 \equiv 10 \equiv 0 \bmod 10$, $5 + 8 \equiv 13 \equiv 3 \bmod 10$, $-6 + 2 \equiv -4 \equiv 6 \bmod 10$, and $3 \cdot 7 \equiv 21 \equiv 1 \bmod 10$.

Example 3.1.3. In $\mathbb{Z}/7\mathbb{Z}$, we have $2^3 \equiv 8 \equiv 1 \bmod 7$ and $3^4 \equiv 9^2 \equiv 2^2 \equiv 4 \bmod 7$.

General advice for doing computations in $\mathbb{Z}/n\mathbb{Z}$ would be to "reduce" as you go along, as seen in the computation of $3^4 \bmod 7$ in the previous example. Other useful tricks are changing to a different representative of the equivalence class to one of smaller absolute value. For example, it might be annoying to try computing $1001^2 \bmod 1003$ by directly squaring 1001, however this is easy if you note that $1001 \equiv -2 \bmod 1003$, so that $1001^2 \equiv (-2)^2 \equiv 4 \bmod 1003$.

Definition 3.1.4. We say that $[a] \in \mathbb{Z}/n\mathbb{Z}$ is a **unit** (or is **invertible**) if there is $[b] \in \mathbb{Z}/n\mathbb{Z}$ such that $[a][b] = [1]$ in $\mathbb{Z}/n\mathbb{Z}$. If such a $[b]$ exists, we write $[b] = [a]^{-1}$ and call $[b]$ the **inverse** of $[a]$. The units of $\mathbb{Z}/n\mathbb{Z}$ are denoted by $(\mathbb{Z}/n\mathbb{Z})^\times$.

Our first goal is the classification of the units mod n, which is a familiar result in disguise.

Theorem 3.1.5. *$[a] \in \mathbb{Z}/n\mathbb{Z}$ is a unit of and only if $\gcd(a, n) = 1$.*

Proof. By definition, $[a]$ is a unit in $\mathbb{Z}/n\mathbb{Z}$ if and only if there is $[b]$ such that $[a][b] = [1]$, i.e. $ab \equiv 1 \bmod n$. By definition of $\mathbb{Z}/n\mathbb{Z}$, this is true if and only if $ab - 1 = ny$ for some $y \in \mathbb{Z}$, i.e. $ab - ny = 1$ in $\mathbb{Z}$. By Bezout's lemma, this is equivalent to saying that $\gcd(a, n) = 1$. □

The units of $\mathbb{Z}/n\mathbb{Z}$ are the elements of $\mathbb{Z}/n\mathbb{Z}$ that we can divide by: dividing by $[a]$ in $\mathbb{Z}/n\mathbb{Z}$ is the same thing as multiplying by $[a]^{-1}$, just like how in $\mathbb{Q}$ dividing by n is the same as multiplying by n^{-1}.

Corollary 3.1.6. *$\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime.*

Proof. If $n = p$ is prime, then $\gcd(a, p) = 1$ for all $1 \leq a \leq p - 1$ so all non-zero elements of $\mathbb{Z}/p\mathbb{Z}$ are units, which by definition means that $\mathbb{Z}/p\mathbb{Z}$ is a field. On the other hand, if $n = ab$ is a non-trivial factorization, then we have $ab \equiv 0 \bmod n$ with $a \not\equiv 0 \bmod n$ and $b \not\equiv 0 \bmod n$, which is not possible in a field. □

The quotients $\mathbb{Z}/p\mathbb{Z}$ for p prime are examples of *finite* fields. When we want to think of $\mathbb{Z}/p\mathbb{Z}$ as a field and not just as a ring, we will write $\mathbb{F}_p$ instead. We'll later see that all finite fields are "built" from the finite fields $\mathbb{F}_p$, so that these are the most important examples of finite fields. We'll also later see how the rings $\mathbb{Z}/p\mathbb{Z}$ are the most important rings for understanding the structure of $\mathbb{Z}/n\mathbb{Z}$, analogously to how understanding $\mathbb{Z}$ means understanding primes.

Since we know how to determine if $a \bmod n$ is a unit or not, how can we find its inverse? This can be done by solving the Bezout equation.

Example 3.1.7. Computing the inverse of $5 \bmod 13$ is the same as trying to solve the equation $5x \equiv 1 \bmod 13$. This means finding integers x, y such that $5x = 1 + 13y$, i.e. $5x - 13y = 1$. By inspection (or by running the usual algorithm), note that $(-5, -2)$ is one such solution, so that $5 \cdot -5 \equiv 1 \bmod 13$. This means that $5^{-1} \equiv -5 \equiv 8 \bmod 13$.

Next, we determine when a linear equation in $\mathbb{Z}/n\mathbb{Z}$ is solvable. Just like in $\mathbb{Z}$, this will not be particularly difficult to do.

Theorem 3.1.8. *The equation* $[a]x = [b]$ *in* $\mathbb{Z}/n\mathbb{Z}$ *is solvable if and only if* $\gcd(a, n) \mid b$. *If* $[x_0]$ *is one solution, then all solutions are given by* $[x_0 + k\frac{n}{d}]$ *for* $0 \leq k \leq d - 1$ *where* $d = \gcd(a, n)$.

Proof. Saying $[a]x = [b]$ in $\mathbb{Z}/n\mathbb{Z}$ is the same thing as saying $ax \equiv b \bmod n$. When is the congruence $ax \equiv b \bmod n$ solvable? This happens if and only if there is $x, y \in \mathbb{Z}$ such that $ax = b + ny$, i.e. $ax - ny = b$. We know this has integer solutions if and only if $\gcd(a, n) \mid b$.

Now, suppose that we have two solutions, $[x]$ and $[x_0]$ in $\mathbb{Z}/n\mathbb{Z}$. This means that $ax \equiv b \bmod n$ and $ax_0 \equiv b \bmod n$, so that $a(x - x_0) \equiv 0 \bmod n$. This means that there is $y \in \mathbb{Z}$ such that $a(x - x_0) = ny$, so dividing through by $d = \gcd(a, n)$ means that $a'(x - x_0) = n'y$ where $a' = \frac{a}{d}$ and $n' = \frac{n}{d}$. Since $\gcd(a', n') = 1$, this means that $n' \mid (x - x_0)$, so that $x = x_0 + n'k$ for some $k \in \mathbb{Z}$. As k runs through the various integers, what are the possible equivalence classes $[x]$? Well, $[x_0 + n'k] = [x_0 + n'\ell]$ is the same as saying that $n'(k - \ell) \equiv 0 \bmod n$, which means that $n'(k - \ell) = nm$ for some m. Dividing through by n' means $k - \ell = dm$, so that $k \equiv \ell \bmod d$. This means that there are d incongruent equivalence classes mod n, coming from taking $k = 0, \ldots, d - 1$, and therefore, by what we've shown, the only *possible* solutions to $[a]x = [b]$. It's clear that each value of k in this range produces distinct equivalence classes mod n, and it's easy to check that they *are* solutions, so therefore these consist of *all* solutions. □

The above theorem highlights some of the strangeness that happens when you work in *rings* and not *fields*. A linear equation in $\mathbb{Z}/n\mathbb{Z}$ can have more than one solution! When $[a]$ is a unit, however, there is indeed a unique solution to $[a]x = [b]$, namely, $x = [a]^{-1}[b]$.

Example 3.1.9. What are the solutions to the equation $[42]x = [12]$ in $\mathbb{Z}/78\mathbb{Z}$? Equivalently, this is the same as asking how do you solve the congruence equation $42x \equiv 12 \bmod 78$? Since $\gcd(42, 78) = 6$, this means there are 6 solutions to this equation. To find one, first we find a solution to $42x + 78y = 12$ in $\mathbb{Z}^2$. Running the Euclidean algorithm, we find that $42 \cdot 2 + 78 \cdot -1 = 6$, so that $42 \cdot 4 + 78 \cdot -2 = 12$. Working mod 78, this means $42 \cdot 4 \equiv 12 \bmod 78$, so that one solution is 4 mod 78. The above theorem then says that all solutions are given by $x \equiv 4, 17, 30, 43, 56, 69 \bmod 78$.

The strength of working inside $\mathbb{Z}/n\mathbb{Z}$ instead of $\mathbb{Z}$ is that it translates problems of divisibility into a *finite* setting. It's very similar to how in linear algebra it's much easier to prove things in a finite dimensional vector space by reducing to just working with a basis. Statements about divisibility get translated to statements in $\mathbb{Z}/n\mathbb{Z}$ by simply realizing that saying $n \mid a$ is the same thing as saying $a \equiv 0 \bmod n$. For example, suppose we wished to prove that $3 \mid n^3 - n$ for any integer n. In $\mathbb{Z}$, one such argument would be to factor $n^3 - n = (n - 1)n(n + 1)$ into a product of three consecutive integers, one of which must therefore be divisible by 3. Working mod 3, however, this just means verifying that $n^3 \equiv n \bmod 3$ for any integer n. The only congruence classes mod 3 are $0, 1, 2 \bmod 3$, so just verifying directly that $0^3 \equiv 0 \bmod 3$, $1^3 \equiv 1 \bmod 3$, and $2^3 \equiv 8 \equiv 2 \bmod 3$ is enough!

In particular, one very important application of passing to the quotient is that it can help detect obstructions to solutions of Diophantine equations.

Example 3.1.10. Does $x^2 + y^2 = 3$ have integer solutions? It's easy to verify by hand that this is not possible, since clearly $0 \leq |x|, |y| \leq 1$. However, a better argument is the following. Suppose that $x^2 + y^2 = 3$ did have a solution (x, y). Then working mod 4, we would have $x^2 + y^2 \equiv 3 \bmod 4$. All integers square to either 0 or 1 mod 4, so the left hand side is congruent to $0, 1, 2 \not\equiv 3 \bmod 4$, which is a contradiction. Therefore, there cannot be any integer solutions.

3.2 Quotients of $\mathbb{Z}[i]$

Next, we'll apply the quotient ring construction to $\mathbb{Z}[i]$. Any ideal of $\mathbb{Z}[i]$ is principal, so $I = (\alpha)$ for some $\alpha \in \mathbb{Z}[i]$. Saying that $\beta \equiv \beta' \bmod \alpha$ means that $\beta = \beta' + \alpha\gamma$ for some $\gamma \in \mathbb{Z}[i]$, or equivalently, that $\beta - \beta' = \alpha\gamma$. What do $\mathbb{Z}[i]$-multiples of α look like? Set $\alpha = a + bi$ and $\gamma = c + di$. By direct computation, one can verify that $\alpha\gamma = c\alpha + d(i\alpha)$. This means that $\alpha\gamma$ is a $\mathbb{Z}$-linear combination of α and $i\alpha$!

In the complex plane, α and $i\alpha$ are orthogonal to each other, and therefore form an $\mathbb{R}$-basis of $\mathbb{C}$. Let $L = \mathrm{Span}_{\mathbb{Z}}\{\alpha, i\alpha\}$ be the lattice in $\mathbb{C}$ generated by α and $i\alpha$. As a set, L consists of all the $\mathbb{Z}[i]$-multiples of α!

Definition 3.2.1. Let $\alpha \in \mathbb{Z}[i]$ and let $L = \mathrm{Span}_{\mathbb{Z}}\{\alpha, i\alpha\}$. The **fundamental parallelogram** of L is the parallelogram in $\mathbb{C}$ whose vertices are given by $0, \alpha, i\alpha, \alpha + i\alpha$.

The fundamental parallelogram is named as such because any $z \in \mathbb{C}$ can be translated to a point inside the fundamental parallelogram by moving along the lattice L. What this means, then, is that for any $\beta \in \mathbb{Z}[i]$, there is $\gamma \in \mathbb{Z}[i]$ inside the fundamental parallelogram of L such that $\beta \equiv \gamma \bmod \alpha$. This means that the equivalence classes inside $\mathbb{Z}[i]/(\alpha)$ come from the Gaussian integers in the fundamental parallelogram.

Example 3.2.2. Let $\alpha = 1 + 2i$, so $i\alpha = -2 + i$. The lattice of interest is $L = \mathrm{Span}_{\mathbb{Z}}\{1 + 2i, -2 + i\}$, and the fundamental parallelogram has vertices $0, 1 + 2i, -2 + i, -1 + 3i$. To find a set of representatives for $\mathbb{Z}[i]/(1 + 2i)$, we need to find the incongruent lattice points in the fundamental parallelogram. The first observation is that all lattice points in the interior are incongruent. The second observation is that we need only consider boundary lattice points on the *lower* half of the parallelogram. This is because boundary points on the upper half can be translated down to a point on the lower half by moving along the boundary. The final observation is that all corner points are congruent to 0, as they are simply just $\mathbb{Z}[i]$-multiples of α. We then find that $\mathbb{Z}[i]/(1 + 2i) = \{[0], [i], [2i], [-1 + i], [-1 + 2i]\}$ so that $\mathbb{Z}[i]/(1 + 2i)$ has size 5.

In general, what is the size of $\mathbb{Z}[i]/(\alpha)$? By our observations, this is determined by the number of lattice points inside the interior of the fundamental parallelogram of the lattice L, and *half* the number of lattice points on the boundary. However, this counts the origin *twice*, so subtracting 1 gives the correct count. There is a result known as *Pick's theorem*, which says that this lattice point count is precisely the area of the parallelogram, which is $N(\alpha)$. We'll later see how Pick's theorem can be proved.

Algebraically, arithmetic in $\mathbb{Z}[i]/(\alpha)$ is similar to arithmetic in $\mathbb{Z}/n\mathbb{Z}$.

Example 3.2.3. In $\mathbb{Z}[i]/(1+2i)$, we have $(-1+2i)(2i) \equiv -4-2i \equiv i \bmod 1+2i$ because $-4-2i = (-2+i)\alpha + i$. Additionally, $(2i)^2 \equiv -4 \equiv -1+i \bmod 1+2i$.

Both of these computations can also be done geometrically by drawing the lattice and seeing which Gaussian integer $-4-2i$ and -4 get translated to, respectively. However, in general, it's going to be significantly easier to just perform the division algorithm to compute the remainder upon division by α.

Proposition 18. *$[\beta]$ is a unit in $\mathbb{Z}[i]/(\alpha)$ if and only if α, β are relatively prime in $\mathbb{Z}[i]$.*

Proof. The proof is the same as in $\mathbb{Z}/n\mathbb{Z}$: $[\beta]x = [1]$ is solvable in $\mathbb{Z}[i]/(\alpha)$ if and only if $\beta x + \alpha y = 1$ is solvable in $\mathbb{Z}[i]^2$, which is equivalent to saying that $\gcd(\alpha, \beta)$ is a unit. □

Just like in $\mathbb{Z}/n\mathbb{Z}$, finding the inverse of $[\beta]$ in $\mathbb{Z}[i]/(\alpha)$ can be done by solving the corresponding Bezout equation in $\mathbb{Z}[i]$. The useful corollary of the above result is the following:

Corollary 3.2.4. *If α is prime in $\mathbb{Z}[i]$, then $\mathbb{Z}[i]/(\alpha)$ is a field.*

Example 3.2.5. With $\alpha = 1+2i$ as before, since $N(\alpha) = 5$ is prime, this means that α is prime in $\mathbb{Z}[i]$. Therefore, $\mathbb{Z}[i]/(\alpha)$ is a field of size 5. One can construct an explicit isomorphism $\mathbb{F}_5 \cong \mathbb{Z}[i]/(1+2i)$ by the map $x \bmod 5 \mapsto x \bmod 1+2i$.

3.3 Quotients of $F[x]$ and finite fields

Similar to before, we will mostly focus our attention on quotients of $F[x]$ instead of arbitrary polynomial rings. This is largely because in $F[x]$, all ideals are principal, and so we only have to consider the quotient rings $F[x]/(p(x))$ for $p(x) \in F[x]$, which will be rather simple. For other rings, such as $\mathbb{Z}[x]$, the structure of ideals is more complicated. For example, $I = (2, x)$ is *not* a principal ideal of $\mathbb{Z}[x]$. For the reader familiar with ring theory, it might be a familiar fact that $\mathbb{Z}[x]/(2, x) \cong \mathbb{F}_2$, explicitly realized by the map $p(x) \bmod I \mapsto p(0) \bmod 2$. Understanding such quotients are important in number theory, as, for example, one may identify $\mathbb{Z}[x]/(x^2+1) \cong \mathbb{Z}[i]$ via $p(x) \bmod x^2+1 \mapsto p(i)$, so that the Gaussians integers are actually just a quotient of $\mathbb{Z}[x]$!

Let $I = (p(x))$ be an ideal of $F[x]$. By the division algorithm, any $f(x) \in F[x]$ can be uniquely written as $f(x) = p(x)q(x) + r(x)$ with $r(x) = 0$ or $\deg(r(x)) < \deg(p(x))$, and so in the quotient ring, $f(x) \equiv r(x) \bmod p(x)$. Since $[r(x)]$ are all *distinct* for $r(x)$ satisfying the remainder bound, similar to as in $\mathbb{Z}$ this tells us that $F[x]/(p(x)) = \{[r(x)] : r(x) = 0 \text{ or } \deg(r(x)) < \deg(p(x))\}$. Unlike with $\mathbb{Z}$ or $\mathbb{Z}[i]$, quotients of $F[x]$ may or may not be infinite, depending on if F is finite or not.

Example 3.3.1. The equivalence classes in the quotient ring $\mathbb{R}[x]/(x^2+1)$ are given by $[a+bx]$ for $a, b \in \mathbb{R}$. There is an isomorphism $\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}$ given by $[a+bx] \mapsto a+bi$, so that $\mathbb{R}[x]/(x^2+1)$ is a field.

Example 3.3.2. In $\mathbb{F}_2[x]/(x^2+x+1)$, the equivalence classes are given by $[0], [1], [x], [x+1]$ where we think of $0, 1$ as living in $\mathbb{F}_2$, and there are four elements in the quotient. In

this quotient, addition of the coefficients works "mod 2". For example, we have $x^2 \equiv x+1 \bmod x^2+x+1$ because $x^2+x+1 \equiv 0 \bmod x^2+x+1$. Similarly, $x(x+1) \equiv 1 \bmod x^2+x+1$ because $x(x+1)+1 \equiv 0 \bmod x^2+x+1$.

Example 3.3.3. $\mathbb{F}_3[x]/(x^2+1)$ has 9 elements: the equivalence classes $[a+bx]$ with $a, b \in \mathbb{F}_3$. As $[x^2+1] = [0]$, this tells us that $[x]^2 = [2]$. We have $(2x+1)+(x+2) \equiv 0 \bmod x^2+1$, and $(2x+1)(x+2) = 2x^2+5x+2 \equiv 2x \bmod x^2+1$.

Unsurprisingly, we have similar results as before:

Proposition 19. *$[f(x)] \in F[x]/(p(x))$ is a unit if and only if $\gcd(f(x), p(x)) = 1$.*

Corollary 3.3.4. *$F[x]/(p(x))$ is a field if and only if $p(x)$ is irreducible in $F[x]$.*

We omit the proofs and leave them to the reader.

Example 3.3.5. We saw before that $\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}$ via an explicit isomorphism, so it's a field. Alternatively, x^2+1 is irreducible in $\mathbb{R}[x]$ because it has no root in $\mathbb{R}[x]$, which tells us that $\mathbb{R}[x]/(x^2+1)$ is a field (but without helping us identify it's isomorphism class). Similarly, x^2+x+1 is irreducible in $\mathbb{F}_2[x]$ because it has no root in $\mathbb{F}_2$. This tells us that $\mathbb{F}_2[x]/(x^2+x+1)$ is a field of size 4.

Example 3.3.6. Since $\gcd(x+2, x^2-1) = 1$ in $\mathbb{Q}[x] = 1$, this means that $[x+2]$ is invertible in $\mathbb{Q}[x]/(x^2-1)$. To find a multiplicative inverse, we run the Euclidean algorithm backwards to find $\frac{1}{3}(x^2-1)+(x+2)(\frac{1}{3}(2-x)) = 1$ is a solution to the Bezout equation. Taking this mod x^2-1, we find $[x+2]^{-1} = [\frac{1}{3}(2-x)]$.

One particular application of the above corollary is that it gives us a method of constructing finite fields! If F is a finite field of size q and $p(x)$ is an irreducible polynomial of degree d in $F[x]$, then $F[x]/(p(x))$ is a field of size p^d elements, because there are p choices for the d coefficients of a polynomial of degree at most $d-1$. It turns out, that *all* finite fields arise from this construction, although we will not prove this. We will, however, prove the following basic fact:

Proposition 20. *Let F be a finite field. Then $|F| = p^n$ for some prime p and $n \geq 1$.*

Proof. Since F is finite, we must have $k \cdot 1 = \underbrace{1+1+\ldots+1}_{k \text{ times}} = 0$ for some integer k. Let c be the smallest positive integer with this property. Then c must be prime, because if $c = ab$ is a non-trivial factorization, we have $c \cdot 1 = (ab) \cdot 1 = (a \cdot 1)(b \cdot 1) = 0$ in F. Since F is a field, this means $a \cdot 1 = 0$ or $b \cdot 1 = 0$, contradicting the definition of c. Writing p instead of c, there is a natural isomorphism of fields $\mathbb{F}_p \cong \langle 1 \rangle$, given by $x \bmod p \mapsto x \cdot 1$, where $\langle 1 \rangle$ is the additive subgroup of F generated by 1. Therefore, we may view $\mathbb{F}_p$ as a sub-field of F by identifying it with $\langle 1 \rangle$. We then see that $\mathbb{F}_p$ acts on F by multiplication, so we may view F as a vector space over $\mathbb{F}_p$. Since F is finite, it must be a finite dimensional vector space, and therefore we must have $F \cong \mathbb{F}_p^n$ for some n by the classification of finite dimensional vector spaces. Taking cardinalities gives us what we want. $\square$

A slightly easier task than classifying all finite fields is showing that for each $n \geq 1$, there *is* a finite field of size p^n (the above proposition just says that it's *possible*, not that there is one!). This task is equivalent to showing that there is at least one irreducible polynomial of degree n in $\mathbb{F}_p[x]$ for each n. Since irreducible polynomials in $\mathbb{F}_p[x]$ play the role of primes in $\mathbb{F}_p[x]$, this is basically asking if there are primes of each possible "size" in $\mathbb{F}_p[x]$. By adapting Euler's argument that $\sum_p \frac{1}{p}$ diverges to the setting of $\mathbb{F}_p[x]$, one can show that such polynomials must exist. It is, however, a much more complicated argument!

3.4 Exercises

1. Compute the following:
   (a) $17^{48} - 5^{24} \bmod 39$
   (b) $84526 \cdot 862967^3 - 448184 \cdot 591183^2 \bmod 15$
   (c) $1477^{-1} \bmod 9235$
   (d) $1769^{234} \bmod 31$
   (e) $1! + 2! + \ldots + 100! \bmod 25$
2. Solve $140x \equiv 133 \bmod 301$.
3. Prove that each of the following Diophantine Equations have no integer solutions.
   (a) $7x^2 + 2 = y^3$
   (b) $x^3 + y^3 + z^3 = 4$
   (c) $x^2 - 2y^2 = 10$
4. Prove that there are no integers m, n such that $3^m + 3^n + 1$ is a perfect square.
5. Prove that $\frac{1}{5}n^5 + \frac{1}{3}n^3 + \frac{7}{15}n$ is an integer for all $n \in \mathbb{Z}$.
6. Let $n = a_n \cdot 10^n + \ldots + a_1 \cdot 10 + a_0$ be the decimal expansion of n. For example, $123 = 1 \cdot 100 + 2 \cdot 10 + 3$.
   (a) Show that n is divisible by 3 if and only if $a_0 + a_1 + \ldots + a_n$ is divisible by 3. Explain why the same condition also holds for divisibility by 9. Similarly, show that n is divisible by 11 if and only if $a_0 - a_1 + \ldots + (-1)^n a_n$ is divisible by 11.
   (b) Let $\gcd(k, 10) = 1$, and choose b so that $10b \equiv 1 \bmod k$. Show that n is divisible by k if and only if $\frac{n-a_0}{10} + ba_0$ is divisible by k. Apply the divisibility test for $k = 29$ to test if 16559 is divisible by 29.
7. (a) Let p be a prime. Show that the only solutions to $x^2 \equiv 1 \bmod p$ are $x \equiv \pm 1 \bmod p$.
   (b) Prove that if p is prime, then $(p-1)! \equiv -1 \bmod p$. *(Hint: pair up integers with their inverses mod p).*

(c) Prove that if $n > 4$ is composite, then $(n-1)! \equiv 0 \bmod n$.

Combining these two parts says n is prime if and only if $n \nmid (n-1)!$. Of course, this is a very *bad* way of checking an integer is prime, because $(n-1)!$ gets very large, very fast!

8. Carefully go through the steps to construct the quotient ring R/I: verify that $\sim$ is an equivalence relation, prove that addition and multiplication are well-defined, and that R/I is actually a ring.

9. Verify that the map in example 3.2.5 is an isomorphism of fields.

10. Find a set of representatives for the ring $\mathbb{Z}[i]/(2+3i)$.

11. Compute $[1+2i]^{-1}$ in $\mathbb{Z}[i]/(2+3i)$.

12. Show that the ideal $(2, x)$ of $\mathbb{Z}[x]$ is not principal.

13. Verify that $\mathbb{Z}[x]/(x^2+1) \cong \mathbb{Z}[i]$ via the map $p(x) \bmod x^2+1 \mapsto p(i)$.

14. Let F be a field and let $p(x)$ be an irreducible polynomial of degree d in $F[x]$, so that $K := F[x]/(p(x))$ is a field.

    (a) Identify F with its image in K. Show that $\alpha = [x] \in K$ is a root of $p(T) \in K[T]$.

    (b) Explain why $\{1, \alpha, \alpha^2, \ldots, \alpha^{d-1}\}$ is a basis for K as an F-vector space.

    (c) Let $p(x) = x^4 + x + 1 \in \mathbb{F}_2[x]$. Prove $p(x)$ is irreducible in $\mathbb{F}_2[x]$ and write down all the elements of K in terms of the basis in part (b). Then, compute $\alpha^6 + \alpha + 1$.

15. Let $p(x)$ be an irreducible polynomial of degree d in $\mathbb{F}_p[x]$, so that $F = \mathbb{F}_p[x]/(p(x))$ is a field of size p^d.

    (a) Prove that for any $\alpha, \beta \in F$, $(\alpha+\beta)^p = \alpha^p + \beta^p$.

    (b) Viewing $\mathbb{F}_p$ as a subfield of F, suppose that $g(x) \in \mathbb{F}_p[x]$ has a root α in F. Prove that α^{p^k} is a root of $g(x)$ for all $k \geq 0$.

    (c) Let $F = \mathbb{F}_2[x]/(x^4+x+1)$, which by the previous problem is a field where x^4+x+1 has a root α. Find all the roots of $f(x)$ in F. Find all the roots of $g(x) = x^4 + x^3 + 1$ in F. Make sure to write your answers in terms of the basis $\{1, \alpha, \alpha^2, \alpha^3\}$!

16. Let α be a root of $x^2 + x + 2$ in $F = \mathbb{F}_3[x]/(x^2+x+2)$. Compute the other root of $x^2 + x + 2$ in F.

Chapter 4

The structure of $(\mathbb{Z}/n\mathbb{Z})^\times$

4.1 Chinese remainder theorem

The Chinese Remainder Theorem is a "structure theorem" for the ring $\mathbb{Z}/n\mathbb{Z}$. It is arguably the most important theorem in all of number theory!

Theorem 4.1.1 (Chinese remainder theorem)**.** *Let n, m be positive integers with $\gcd(n, m) = 1$. Then $\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ as rings.*

Proof. Define a map $f : \mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ by $x \bmod mn \mapsto (x \bmod m, x \bmod n)$. First, we must check that f is well-defined. Suppose that $x \equiv x' \bmod mn$, so that $mn \mid x - x'$. Then we certainly have that $m \mid x - x'$ and $n \mid x - x'$ so that $x \equiv x' \bmod m$ and $x \equiv x' \bmod n$, meaning $f(x \bmod mn) = f(x' \bmod mn)$.

Next, we check the map is injective. If $f(x \bmod mn) = f(y \bmod mn)$, then $x \equiv y \bmod m$ and $x \equiv y \bmod n$. This means $m \mid x - y$ and $n \mid x - y$, which means that $mn \mid x - y$ because m, n are relatively prime. Therefore, $x \equiv y \bmod mn$. To show that f is surjective, for any pair $(a \bmod m, b \bmod n)$ we wish to find $x \bmod mn$ such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$. We do this as follows. From $x \equiv a \bmod m$ we must have $x = a + mk$ for some integer k. Therefore, $a + mk \equiv b \bmod n$, so that $mk \equiv b - a \bmod n$. Since $\gcd(m, n) = 1$, m is invertible mod n, and so $k \equiv m^{-1}(b - a) \bmod n$. Let ℓ be any integer with $\ell \equiv m^{-1} \bmod n$. Then we see that $a + m\ell(b - a) \bmod mn$ is the desired class that works.

That f is a ring homomorphism easily follows from the definitions of addition and multiplication in the product rings. □

By writing n in terms of its prime factorization and repeatedly applying the Chinese Remainder Theorem, we get the following corollary (often times *also* called the Chinese Remainder Theorem).

Corollary 4.1.2. *Let $n = p_1^{e_1} \cdots p_k^{e_k}$ be the prime factorization of n. Then as rings, $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \ldots \times \mathbb{Z}/p_k^{e_k}\mathbb{Z}$.*

The Chinese Remainder Theorem is important because it says that if you understand whats happening in $\mathbb{Z}/p^e\mathbb{Z}$ for prime p and $e \geq 1$, then you can "glue" all the information

together to understand what's happening in $\mathbb{Z}/n\mathbb{Z}$. The proof of the Chinese Remainder Theorem is *constructive*, and so by following the proof we can explicitly solve systems of congruences.

Example 4.1.3. Suppose we wish to find all integers that satisfy the system of congruences

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

The first equation says that $x = 1 + 2k$ for some integer k, so plugging into the second yields $1 + 2k \equiv 2 \bmod 3$. This means that $k \equiv 2 \bmod 3$, so $k = 2 + 3\ell$ for some ℓ. Plugging back in, we get that $x = 1 + 2(2 + 3\ell) = 5 + 6\ell$. Finally, plugging this into the last equation, $5 + 6\ell \equiv 3 \bmod 5$ means that $\ell \equiv 3 \bmod 5$, so that $\ell = 3 + 5s$ for some s. Plugging this back in, $x = 5 + 6(3 + 5s) = 23 + 30s$. This says that $x \equiv 23 \bmod 30$, and clearly any such choice of integer actually works. Therefore, any x with $x \equiv 23 \bmod 30$ solves the system. Ring theoretically, this computation can be interpreted as computing the pre-image of $(1 \bmod 2, 2 \bmod 3, 3 \bmod 5)$ under the isomorphism $\mathbb{Z}/30\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$ from the Chinese Remainder Theorem.

Since the Chinese remainder theorem is an isomorphism of rings, it induces an isomorphism on unit groups, too:

Corollary 4.1.4. *Let n, m be positive integers with $\gcd(n, m) = 1$. Then $(\mathbb{Z}/mn\mathbb{Z})^\times \cong (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$*

Proof. The Chinese Remainder Theorem says that $\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, so $(\mathbb{Z}/mn\mathbb{Z})^\times \cong (\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z})^\times$. By definition of multiplication in the product ring, we see that ($a \bmod m, b \bmod n$) is a unit if and only if there is ($x \bmod m, y \bmod n$) such that $ax \equiv 1 \bmod m$ and $by \equiv 1 \bmod n$, which is equivalent to saying that $a \bmod m$ and $b \bmod n$ are units in $\mathbb{Z}/m\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$ respectively. This means that $(\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z})^\times = (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$. □

Definition 4.1.5. We define $\varphi(n) = \#\{1 \leq x \leq n : \gcd(x, n) = 1\}$.

From what we know about $\mathbb{Z}/n\mathbb{Z}$, saying that $\gcd(x, n) = 1$ is the same as saying that x is a unit mod n, so this means that $\varphi(n) = |(\mathbb{Z}/n\mathbb{Z})^\times|$ counts the size of the unit group mod n.

Proposition 21. *Let m, n be positive integers with $\gcd(m, n) = 1$. Then $\varphi(mn) = \varphi(m)\varphi(n)$.*

Proof. This follows immediately from the above corollary by definition of φ and noting that the cardinality of a product ring is the product of cardinalities. □

Writing n in terms of its prime factorization, we get an immediate corollary.

Corollary 4.1.6. *Let $n = p_1^{e_1} \cdots p_k^{e_k}$ be the prime factorization of n. Then $\varphi(n) = \varphi(p_1^{e_1}) \cdots \varphi(p_k^{e_k})$.*

To compute $\varphi(n)$, then, means we just have to compute $\varphi(p^e)$ for prime p and $e \geq 1$. However, this is rather simple.

Proposition 22. *let p be a prime and $e \geq 1$. Then $\varphi(p^e) = p^{e-1}(p-1)$.*

Proof. $\varphi(p^e)$ counts, by definition, the number of integers between 1 and p^e that are co-prime to p^e. There are p^{e-1} integers in this range *not* co-prime to p^e, namely, the p^{e-1} multiples of p. Therefore, there are $p^e - p^{e-1} = p^{e-1}(p-1)$ integers that *are* co-prime. □

Corollary 4.1.7. *$\varphi(n) = n\prod_{p|n}(1-\frac{1}{p})$, where the product is taken over all primes p that divide n.*

Proof. Write $\varphi(p^e) = p^e(1-\frac{1}{p})$, from which it follows immediately by writing n in terms of a prime factorization and using the multiplicativity of φ. □

Example 4.1.8. $360 = 2^3 \cdot 3^2 \cdot 5$, so $\varphi(360) = \varphi(8)\cdot\varphi(9)\cdot\varphi(5) = 4\cdot 6\cdot 4 = 96$.

4.2 Euler's theorem and orders mod n

Theorem 4.2.1 (Euler)**.** *Let $a, n \in \mathbb{Z}$ with $n > 1$ and $\gcd(a,n) = 1$. Then $a^{\varphi(n)} \equiv 1 \bmod n$.*

Proof. Consider the map $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ give by $f([x]) = [a][x]$. This is a bijection, with inverse given by $f^{-1}([x]) = [a]^{-1}[x]$. There are $\varphi(n)$ units in $\mathbb{Z}/n\mathbb{Z}$, call them $u_1, \ldots, u_{\varphi(n)}$. Since $[a]$ is a unit mod n and the product of two units is a unit, this means that f bijectively maps the set of units onto itself, so that the sets $\{u_1, \ldots, u_{\varphi(n)}\}$ and $\{[a]u_1, \ldots, [a]u_{\varphi(n)}\}$ are the same. Taking the product of all elements in these sets, we see that $\prod_{i=1}^{\varphi(n)} u_i \equiv \prod_{i=1}^{\varphi(n)} au_i \bmod n$. Note that $\prod_{i=1}^{\varphi(n)} u_i$ is invertible because it's a product of units, and therefore dividing through on both sides we find $a^{\varphi(n)} \equiv 1 \bmod n$ as desired. □

When $n = p$ is prime, this result is usually known as *Fermat's little theorem*.

Corollary 4.2.2 (Fermat's little theorem)**.** *Let p be a prime and $p \nmid a$. Then $a^{p-1} \equiv 1 \bmod p$.*

Both Euler's theorem and Fermat's little theorem are special cases of the more general *Lagrange's theorem* in group theory, which says that for any finite group G, and $a \in G$, one has $a^{|G|} = e$, where e is the identity element of G. The proof of Lagrange's theorem is the *same* as the proof as Euler's theorem, proven by showing that the multiplication by a map is a bijection on G. Euler and Fermat's results are the cases where $G = (\mathbb{Z}/n\mathbb{Z})^\times$ and $(\mathbb{Z}/p\mathbb{Z})^\times$, respectively.

Euler's theorem is a very useful computational tool for simplifying exponentials mod n, as it says that the exponent only matters mod $\varphi(n)$.

Example 4.2.3. Suppose we wanted to compute $3^{256} \bmod 100$. Since $\varphi(100) = \varphi(25)\cdot\varphi(4) = 20\cdot 2 = 40$, by Euler's theorem, we have $3^{40} \equiv 1 \bmod 100$. Since $256 \equiv 16 \bmod 40$, this means that $3^{256} \equiv 3^{16} \bmod 100$. This is much more manageable, and we handle this with repeated squaring: $3^2 \equiv 9 \bmod 100$, $3^4 \equiv 81 \equiv -19 \bmod 100$, $3^8 \equiv 361 \equiv 61 \bmod 100$, $3^{16} \equiv 3721 \equiv 21 \bmod 100$.

Euler's theorem is also quite useful for computing k-th roots mod n:

Example 4.2.4. Suppose we wish to solve $x^5 \equiv 2 \bmod 7$. By Fermat's little theorem, any $x \bmod 7$ satisfies $x^6 \equiv 1 \bmod 7$. If we can find k such that $5k \equiv 1 \bmod 6$, then we must have $x \equiv 2^k \bmod 7$ by exponentiating. This is a standard inverse computation, and we find that $k \equiv 5 \bmod 6$, works. This means that $x \equiv 2^5 \equiv 4 \bmod 7$ the unique solution to this congruence equation.

Euler's theorem can even be used to compute inverses, although it's not generally an efficient way to do so by hand.

Example 4.2.5. By Euler's theorem, we have $3^{58} \equiv 1 \bmod 118$, which means that $3^{-1} \equiv 3^{57} \bmod 118$. One may compute that $3^5 \equiv 7 \bmod 118$, $3^{25} \equiv 51 \bmod 118$, $3^{29} \equiv 1 \bmod 118$, so that $3^{57} \equiv 3^{25} \cdot 3^3 \equiv 79 \bmod 118$.

Definition 4.2.6. Let $a, n \in \mathbb{Z}$ with $n > 1$ and $\gcd(a, n) = 1$. The **order mod** n of a, denoted $\mathrm{ord}_n(a)$ is the smallest positive integer k such that $a^k \equiv 1 \bmod n$.

Since $a^{\varphi(n)} \equiv 1 \bmod n$ by Euler's theorem, this tells us that $\mathrm{ord}_n(a)$ is actually defined, and that $\mathrm{ord}_n(a) \leq \varphi(n)$.

Example 4.2.7. We have $\mathrm{ord}_{10}(7) = 4$ because $7^4 \equiv 1 \bmod 10$ and $7, 7^2, 7^3 \not\equiv 1 \bmod 10$.

Proposition 23. *Let $a, n, k \in \mathbb{Z}$ with $n > 1$. Then $a^k \equiv 1 \bmod n$ if and only if $ord_n(a) \mid k$.*

Proof. if $\mathrm{ord}_n(a) \mid k$ then $k = \mathrm{ord}_n(a)\ell$ for some ℓ, from which we see that $a^k \equiv (a^{\mathrm{ord}_n(a)})^\ell \equiv 1 \bmod n$. Conversely, suppose that $a^k \equiv 1 \bmod n$. By the division algorithm, we may write $k = \mathrm{ord}_n(a)q + r$ with $0 \leq r < \mathrm{ord}_n(a)$. Plugging this in, we have $a^k \equiv a^r \equiv 1 \bmod n$. By definition of $\mathrm{ord}_n(a)$, this is possible only if $r = 0$, in which case we see that $\mathrm{ord}_n(a) \mid k$. □

The above proposition is useful for attempting to *compute* $\mathrm{ord}_n(a)$, because in combination with Euler's theorem, this means that $\mathrm{ord}_n(a)$ must be a divisor of $\varphi(n)$. This means that one can compute $\mathrm{ord}_n(a)$ by ruling out divisors one by one.

Example 4.2.8. What is $\mathrm{ord}_{23}(5)$? By Fermat's little theorem, we have $\varphi(23) = 22$, so $\mathrm{ord}_{23}(5)$ must divide 22, and therefore we must have $\mathrm{ord}_{23}(5) = 2, 11, 22$. It's clearly not 2, and we compute $5^{11} \equiv 5 \cdot (25)^5 \equiv 5 \cdot 2^5 \equiv 45 \equiv -1 \bmod 23$, so that it's not 11 either. This then tells us that $\mathrm{ord}_{23}(5) = 22$.

Proposition 24. *Let $a, n, k, \ell \in \mathbb{Z}$ with $n > 1$ and $\gcd(a, n) = 1$. Then $a^k \equiv a^\ell \bmod n$ if and only if $k \equiv \ell \bmod ord_n(a)$.*

Proof. Suppose that $a^k \equiv a^\ell \bmod n$. Multiplying through by $a^{-\ell}$ we see that $^{k-\ell} \equiv 1 \bmod n$, so that by the above proposition, $\mathrm{ord}_n(a) \mid k - \ell$ so that $k \equiv \ell \bmod \mathrm{ord}_n(a)$. Conversely, suppose that $k \equiv \ell \bmod \mathrm{ord}_n(a)$, so that $k = \ell + \mathrm{ord}_n(a)t$ for some t. It's then clear that $a^k \equiv a^\ell \cdot (a^{\mathrm{ord}_n(a)})^t \equiv a^\ell \bmod n$. □

The above proposition says that exponentiation only matters modulo $\mathrm{ord}_n(a)$, which means that knowing the value of $\mathrm{ord}_n(a)$ is of computational interest. Our next proposition tells us how to compute the order of a power.

Proposition 25. *Let $a, n, k \in \mathbb{Z}$ with $n > 1$, $k \geq 1$, and $\gcd(a, n) = 1$. Then $ord_n(a^k) = \frac{ord_n(a)}{\gcd(k, ord_n(a))}$.*

Proof. Set $t = \mathrm{ord}_n(a)$ and $d = \gcd(k, t)$. Then we may write $t = t'd$ and $k = k'd$ for some k', t'. We have $(a^k)^{t/d} \equiv a^{kt/d} \equiv a^{k't'd} \equiv a^{k't} \equiv (a^t)^{k'} \equiv 1 \bmod n$, which says that $\mathrm{ord}_n(a^k) \mid t/d = t'$. On the other hand, $(a^k)^{\mathrm{ord}_n(a^k)} \equiv a^{k \cdot \mathrm{ord}_n(a^k)} \equiv 1 \bmod n$, so this means that $t \mid k\mathrm{ord}_n(a^k)$. Dividing through by d, this means $t' \mid k' \cdot \mathrm{ord}_n(a^k)$, and since $\gcd(k', t') = 1$ this means that $t' \mid \mathrm{ord}_n(a^k)$. Combining the two divisibilities gives us what we want. □

Since Euler's theorem tells us that $\mathrm{ord}_n(a) \leq \varphi(n)$, a natural question is, when does equality happen?

Definition 4.2.9. Let $n > 1$ be an integer and $\gcd(a, n) = 1$. We say that a is a **generator mod** n if $\mathrm{ord}_n(a) = \varphi(n)$.

The term "generator" is because if $\mathrm{ord}_n(a) = \varphi(n)$, the powers $a, a^2, \ldots, a^{\varphi(n)}$ must all be *distinct* residues modulo n. Since there are $\varphi(n)$ units mod n, this means that each unit is congruent to some power of a modulo n. Group theoretically, this says $\langle a \bmod n \rangle = (\mathbb{Z}/n\mathbb{Z})^\times$. Since cyclic groups are significantly easier to work in, it will be rather useful to understand when generators exist. The Chinese Remainder Theorem tells us that $(\mathbb{Z}/n\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z})^\times \times \ldots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})^\times$ for $n = p_1^{e_1} \cdots p_k^{e_k}$, so we'll start by investigating when $(\mathbb{Z}/p^e\mathbb{Z})^\times$ is cyclic.

4.3 Cyclicity of $(\mathbb{Z}/p\mathbb{Z})^\times$

The first step in our investigation of when $(\mathbb{Z}/p^e\mathbb{Z})^\times$ is cyclic for prime p and $e \geq 1$ is to start with the $e = 1$ case. As it turns out, there is always a generator mod p. We'll deduce this by proving a *stronger* result, that the multiplicative group of any finite field is cyclic.

Proposition 26. *Let F be a field and $f(x) \in F[x]$ a polynomial of degree $d \geq 1$. Then $f(x)$ has at most d distinct roots in d.*

Proof. If $d = 1$, then $f(x) = ax + b$ for some $a, b \in F$ and clearly the only root of $f(x)$ in F is $-b/a$. Now suppose we know the result is true for any polynomial of degree k. Let $f(x)$ be an arbitrary polynomial of degree $k + 1$. If $f(x)$ has no root in F, then we're done. Otherwise, it has some root $\alpha \in F$. By the division algorithm, we may write $f(x) = (x - \alpha)g(x) + r(x)$ where $r(x) = 0$ or $r(x) = c$ for some $c \in F$. Plugging in α shows that $r(\alpha) = 0$, so that $r(x) = 0$ for all $x \in F$. This means that $f(x) = (x - \alpha)g(x)$, and $\deg(g(x)) = k - 1$. By induction hypothesis, $g(x)$ has at most $k - 1$ distinct roots in F, which means that $f(x)$ has at most $(k - 1) + 1 = k$ roots in F. By induction, we're done. □

Proposition 27. *Let F be a finite field of size q. Then for any $\alpha \in F^\times$, $\alpha^{q-1} = 1$.*

Proof. The proof is the same as the proof of Euler's theorem: consider the map $f : F^\times \to F^\times$ given by $f(x) = \alpha x$. This is a bijection with inverse map $f^{-1}(x) = \alpha^{-1}x$. Therefore, $\prod_{x \in F^\times} \alpha x = \prod_{x \in F^\times} x$ which says $\alpha^{q-1} = 1$ after canceling the product from both sides. □

Proposition 28. *Let F be a finite field of size q, and define $S_d = \#\{\alpha \in F^\times : ord(\alpha) = d\}$. Then if $S_d > 0$, we have $S_d = \varphi(d)$.*

Proof. Suppose that $\text{ord}(\alpha) = d$. Then $\alpha^d = 1$, so α is a root of $f(x) = x^d - 1 \in F[x]$. Since α has order d, this means the powers $1, \alpha, \alpha^2, \ldots, \alpha^{d-1}$ are all *distinct* elements of F, and note that $f(\alpha^k) = (\alpha^d)^k - 1 = 0$. Since $f(x)$ has at most d roots in F and we've found d distinct roots, this must be all of them. Which roots have order d? Using the formula for the order of a power, $\text{ord}(\alpha^k) = d$ if and only if $\gcd(d, k) = 1$, and there are precisely $\varphi(d)$ values of k in the range $1 \leq k \leq d$ with this property. Since any element of order d is a root of $f(x)$, this shows that if an element of order d exists, there are exactly $\varphi(d)$ elements of order d. □

Proposition 29. *For any integer $n \geq 1$, we have $\sum_{d|n} \varphi(d) = n$.*

Proof. The trick is to compute the sum *backwards*: $\sum_{d|n} \varphi(d) = \sum_{d|n} \varphi(n/d)$. By definition, $\varphi(n/d)$ counts the number of integers in the range $1 \leq k \leq n/d$ with $\gcd(k, n/d) = 1$. Now, saying that $\gcd(k, n/d) = 1$ is the same as saying that $\gcd(dk, n) = d$. If $1 \leq m \leq n$ and $\gcd(m, n) = d$, then $m = dk$ for some $1 \leq k \leq n/d$ *anyway*, so this means that $\varphi(n/d)$ actually counts the number of integers in the range $1 \leq m \leq n$ with $\gcd(m, n) = d$. Now, any such integer $1 \leq m \leq n$ has $\gcd(m, n) = d$ for some $d \mid n$. Therefore, the sets $A_d = \{1 \leq m \leq n : \gcd(m, n) = d\}$ for $d \mid n$ form a partition of $\{1, \ldots, n\}$. By our observation, $|A_d| = \varphi(n/d)$, so taking cardinalities yields $\sum_{d|n} \varphi(n/d) = \sum_{d|n} |A_d| = n$ as desired. □

Theorem 4.3.1. *Let F be a finite field of size q. Then $F^\times$ is a cyclic group. In particular, $(\mathbb{Z}/p\mathbb{Z})^\times$ always has a generator for p prime.*

Proof. For any $\alpha \in F^\times$ we must have $\text{ord}(\alpha) \mid q-1$, since $\alpha^{q-1} = 1$. By the previous lemma, this says that $q - 1 = \sum_{d|q-1} \varphi(d)$. Using the notation of proposition 25, $S_d \leq \varphi(d)$ for all d (there could be *no* elements of order d!). Therefore, $q-1 \leq \sum_{d|q-1} S_d \leq \sum_{d|q-1} \varphi(d) = q-1$. This forces equality, so that $q - 1 = \sum_{d|q-1} S_d$. This forces $S_d > 0$ for all $d \mid q - 1$, so that in particular, $S_{q-1} = \varphi(q - 1)$ as desired. □

It's worth pointing out that this argument really does use the fact that F is a *field* in a non-trivial way. The crux of the argument relies on the fact that over a field, the number of roots of a polynomial are bounded by the degree. Over a ring, this need not happen. For example, in $(\mathbb{Z}/8\mathbb{Z})[x]$, the polynomial $x^2 - 1$ has 4 roots, so the argument cannot generalize.

One very useful application of our main theorem is being able to detect if elements of finite fields are n-th powers or not.

Proposition 30. *Let F be a finite field of size q and let $\alpha \in F^\times$. Then the equation $x^n = \alpha$ is solvable in F if and only if $\alpha^{q-1/d} = 1$, where $d = \gcd(q - 1, n)$. When there is a solution, there are exactly d solutions.*

Proof. Let g be a generator of $F^\times$. Write $\alpha = g^\ell$ for some ℓ, and let $x = g^k$. Then solving $x^n = a$ is the same as solving $g^{nk} = g^\ell$ in F. This is equivalent to the solvability of the

congruence $nk \equiv \ell \bmod q-1$, which we know is solvable if and only if $\gcd(n, q-1) \mid \ell$, and has exactly $d = \gcd(n, q-1)$ solutions. Write $q-1 = dm$ for some m. If $d \mid \ell$, then $\alpha^{q-1/d} = g^{(q-1)\ell/d} = (g^{q-1})^{\ell/d} = 1$. Conversely, if $\alpha^{q-1/d} = 1$, this means that $g^{(q-1)\ell/d} = 1$, so $q-1 \mid (q-1)\ell/d$. This means $d(q-1) \mid (q-1)\ell$, so that $d \mid \ell$ as desired. □

Example 4.3.2. Suppose we wanted to solve $x^{44} \equiv 81 \bmod 257$. First, note that 3 is a generator mod 257. We must have $\mathrm{ord}_{257}(3) \mid 256$, so so we just need to rule out all lower powers of 2. One may compute $3^5 \equiv -14 \bmod 257$, $3^{10} \equiv 196 \bmod 257$, $3^{11} \equiv 74 \bmod 257$, $3^{22} \equiv 79 \bmod 257$, $3^{44} \equiv 73 \bmod 257$, $3^{128} \equiv \frac{176}{81} \equiv 2 + \frac{14}{81} \equiv -1 \bmod 257$. This shows that we must have $\mathrm{ord}_{257}(3) = 256$. Therefore, we want to solve $3^{44k} \equiv 3^4 \bmod 257$. This is the same as solving $44k \equiv 4 \bmod 256$, i.e. $11k \equiv 1 \bmod 64$. This means $k \equiv 35 \bmod 64$, so that $k \equiv 35, 99, 163, 227 \bmod 256$. This yields the four solutions $x \equiv 186, 108, 8, 149 \bmod 257$.

4.4 Hensel's lemma

If the Chinese Remainder Theorem is in the running for the most important theorem in number theory, then Hensel's lemma might be a contender for the *second* most important theorem. Hensel's lemma is a "lifting" theorem: it says that if you understand what's happening in $\mathbb{Z}/p\mathbb{Z}$ for prime p, then you can often times "lift" up to understand what's happening in $\mathbb{Z}/p^e\mathbb{Z}$ for $e \geq 1$.

Theorem 4.4.1 (Hensel's lemma)**.** *Let p be a prime. Suppose that $f(x) \in \mathbb{Z}[x]$ and that $f(c) \equiv 0 \bmod p$ and $f'(c) \not\equiv 0 \bmod p$ for some $c \in \mathbb{Z}$. Then for any $k \geq$, there is $c_k \in \mathbb{Z}$ such that $c_k \equiv c \bmod p$ with $f(c_k) \equiv 0 \bmod p^k$, and c_k is unique modulo p^k.*

Proof. We will explicitly construct a root of $f(x) \bmod p^k$ for all $k \geq 1$. By assumption, we may take $c_1 = c$, and certainly there is a unique root of $f(x) \bmod p$ with this property. Now, suppose that we have constructed c_k with the desired properties. We wish to construct c_{k+1}. Consider $c_k + tp^k$ for $t \in \mathbb{Z}$. We'll show that a choice of t gives the desired properties. We have $f(c_k + tp^k) = f(c_k) + tp^k f'(c_k) + \frac{1}{2}t^2p^{2k}f''(c_k) + \dots$ by Taylor expanding $f(x)$ around c_k. Working mod p^k, we see that $f(c_k + tp^k) \equiv f(c_k) + tp^k f'(c_k) \bmod p^{k+1}$, because all higher terms are divisible by a larger power of p. We wish to find t such that $0 \equiv f(c_k) + tp^k f'(c_k) \bmod p^{k+1}$. By induction hypothesis, $f(c_k) \equiv 0 \bmod p^k$, so dividing through, we want to solve $0 \equiv \frac{f(c_k)}{p^k} + tf'(c_k) \bmod p$. Since $c_k \equiv c \bmod p$ and $f'(c) \not\equiv 0 \bmod p$ by assumption, we can indeed solve for t: $t \equiv -\frac{f(c_k)}{p^k} f'(c)^{-1} \bmod p$. We then set $c_{k+1} = c_k + tp^k$, from which $f(c_k) \equiv 0 \bmod p^{k+1}$ by construction, and clearly $c_{k+1} \equiv c \bmod p$. By induction, this shows there is a root mod p^k for all $k \geq 1$.

This shows existence, so now we need to prove that there is a unique choice of lift of the root mod p^k. Once again, we proceed by induction. For $k = 1$, this is clear. Now suppose that we know there's a unique lift to a root mod p^k, and suppose we have two roots c_1, c_2 such that $f(c_1) \equiv f(c_2) \equiv 0 \bmod p^{k+1}$ that both satisfy $c_1 \equiv c_2 \equiv c \bmod p$. We need to show that $c_1 \equiv c_2 \bmod p^{k+1}$. We prove *this* by induction as well! Suppose we know that $f(c_1) \equiv f(c_2) \equiv 0 \bmod p^{k+1}$ and $c_1 \equiv c_2 \bmod p \implies c_1 \equiv c_2 \bmod p^k$ for some k. If $f(c_1) \equiv f(c_2) \equiv 0 \bmod p^{k+1}$ and $c_1 \equiv c_2 \bmod p$. Then $f(c_1) \equiv f(c_2) \equiv 0 \bmod p^k$ as

well, so by induction hypothesis, $c_1 \equiv c_2 \bmod p^k$. Then we can write $c_2 = c_1 + tp^k$ for some t. Plugging in and Taylor expanding at c_1, we have $f(c_2) \equiv f(c_1) + tp^k f'(c_1) \bmod p^{k+1}$, and since $f(c_1) \equiv f(c_2) \equiv 0 \bmod p^k$, this in fact means that $0 \equiv tf'(c_1) \bmod p$. Since $f'(c_1) \equiv f'(c) \not\equiv 0 \bmod p$, this means $t \equiv 0 \bmod p$, so that $c_1 \equiv c_2 \bmod p^{k+1}$ as desired. Therefore by induction, we're done. □

The condition that $f'(c) \not\equiv 0 \bmod p$ means that c is not a repeated root of $f(x)$ in $\mathbb{F}_p$. Therefore, what Hensel's lemma says is that *simple* roots of polynomials $\bmod p$ uniquely lift up to roots mod p^k for any $k \geq 1$.

Example 4.4.2. Let $f(x) = x^2 - 2 \in \mathbb{Z}[x]$. Note that $f(3) \equiv 0 \bmod 7$ and $f'(3) \equiv 6 \not\equiv 0 \bmod 7$. By Hensel's lemma, this means for any $k \geq 1$ we can find a unique lift of 3 mod 7 to a root of $f(x)$ mod p^k. The proof of Hensel's lemma explicitly tells us how to do this!

To find a root mod 49, our lift is of the form $c_2 = 3 + 7t$ for some t. Therefore, we need to solve $\frac{f(3)}{7} + tf'(3) \equiv 0 \bmod 7$. Since $f'(3) \equiv -1 \bmod 7$, this means we need $1 - t \equiv 0 \bmod 7$, so $t \equiv 1 \bmod 7$. Thus, 10 mod 49 is the unique lift of 3 mod 7 to a root of $f(x)$ mod p^k.

We can keep repeating this process to lift up roots further. For example, to compute the lift of the root to a root mod 343, we have $c_3 = 10 + 49t$ for some t. To compute t, we solve $\frac{f(10)}{49} + tf'(49) \equiv 0 \bmod 7$. Since $f'(49) \equiv f'(7) \equiv -1 \bmod 7$, this means we solve $2 - t \equiv 0 \bmod 7$, so that $t \equiv 2 \bmod 7$. Then 108 mod 343 is the unique lift of 3 mod 7 to a root of $f(x)$ mod 343.

By combining Hensel's lemma with the Chinese Remainder Theorem, we get a general process for trying to solve polynomial congruences $f(x) \equiv 0 \bmod n$ for $f(x) \in \mathbb{Z}[x]$.

- Write $n = p_1^{e_1} \cdots p_k^{e_k}$ as its prime factorization.
- If applicable, try to lift roots of $f(x) \bmod p_i$ to roots of $f(x) \bmod p_i^{e_i}$ using Hensel's lemma.
- Use the Chinese Remainder Theorem to glue the roots mod $p_i^{e_i}$ to a root mod n.

That this last steps works follows from the next proposition:

Proposition 31. *Suppose that $f(x) \in \mathbb{Z}[x]$. Let $m, n > 1$ with $\gcd(m, n) = 1$. If $f(r) \equiv 0 \bmod m$ and $f(s) \equiv 0 \bmod n$ then $f(c) \equiv 0 \bmod mn$, where $c \equiv r \bmod m$ and $c \equiv s \bmod n$.*

Proof. Exercise. □

Example 4.4.3. Suppose we wanted to find all solutions to the congruence $x^3 - 6 \equiv 0 \bmod 245$. We note that $245 = 5 \cdot 7^2$, so solving $x^3 \equiv 6 \bmod 245$ is equivalent to solving

$$\begin{cases} x^3 \equiv 1 \bmod 5 \\ x^3 \equiv 6 \bmod 49 \end{cases}$$

There is a unique solution to the first congruence, namely, $x \equiv 1 \bmod 5$. For the second congruence, note that there are three solutions to $x^3 \equiv 6 \bmod 7$, given by $x \equiv 3, 5, 6 \bmod 7$.

Since none of these are repeated roots of $f(x)$ mod 7, by Hensel's lemma, they all uniquely lift to roots mod 49. One may run through the lifting computation and verify that these lifts are $x \equiv 24, 34, 40 \bmod 49$. Since any root of $f(x)$ mod 49 *is* a lift of some root of $f(x)$ mod 7, these are *all* the roots of $f(x)$ mod 49. We then get three systems of congruences

$$\begin{cases} x \equiv 1 \bmod 5 \\ x \equiv 23 \bmod 49 \end{cases} \quad \begin{cases} x \equiv 1 \bmod 5 \\ x \equiv 34 \bmod 49 \end{cases} \quad \begin{cases} x \equiv 1 \bmod 5 \\ x \equiv 40 \bmod 49 \end{cases}$$

which glue together to the roots $x \equiv 171, 181, 236 \bmod 245$ by the Chinese Remainder Theorem. Since any root of $f(x)$ mod 245 corresponds to a pair of roots mod 5 and 49, these are *all* roots of $f(x)$ mod 245.

What happens when the condition $f'(c) \not\equiv 0 \bmod p$ is not met? Let $r = c + tp^k$ be a lift of $c \bmod p^k$ to a congruence class mod p^{k+1}. Then we have $f(r) \equiv f(c) + tp^k f'(c) \equiv f(c) \bmod p^{k+1}$. This means that either *all* lifts are roots, if c is a root of $f(x)$ mod p^k, or *no* lifts are roots, if c is not a root of $f(x)$ mod p^k.

Example 4.4.4. Consider $f(x) = x^2 + 1 \in \mathbb{Z}[x]$. Then $f(1) \equiv 0 \bmod 2$ and $f'(1) \equiv 0 \bmod 2$, so Hensel's lemma doesn't apply. The lifts of 1 mod 2 to a congruence class mod 4 are $1, 3 \bmod 4$, and neither of these are roots because $f(1) \equiv 2 \not\equiv 0 \bmod 4$.

Hensel's lemma can also be used to compute inverse mod p^k!

Example 4.4.5. Suppose we wanted to compute 5^{-1} mod 343. This is equivalent to finding the root of $f(x) = 5x - 1$ mod 343. Note that $f(3) \equiv 0 \bmod 7$ and $f'(3) \equiv 5 \bmod 7$, so we can lift 3 mod 7 to a root mod 343. One may verify that 10 mod 49 is the lift to a root mod 49, and that 206 mod 343 is the lift to a root mod 343. This means that $5^{-1} \equiv 206 \bmod 343$.

Since Hensel's lemma can be used to compute inverses mod p^k, one may then combine Hensel's lemma with the Chinese Remainder Theorem to compute inverses mod n for any n. However, this is, in general, going to be slower than just running the Euclidean Algorithm.

The astute reader may have noticed the following recursive formula for computing lifts of roots:

Proposition 32. *Suppose that $f(x) \in \mathbb{Z}[x]$ satisfies the conditions of Hensel's lemma, and that c_k is a root of $f(x) \bmod p^k$ for $k \geq 1$. Then $c_{k+1} \equiv c_k - f(c_k)\overline{f'(c)} \bmod p^{k+1}$ is the unique lift of $c_k \bmod p^k$ to a root of $f(x) \bmod p^{k+1}$, where $\overline{f'(c)} \equiv f'(c)^{-1} \bmod p$.*

This recursion looks suspiciously similar to *Newton's method* from calculus, which says that the successive iterates $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ provide good approximations to a root of $f(x)$. This is not coincidental!

For each prime p and $k \geq 1$, there is a natural ring homomorphism $\pi_{k+1} : \mathbb{Z}/p^{k+1}\mathbb{Z} \to \mathbb{Z}/p^k\mathbb{Z}$ given by $x \bmod p^{k+1} \mapsto x \bmod p^k$. This produces an infinite sequence of maps

$$\ldots \to \mathbb{Z}/p^k\mathbb{Z} \xrightarrow{\pi_k} \mathbb{Z}/p^{k-1}\mathbb{Z} \to \ldots \to \mathbb{Z}/p^2\mathbb{Z} \xrightarrow{\pi_2} \mathbb{Z}/p\mathbb{Z}$$

There is a formal algebraic construction called the *inverse limit*. One may form the inverse limit $\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^k\mathbb{Z}$ with respect to these projection maps, which is called the ring of *p-adic integers*.

Algebraically, one may think of $\mathbb{Z}_p$ as an "infinite" version of $\mathbb{Z}/p^k\mathbb{Z}$. By the construction of the inverse limit, there are projection maps $\mathbb{Z}_p \to \mathbb{Z}/p^k\mathbb{Z}$ for any $k \geq 1$. Elements of $\mathbb{Z}_p$ look like "power series in p": for $\alpha \in \mathbb{Z}_p$ one can write $\alpha = c_0 + c_1 p + c_2 p^2 + \ldots$ where $c_i \in \{0, \ldots, p-1\}$ (one has to make sense of what it means for this infinite sum to converge in $\mathbb{Z}_p$, but we won't get into this).

Hensel's lemma is then a p-adic version of Newton's method! At each step, it produces a root r_k of $f(x) \mod p^k$, which looks like something of the form $r_k = r_0 + r_1 p + \ldots + r_{k-1}p^{k-1}$, coming from lifting an initial root $r_0 \bmod p$. Taking a limit as $k \to \infty$ will produce a root $\alpha \in \mathbb{Z}_p$ of $f(x)$. The k-th iterate r_k is a p-adic approximation of α that agrees mod p^k.

4.5 Cyclicity of $(\mathbb{Z}/n\mathbb{Z})^\times$

We've seen that $(\mathbb{Z}/p\mathbb{Z})^\times$ for p prime is a cyclic group. Using the lifting philosophy of Hensel, we're going to show that a generator mod p can be lifted to a generator mod p^k for all $k \geq 1$, so that $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is cyclic. Once we've proven this, we'll be ready to tackle the general case of when $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic.

Proposition 33. *Suppose that $\mathrm{ord}_n(a) = m$ and $\mathrm{ord}_n(b) = \ell$. If $\gcd(m, \ell) = 1$ then $\mathrm{ord}_n(ab) = m\ell$.*

Proof. Set $t = \mathrm{ord}_n(ab)$. We see that $(ab)^{m\ell} = (a^m)^\ell (b^\ell)^m \equiv 1 \bmod n$, so that $t \mid m\ell$. On the other hand, we have $(ab)^t \equiv 1 \bmod n$, so raising to the m-th power yields $b^{tm} \equiv 1 \bmod n$. Similarly, $a^{t\ell} \equiv 1 \bmod n$. This means that $\ell \mid tm$ and since $\gcd(\ell, m) = 1$ this means that $\ell \mid t$. Similarly, $m \mid t$, and once again because $\gcd(\ell, m) = 1$ we conclude that $\ell m \mid t$, so that $t = m\ell$. □

The proof really requires that $\gcd(m, \ell) = 1$ (it was used *twice*!). Without this condition, it's not even true that $\mathrm{ord}_n(ab) = \mathrm{lcm}(\mathrm{ord}_n(a), \mathrm{ord}_n(b))$ like one might expect!

Proposition 34. *There is a generator mod p^2 for odd prime p.*

Proof. Let g be a generator mod p. Then g is a root mod p of $f(x) = x^{p-1} - 1 \in \mathbb{Z}[x]$. Since $f'(g) \equiv -g^{p-2} \not\equiv 0 \bmod p$, by Hensel's lemma, there is a *unique* lift $g + cp$ of g to a root of $f(x) \bmod p^2$. Note that $\varphi(p^2) = p(p-1)$, so this means that $\mathrm{ord}_{p^2}(g + cp) = p - 1$ or $p(p-1)$ since by virtue of being a root of $x^{p-1} - 1$, the order has to be divisible by $p - 1$. We claim that it cannot be $p(p-1)$. For any d with $(g + cp)^d \equiv 1 \bmod p^2$, we must have $g^d \equiv 1 \bmod p$. In particular, this means any *other* lift $g + c'p$ *cannot* be a root of $x^{p-1} - 1 \bmod p^2$, because there's a *unique* lift of g to a root mod p^2. This means that all *other* lifts must have order $p(p-1)$, so we're done. □

In particular, this means that for each generator g mod p, there are $p-1$ lifts of g that are a generator mod p^2. Since there are $\varphi(p-1)$ generators mod p, there are $(p-1)\varphi(p-1) = \varphi(\varphi(p^2))$ generators mod p^2 (note that a generator mod p^2 must also be a generator mod p!).

Proposition 35. *If g is a generator mod p^2 for odd prime p then g is a generator mod p^k for all $k \geq 2$.*

Proof. We already did the case $k = 2$. Suppose that g is a generator mod p^k, and set $t = \text{ord}_{p^{k+1}}(g)$. Since $\varphi(p^{k+1}) = p^k(p-1)$ and both $g^t \equiv 1 \bmod p^{k+1}$ and $g^{\varphi(p^{k+1})} \equiv 1 \bmod p^{k+1}$, this means $t = p^k(p-1)$ or $t = p^{k-1}(p-1)$. Since $g^{p^k(p-1)} \equiv 1 \bmod p^{k+1}$, this means that $g^{p-1} \equiv 1 \bmod p$, so we may write $g^{p-1} = 1 + p\ell$ for some ℓ with $p \nmid \ell$ (because g is a generator mod p^2, and so $g^{p-1} \not\equiv 1 \bmod p^2$). Then $g^{p(p-1)} = (g^{p-1})^p = (1+p\ell)^p \equiv 1 + p\ell^2 \bmod p^3$ because the binomial coefficients $\binom{p}{k}$ are divisible by p for $k = 1, \ldots, p-1$. Therefore, we may write $g^{p(p-1)} = 1 + \ell_2 p^2$ with $p \nmid \ell_2$. Taking a p-th power once more, we find $g^{p^2(p-1)} \equiv 1 + \ell_2 p^3 \bmod p^4$. Repeating this process, we eventually find that $g^{p^{k-1}(p-1)} \equiv 1 + \ell_{k-1}p^k \bmod p^{k+1}$ with $p \nmid \ell_{k-1}$. This means that $\text{ord}_{p^{k+1}}(g) = p^k(p-1)$ must be the only possibility, so that g is a generator mod p^{k+1}. By induction, we're done. □

Note that the proof *explicitly* tells us what a generator is mod p^k for all $k \geq 2$. Since $g^p \equiv g \bmod p$, $g^p \bmod p^2$ is a lift of g mod p. Then note that $(g^p)^{p-1} \equiv 1 \bmod p^2$ by Euler's theorem, so g^p is the "bad" lift of g that cannot be a generator mod p^2. This means that any *other* lift of g mod p that's not congruent to g^p mod p^2 *is* a generator mod p^2! In particular, note that $g^{p-1}(g+p) \equiv g \bmod p$, and $g^{p-1}(g+p) \not\equiv g^p \bmod p^2$, so that $g^{p-1}(g+p)$ is always a generator mod p^2.

Example 4.5.1. 3 is a generator mod 7, so $3^6 \cdot 10 \equiv 38 \bmod 49$ is a generator mod 7^k for all $k \geq 2$.

We now have to handle the case of $p = 2$, for which Hensel's lemma obviously does not apply, because $x^{p-1} - 1 = x - 1$ always has a single root!

Proposition 36. *There is a generator mod 2^k if and only if $k = 1, 2$.*

Proof. That there is a generator mod 2 and a generator mod 4 is obvious: in both cases, 1 mod 2 and 3 mod 4 are generators. Therefore, suppose that $k \geq 3$ and g is a generator mod 2^k. Then the set of powers $\{1, g, \ldots, g^{2^{k-1}}\}$ must hit all 2^{k-1} units mod 2^k. Therefore, there is ℓ such that $g^\ell \equiv -1 \bmod 2^k$. Squaring, $g^{2\ell} \equiv 1 \bmod 2^k$ so this means $2^{k-1} \mid 2\ell$ so that $2^{k-2} \mid \ell$, which forces $2^{k-2} = \ell$ since $1 \leq \ell \leq 2^k$. On the other hand, note that $g^2 \equiv 1 \bmod 8$ since all units square to 1 mod 8. This means that $g^4 \equiv 1 \bmod 16$ using the Binomial theorem, and so repeatedly squaring, we arrive at $g^{2^{k-2}} \equiv 1 \bmod 2^k$. However, this means that $1 \equiv -1 \bmod 2^k$, which is a contradiction. Therefore, there is no generators mod 2^k for $k \geq 3$. □

We need one more property of orders before we're ready to tackle our goal.

Proposition 37. *Suppose that* $\gcd(m, n) = 1$. *Then* $ord_{mn}(a) = lcm(ord_m(a), ord_n(a))$.

Proof. Let $t = \text{ord}_{mn}(a)$ and $k = \text{lcm}(\text{ord}_m(a), \text{ord}_n(a))$. Then $a^t \equiv 1 \bmod mn$, so $a^t \equiv 1 \bmod m$ and $a^t \equiv 1 \bmod n$, so this means that $\text{ord}_m(a) \mid t$ and $\text{ord}_n(a) \mid t$ so $k \mid t$. On the other hand, $a^k \equiv 1 \bmod m$ and $a^k \equiv 1 \bmod n$, so in particular since m, n are relatively prime this means that $a^k \equiv 1 \bmod mn$, so that $t \mid k$. This proves $t = k$ as we wanted. □

Theorem 4.5.2. *There is a generator mod n if and only if $n = 2, 4, p^k, 2p^k$ for p an odd prime and $k \geq 1$.*

Proof. We've already shown the result for $n = 2, 4, p^k$, so it remains to see that there is a generator mod $2p^k$. To see this, suppose that g is a generator mod p^k. If g is odd, then g has order 1 mod 2, so applying the previous proposition says that $\text{ord}_{2p^k}(g) = \varphi(p^k) = \varphi(2p^k)$, so that g is a generator. If g is even, then $g + p^k$ is odd, and $\text{ord}_{p^k}(g + p^k) = \text{ord}_{p^k}(g)$, so $g + p^k$ is a generator mod $2p^k$.

Conversely, write $n = 2^e p_1^{e_1} \cdots p_k^{e_k}$ as a product of primes where p_i are odd. Then $\text{ord}_n(a) = \text{lcm}(\text{ord}_{2^e}(a), \ldots, \text{ord}_{p_k^{e_k}}(a))$. If $e \geq 3$, then $\text{ord}_{2^e}(a) \leq 2^{e-2}$ which means that $\text{ord}_n(a) < 2^{e-2}\varphi(p_1^{e_1} \cdots p_k^{e_k}) < \varphi(n)$. If $k \geq 2$, then $\varphi(p_1^{e_1})$ and $\varphi(p_k^{e_k})$ are both even, and so this means that $\text{lcm}(\text{ord}_{p_1^{e_1}}(a), \text{ord}_{p_2^{e_2}}(a)) < \varphi(p_1^{e_1} p_2^{e_2})$, which one again results in $\text{ord}_n(a) < \varphi(n)$. So it remains to rule out $n = 4p^k$, but the same argument works because both $\varphi(4)$ and $\varphi(p^k)$ are even. This leaves the only possibilities as $n = 2, 4, p^k, 2p^k$. □

Earlier, we gave a criteria for when an element in a finite field is an n-th power. Since we now know when there is a generator mod n, we can give a similar criteria for such values.

Proposition 38. *Suppose that there is a generator mod m, and $\gcd(a, m) = 1$. Then $x^n \equiv a \bmod m$ is solvable if and only if $a^{\varphi(m)/d} \equiv 1 \bmod m$ with $d \gcd(n, \varphi(m))$.*

The proof is identical to the finite field case, and so we omit it. A particular test of interest will be the following special case:

Corollary 4.5.3. *Let p be a prime and $p \nmid a$. Then $x^2 \equiv a \bmod p$ is solvable if and only if $a^{\frac{p-1}{2}} \equiv 1 \bmod p$.*

4.6 Application: Cryptography

As an application of all this theory, we'll give a short discussion about cryptography.

Definition 4.6.1. Let g be a generator mod n. For any a with $\gcd(a, n) = 1$, the **base g logarithm of** a, $\log_g(a)$, is the congruence class mod $\varphi(n)$ that solves $g^x \equiv a \bmod n$.

This is, of course, similar to how logarithms in $\mathbb{R}$ are defined: $\log_b(a)$ is the real number such that $b^x = a$ in $\mathbb{R}$. Here the same thing happens, except because exponents are only unique modulo $\text{ord}_n(g) = \varphi(n)$, the logarithm is a *congruence class*, and not an integer. We often refer to $\log_g(a)$ as a "discrete logarithm"

Example 4.6.2. 3 is a generator mod 7, and $3^4 \equiv 4 \bmod 7$, so $\log_3(4) \equiv 4 \bmod 6$.

Discrete logarithms have all the properties that you would expect.

Proposition 39. *Let g be a generator mod n. Then for any integers a, b, k we have:*

1. $\log_g(1) \equiv 0 \bmod \varphi(n)$.
2. $\log_g(ab) \equiv \log_g(a) + \log_g(b) \bmod \varphi(n)$.
3. $\log_g(a^k) \equiv k \log_g(a) \bmod \varphi(n)$.

Proof. These all follow rather easily. The first is obvious, so we just prove the second. The third then follows similarly. By definition, $g^{\log_g(ab)} \equiv ab \bmod n$, and $a \equiv g^{\log_g(a)} \bmod n$ and $b \equiv g^{\log_g(b)} \bmod n$, so this means $g^{\log_g(ab)} \equiv g^{\log_g(a)+\log_g(b)} \bmod n$ which then immediately tells us that $\log_g(ab) \equiv \log_g(a) + \log_g(b) \bmod \varphi(n)$. □

Just as logarithms are useful for solving equations, so are discrete logarithms.

Example 4.6.3. Suppose we wanted to solve the congruence $6x^{12} \equiv 11 \bmod 17$. One may check that 3 is a generator mod 17, so taking logarithms of both sides, $\log_3(6) + 12\log_3(x) \equiv \log_3(11) \bmod 16$. Since $\log_3(6) \equiv 1 + \log_3(2) \bmod 16$, this means $1 + 12\log_3(x) \equiv \log_3(\frac{11}{2}) \equiv \log_3(14) \bmod 16$, because $\frac{11}{2} \equiv 14 \bmod 17$. Note that $14 \equiv -3 \bmod 17$, so $\log_3(14) \equiv \log_3(-3) \equiv \log_3(-1) + 1 \equiv 9 \bmod 16$, as $\log_3(-1) = 8$ because 3 is a generator mod 17. Therefore, $12\log_3(x) \equiv 8 \bmod 16$, and so this tells us that $3\log_3(x) \equiv 2 \bmod 4$. Solving says $\log_3(x) \equiv 2, 6, 10, 14 \bmod 16$, so that $x \equiv 9, 15, 8, 2 \bmod 17$.

Classical algorithms in cryptography are based on "hard" problems in number theory. In particular, we'll discuss the following two algorithms, each based on the following ideas:

- RSA: Taking k-th roots mod$n$ is hard!
- ElGamal: Computing $\log_g(a)$ is hard!

It's worth pointing out that discrete logarithms make sense in any cyclic group, since the definition only depends on the existence of a generator. Classically, ElGamal was done in $(\mathbb{Z}/p\mathbb{Z})^\times$, although other common choices are are $F^\times$ for F a finite field of size q, or the group of rational points on certain elliptic curves. It's more difficult to do computations in these settings, so it increases the security of the crypto system.

We'll start with RSA. We have two participants, Alice and Bob, who wish to secretly communicate with each other over a public channel. Each participant has a **public key**, and a **private key**. The data of each of these is as follows:

- The public key is a pair (e, m) of integers such that $\gcd(e, \varphi(m)) = 1$.
- The private key is an integer d such that $ed \equiv 1 \bmod \varphi(m)$.

The idea behind RSA is quite straightforward. Let the public keys of Alice and Bob be (e_A, m_A) and (e_B, m_B) respectively, with private keys d_A and d_B. Alice takes the message she wants to send, and converts it via some shared scheme between Alice and Bob into an integer M (by, say, associating letters to numbers via a cipher). Alice takes her message M, and encrypts it by computing $M^{e_B} \bmod m_B$. Bob can then decrypt the message by taking

a d_B-th power, as $(M^{e_B})^{d_B} \equiv M^{e_B d_B} \equiv M \bmod m_B$ because $e_B d_B \equiv 1 \bmod \varphi(m_B)$. Where does the security come from? Suppose that Eve the spy wants to listen in on their conversation. Eve always knows (e_A, m_A) and (e_B, m_B). If she intercepts the encrypted message E, she needs to either solve $e_A x \equiv 1 \bmod \varphi(m_A)$ or $e_B x \equiv 1 \bmod \varphi(m_B)$, depending on if the message is getting sent to Alice or Bob. The point is that doing so *generally* requires *computing* $\varphi(m_i)$, and this is not really feasible without knowing the factorization of the modulus!

Traditionally, one takes $m = pq$ for distinct primes p, q. Then $\varphi(m) = (p-1)(q-1)$. Note that $x^2 - (m+1-\varphi(m))x + m = (x-p)(x-q)$, so if you know both $m, \varphi(m))$ it's equivalent to knowing p, q, and therefore how m factors. The security of RSA then, relies of the fact that there's no efficient algorithm for factoring integers. With modern computational power, this is not yet possible. However with *quantum computers*, there *is* an efficient factoring algorithm known as *Schor's algorithm*. The world is still quite a ways away from RSA no longer being actually secure, but the existence of this algorithm might make it less attractive a choice of cryptosystem. We give a numerical example to illustrate how RSA works.

Example 4.6.4. Suppose that you are Alice, and Bob's public key is $(3, 2823907)$. You wish to send the message "MOM" to Bob. Let $A = 01, \ldots, Z = 26$ and let a space correspond to 27. Using this rule, the message "MOM" becomes the integer 131513. We encrypt the message by computing $131513^3 \equiv 1842379 \bmod 2823907$, and we send Bob the integer 1842379. Bob's private key is 1880251 (verify this using WolfamAlpha!). Bob can then decrypt the received message by computing $1842379^{1880251} \equiv 131513 \bmod 2823907$, and translating back into text using the predetermined association of letters and numbers.

Some remarks about RSA:

- If the message is long, it needs to be broken up into chunks before it can be sent over. For example, if the modulus in Bob's public key is 7 digits long, then using the scheme in the above example, Alice can only send over messages consisting of at most three letters at a time. This is because any message of at least four letters will correspond to an 8 digit integer, which would be reduced, losing information.

- If the modulus was, say, 12964553, then the message can have at most 4 letters, as long as the starting letter is not after "K". For example, any four letter message starting with "M" would be able to be sent, because we would get an 8 digit number that gets reduced mod 12964553.

Example 4.6.5. Assume that Bob has the same public key as in the previous example. If Bob receives the block message $(1294545, 1214153)$ from Alice, this decodes to $11209, 2205$. Since any message needs to have an *even* number of digits, we need to insert a leading 0 into the first block, leaving Bob with $011209, 2205$. This decodes to the message "ALIVE".

Next, we take a look at the ElGamal cryptosystem. Once again, we assume a similar set up: Alice and Bob wish to secretly communicate over a public channel. Once again, the system consists of a **public key** and a **private key**. The data is as follows:

- The public key is a triple (p, g, h) such that p is a prime, g is a generator mod p, and $h \equiv g^a \bmod p$.

- The private key is any integer a.

Suppose that the public keys of Alice and Bob are (p, g, h) and (p, g, h') respectively. Alice and Bob share a secret! Alice can compute $(h')^a \equiv (g^b)^a \equiv g^{ab} \bmod p$ and Bob can compute $h^b \equiv (g^a)^b \equiv g^{ab} \bmod p$. We set $s \equiv g^{ab} \bmod p$. As before, Alice converts her message into an integer M, and she can encrypt it by sending $Ms \bmod p$ over to Bob. Since Bob knows $s \bmod p$, Bob can compute its inverse, and decrypt the message by computing $(Ms) \cdot s^{-1} \bmod p \equiv M \bmod p$. Why is this secure? If Eve the spy tries to listen in to their conversation, Eve will know g, g^a, g^b. In order to decrypt the message that's getting send, she'd need to know how to compute g^{ab} using only these quantities. Currently, there is no efficient algorithm for doing so!

4.7 Application: Decimal expansions

In elementary school, you learn that a rational number x between 0 and 1 has either a finite decimal expansion $x = .c_1c_2 \dots c_d$, or is eventually *periodic*, that is $x = .b_1b_2 \dots b_m\overline{c_1c_2 \dots c_d}$. The goal of this handout will be to explain this phenomenon, and determine algorithms for determining decimal expansions. Surprisingly, the key to this will be Euler's theorem.

We'll start by showing that rational numbers are precisely those with eventually periodic decimal expansions.

Theorem 4.7.1. *Let x be a real number with $0 < x < 1$. Then x is rational if and only if the decimal expansion of x is eventually periodic.*

Proof. First, suppose that x is rational. Write $x = \frac{a}{b}$ with $(a, b) = 1$, and suppose the decimal expansion of x is $x = .c_1c_2c_3 \dots$. Then $10^k x = c_1c_2 \dots c_k.c_{k+1} \dots$. By the division algorithm, write $10^k a = bq_k + r_k$ where $0 \leq r_k < b$. Thus, $c_1c_2 \dots c_k.c_{k+1} = 10^k x = \frac{10^k a}{b} = q_k + \frac{r_k}{b}$, so that $q_k = c_1 \dots c_k$ and $\frac{r_k}{b} = .c_{k+1}c_{k+2} \dots$. Since there are only finitely many possible values for r_k, there exist some integers m, n with $m < n$ such that $r_m = r_n$, so that $.c_{m+1}c_{m+2} \dots = .c_{n+1}c_{n+2} \dots$. This says the decimal expansion of x is $.c_1c_2 \dots \overline{c_{m+1} \dots c_n}$.

Next, assume that $x = .b_1b_2 \dots b_m\overline{c_1 \dots c_d}$. Then $10^m x = b_1 \dots b_m.\overline{c_1 \dots c_d}$. Therefore if we can show $\overline{c_1 \dots c_d}$ is rational, we're done, as $10^m x$ will then be an integer plus a rational number, and therefore solving for x says x is rational. Set $y = .\overline{c_1 \dots c_d}$. Then $10^d y = c_1 \dots c_d.\overline{c_1 \dots c_d}$, so $(10^d - 1)y = c_1 \dots c_d$ says $y = \frac{c_1 \dots c_d}{10^d - 1}$, so that y is rational as desired. □

Saying the decimal expansion is finite means that the repeating part is a block of 0's, so we really have proved what we wanted: (the repeating block is either all 0's or it isn't). Note that the above proof is *constructive*: given a rational number, it gives us an algorithm for computing it's decimal expansion, and given a decimal expansion, it tells us what rational number it comes from.

Example 4.7.2. Let $x = .11\overline{123}$. Then $100x = 11.\overline{123}$, so we need to compute $y = .\overline{123}$. We have $1000y = 123.\overline{123}$, so $999y = 123$ says $y = \frac{123}{999}$. Therefore, $100x = 11 + \frac{123}{999} = \frac{11112}{999}$, so $x = \frac{11112}{99900} = \frac{926}{8325}$.

Example 4.7.3. Let $x = \frac{1}{303}$. To compute the decimal expansion of x, we follow the proof:

$$\begin{aligned}
10^1 \cdot 1 &= 303 \cdot 0 + 10 \\
10^2 \cdot 1 &= 303 \cdot 0 + 100 \\
10^3 \cdot 1 &= 303 \cdot 3 + 91 \\
10^4 \cdot 1 &= 303 \cdot 33 + 1 \\
10^5 \cdot 1 &= 303 \cdot 330 + 10
\end{aligned}$$

We have found two integers m and n with $r_m = r_n$, namely, $m = 1$ and $n = 5$. This says $x = .c_1\overline{c_2c_3c_4c_5}$. We can now read off the digits by looking at the remainders:

$$\begin{aligned}
c_1 = q_1 &= 0 \\
10c_1 + c_2 = q_2 &= 0 \implies c_2 = 0 \\
100c_1 + 10c_2 + c_3 = q_3 &= 3 \implies c_3 = 3 \\
1000c_1 + 100c_2 + 10c_3 + c_4 = q_4 &= 33 \implies c_4 = 3 \\
10000c_1 + 1000c_2 + 100c_3 + 10c_4 + c_5 = q_5 &= 330 \implies c_5 = 0
\end{aligned}$$

Therefore, $x = .0\overline{0330} = .\overline{0033}$.

This algorithm is not terribly useful: the output for $\frac{1}{303}$ had an initial non-repeating block, however we could actually write it as a purely repeating decimal!

Our next goal will be to determine whether a rational number $x = \frac{a}{b}$ has a finite decimal expansion, or an eventually periodic decimal expansion (with non-zero repeating block), and to determine a better algorithm for computing the decimal expansion. The key step in our proof was that there are finitely many remainders upon division by b, so that $r_m = r_n$ for some integers $m < n$. Translated into a statement about modular arithmetic, there are integers m, n such that $10^m \equiv 10^n \bmod b$. If 10 is invertible mod b, this is the same thing as saying $10^{n-m} \equiv 1 \bmod b$, so that there is a solution to $10^d \equiv 1 \bmod b$.

We'll first tackle the case where x has a *purely periodic* decimal expansion, i.e. $x = .\overline{c_1 \dots c_d}$.

Theorem 4.7.4. *Let $x = \frac{a}{b}$ with $(a, b) = 1$ be a rational number between 0 and 1. Then the decimal expansion of x is purely periodic if and only if $(10, b) = 1$. In particular, the period length is given by $\mathrm{ord}_b(10)$.*

Proof. First, suppose that $x = .\overline{c_1 \dots c_d}$ is purely periodic. Then $10^d x = c_1c_2 \dots c_d.\overline{c_1 \dots c_d}$, so $x = \frac{c_1 \dots c_d}{10^d - 1}$. Since $10^d - 1 \equiv 1 \bmod 10$, the denominator remains co-prime to 10 even after canceling common factors with the numerator.

Now suppose that $(10, b) = 1$, and let $d = \mathrm{ord}_b(10)$. Then $10^d \equiv 1 \bmod b$, so $10^d - 1$ is a multiple of b. Write $10^d - 1 = bk$ for some k, so that $x = \frac{a}{b} = \frac{ak}{bk} = \frac{ak}{10^d - 1}$. Since $\frac{a}{b} < 1$ we have $ak < 10^d - 1$, so the decimal expansion of ak requires at most d digits. Write

$ak = c_1 \dots c_d$, so $x = \frac{c_1 \dots c_d}{10^d - 1} = .\overline{c_1 \dots c_d}$.

Finally, suppose that x can be written as a repeating decimal with block length ℓ. The above argument shows that x can be written as a fraction with denominator $10^\ell - 1$, so that $10^\ell \equiv 1 \bmod b$. This means $d = \mathrm{ord}_b(10)$ satisfies $d \mid \ell$, so $d \leq \ell$. Since we've shown that x *can* be written as a repeating decimal with block length d, this shows it is the minimal such length, and therefore the period. □

Now, we'll tackle when x has a finite decimal expansion.

Theorem 4.7.5. *Let $x = \frac{a}{b}$ with $(a, b) = 1$ be a rational number between 0 and 1. Then the decimal expansion of x is finite if and only if the only possible prime factors of b are 2 and 5.*

Proof. First, suppose that $b = 2^e 5^f$ for some e, f. Let $d = \max\{e, f\}$. Then $10^d x = ka$ for some integer k, so $x = \frac{ka}{10^d}$ says the decimal expansion of x is an integer with some number of zeros before it, i.e. is finite.

Next, suppose that x has a finite decimal expansion, $x = .c_1 c_2 \dots c_d$. Then $10^d x = c_1 \dots c_d$, so $x = \frac{c_1 \dots c_d}{10^d}$. Canceling common factors from the numerator and denominator to reduce to common form, this says the only prime factors of the b must divide 10^d, i.e. must be 2 or 5. □

If we combine the two statements, we get the following theorem:

Theorem 4.7.6. *Let $x = \frac{a}{b}$ where $(a, b) = 1$ be a rational number between 0 and 1. Depending on the prime factorization of b, exactly one of the following holds:*

(1) x has a finite decimal expansion, if and only if $b = 2^e 5^f$ for some e, f not both 0.

(2) x is purely periodic with period length $ord_b(10)$ if and only if $(10, b) = 1$.

(3) x is eventually periodic with an initial non-repeating block if and only if $(10, b) \neq 1$ and b is divisible by a prime other than 2 or 5. If $b = 2^e 5^f b'$, then the period length is $ord_{b'}(10)$.

Proof. Everything is immediate from what we have done so far, except the claim about the period length in statement (3). Let $k = \max\{e, f\}$, then $10^k x = \frac{ma}{b'}$ for some integer m. Writing $ma = b'q + r$, with $0 \leq r < b'$, we have $\frac{ma}{b'} = q + \frac{r}{b'}$. Since $(10, b') = 1$, this says $\frac{r}{b'}$ is periodic of length $\mathrm{ord}_{b'}(10)$, so $10^k x$ has a purely periodic fractional part of period $\mathrm{ord}_{b'}(10)$. Dividing by 10^k shifts the decimal point left by k places, so that x has an initial non-repeating block (the digits of q) followed by a repeating block. □

By using modular arithmetic, we can improve the algorithm from our earlier example.

Example 4.7.7. Let $x = \frac{1}{303}$. Since $(10, 303) = 1$, the above theorem says x is purely periodic with period length $d = \mathrm{ord}_{303}(10)$. By Euler's theorem, $\varphi(303) = \varphi(3)\varphi(101) = 200$, so $d \mid 200$. One can check manually that $10^4 \equiv 1 \bmod 303$, so $d = 4$. We have $10^4 - 1 = 303 \cdot 33$, so $\frac{1}{303} = \frac{33}{10^4 - 1} = .\overline{0033}$.

Example 4.7.8. Let $x = \frac{1}{200}$. Then $200 = 2^3 \cdot 5^2$, so x has a finite decimal expansion. We have $\max\{3, 2\} = 3$, so $1000x = 5$. Shifting the decimal to the left 3 places says $x = .005$.

Example 4.7.9. Let $x = \frac{926}{8325}$. We have $8325 = 3^2 \cdot 5^2 \cdot 37$, so x has an initial non-repeating block following by a repeating block. To compute the decimal expansion of x, we'll use a combination of the previous two methods. Multiply x by 100, so that $100x = \frac{926 \cdot 4}{9 \cdot 37} = \frac{3704}{333} = 11 + \frac{41}{333}$. Since $(10, 333) = 1$, $\frac{41}{333}$ is purely periodic. We find $\varphi(333) = 216$ so $d = \mathrm{ord}_{333}(10) \mid 216$. One can check that $10^3 \equiv 1 \bmod 333$, so $d = 3$. We have $10^3 - 1 = 333 \cdot 3$, so $\frac{1}{333} = \frac{3}{10^3 - 1}$. This says $\frac{41}{333} = \frac{123}{10^3 - 1}$, so $\frac{41}{333} = .\overline{123}$. Therefore, $100x = 11.\overline{123}$, so shifting the decimal two places left gives $x = .11\overline{123}$.

Example 4.7.10. As a final example, the fractions $\frac{1}{7}, \frac{2}{7}, \ldots, \frac{6}{7}$ are all purely periodic because $(10, 7) = 1$. One can check that $\mathrm{ord}_7(10) = 6$, so that each of these fractions are periodic of length 6. As you're probably aware, these fractions are *cyclic* shifts of each other:

$$\begin{array}{lll} \frac{1}{7} = .\overline{142857} & \frac{2}{7} = .\overline{285714} & \frac{3}{7} = .\overline{428571} \\ \frac{4}{7} = .\overline{571428} & \frac{5}{7} = .\overline{714285} & \frac{6}{7} = .\overline{857142} \end{array}$$

Why does this happen? Since $\mathrm{ord}_7(10) = 6$, the powers $10^k \bmod 7$ for $1 \leq k \leq 6$ must all be distinct. Since there are 6 non-zero elements mod 7, we actually hit all of them: for any $1 \leq m \leq 6$, there is k such that $m \equiv 10^k \bmod 7$. This says that $\frac{m}{7}$ and $\frac{10^k}{7}$ have the same fractional part, but the latter we can compute by just shifting the decimal point! To illustrate this, suppose we wanted to compute $\frac{5}{7}$. One can check that $10^5 \equiv 5 \bmod 7$, so $\frac{5}{7}$ and $\frac{10^5}{7}$ have the same fractional part. From $\frac{1}{7} = .\overline{142857}$, we find $\frac{10^5}{7} = 14285.\overline{714285}$ (we know this repeats because the period length is *independent* of the numerator, so the first 6 digits must be the repeating block), which gives $\frac{5}{7} = .\overline{714285}$.

Integers for which $\frac{1}{n}$ have the cyclic shifting property listed above are quite rare: it turns out, these are precisely the integers n such that $\mathrm{ord}_n(10) = \varphi(n)$. The fractions with $n \leq 100$ for which this holds are $\frac{1}{7}, \frac{1}{17}, \frac{1}{19}, \frac{1}{23}, \frac{1}{29}, \frac{1}{47}, \frac{1}{49}, \frac{1}{59}, \frac{1}{61}, \frac{1}{97}$. In general, finding such an n where this condition holds is very hard!

4.8 Exercises

1. Solve the following systems of congruences.

   (a)
   $$\begin{cases} x \equiv 3 \bmod 5 \\ x \equiv 1 \bmod 7 \\ x \equiv 4 \bmod 9 \end{cases}$$

(b)

$$\begin{cases} x \equiv 5 \bmod 6 \\ x \equiv 14 \bmod 15 \\ x \equiv 24 \bmod 35 \end{cases}$$

(Note: since the moduli are not co-prime, you need to split each single congruence into a system of congruences, first!)

(c) Show that the system below does *not* have a solution.

$$\begin{cases} x \equiv 5 \bmod 6 \\ x \equiv 8 \bmod 15 \\ x \equiv 6 \bmod 35 \end{cases}$$

2. Solve the system of congruences

$$\begin{cases} x^2 \equiv 2 \bmod 7 \\ x^2 \equiv 3 \bmod 11 \\ x^2 \equiv 4 \bmod 13 \end{cases}$$

3. Compute $50^{51^{52}} \bmod 101$.

4. Solve $x^{11} \equiv 7 \bmod 18$.

5. Prove that the map $f : \mathbb{F}_{29}^{\times} \to \mathbb{F}_{29}^{\times}$ given by $x \bmod 29 \mapsto x^{17} \bmod 29$ is a bijection.

6. Find all solutions to $\varphi(n) = 16$.

7. Prove there is no solution to $\varphi(n) = 14$.

8. Find all solutions to $x^5 - 2x - 42 \equiv 0 \bmod 1323$ (Note: $1323 = 3^3 \cdot 7^2$).

9. Create a table of the values of $\log_5(a)$ for all units mod 23. Use your table and properties of the discrete logarithm to solve the equations $3^x \equiv 2 \bmod 23$ and $3x^{14} \equiv 2 \bmod 23$.

10. Find all generators of $F^{\times}$ for the field $F = \mathbb{F}_3[x]/(x^2 + 1)$.

11. Suppose you know that $a^{40} \equiv 1 \bmod n$, $a^{72} \equiv 1 \bmod n$, and $a^4 \not\equiv 1 \bmod n$. What is $\mathrm{ord}_n(a^2)$?

12. Find all primes p such that $\mathrm{ord}_p(2) = 10$.

13. Prove for integers $m, n > 1$ that $\varphi(mn) = \varphi(m)\varphi(n)\frac{d}{\varphi(d)}$ where $d = \gcd(m, n)$.

14. Let p, q be distinct odd primes such that $p - 1 \mid q - 1$. If $\gcd(a, pq) = 1$, prove that $a^{q-1} \equiv 1 \bmod pq$.

15. Prove that $n \mid \varphi(2^n - 1)$ for all $n \geq 1$.

16. *Mersenne primes* are primes of the form $2^p - 1$ for prime p, and *Fermat primes* are primes of the form $2^{2^n} + 1$.

    (a) Let p be an odd prime, and let q be a prime divisor of $2^p - 1$. Prove that $\text{ord}_q(2) = p$. Deduce that if q is a prime divisor of $2^p - 1$, then $q = 2pk + 1$ for some integer k.

    (b) Similarly, for odd prime p prove that if $p \mid 2^{2^n} + 1$ then $\text{ord}_p(2) = 2^{n+1}$. Deduce that if p is a prime divisor of $2^{2^n} + 1$, that p must be of the form $2^{n+1}k + 1$ for some integer k.

    (c) Using WolframAlpha, $2^{32}+1 = 641 \cdot 6700417$. Use the previous part to explain how one could identify 641 as a possible factor, and prove by hand that $641 \mid 2^{32} + 1$ using modular arithmetic, so that 641 is the smallest prime divisor of $2^{32} + 1$. Similarly, find by hand the smallest prime divisor of $2^{29} - 1$.

17. Let $a > 0$ with a not divisible by p for prime p. For $1 \leq j \leq p-1$, write $a \cdot j = pq_j + r_j$ by the division algorithm. Fermat's little theorem was proven by multiplying these equations together.

    (a) Adding them instead, prove that $\sum_{j=1}^{p-1} \lfloor \frac{aj}{p} \rfloor = \frac{(a-1)(p-1)}{2}$.

    (b) Consider the triangle with vertices $(0, 0)$, $(p, 0)$ and (p, a). Show that the expression in part (a) counts the number of lattice points in the interior of this triangle.

18. This exercise is a primality test based on Fermat's little theorem. For any integer a and prime p with $p \nmid a$, Fermat's little theorem says $a^{p-1} \equiv 1 \bmod p$, so taking the contrapositive of this statement says that if $a^{p-1} \not\equiv 1 \bmod p$, then p is composite! This gives rise to an algorithm for testing if an integer n is prime or not:

    - Randomly pick an integer a with $n \nmid a$.
    - If $a^{n-1} \not\equiv 1 \bmod n$, then n is composite and we're done!
    - Otherwise, if $a^{n-1} \equiv 1 \bmod n$, pick a different integer a and repeat.

    If a is an integer such that $a^{n-1} \not\equiv 1 \bmod n$, we call a a *Fermat witness* for the compositeness of n.

    (a) Using WolframAlpha or any other computer algebra system, find the smallest Fermat witness for 2821.

    (b) Let $m = 56052361$. Using WolframAlpha or any other computer algebra system, determine if $2, 3, 5, 6, 7, 10$, or 11 are Fermat witnesses for m. What do you find? Is this enough information to tell you with certainty if m is prime or composite, and why?

    We call n a *Carmichael number* if $a^{n-1} \equiv 1 \bmod n$ for all integers a with $\gcd(a, n) = 1$. Carmichael numbers are the integers for which our primality test will never give us any information.

(c) Prove that 561 is a Carmichael number. *(Hint:* $561 = 3 \cdot 11 \cdot 17$. *Euler's theorem might be useful.)*

19. Euler's theorem says for any integer a with $\gcd(a, n) = 1$, that $a^{\varphi(n)} \equiv 1 \bmod n$. However, $\varphi(n)$ is often times not the smallest exponent we can choose with this property. The *Carmichael function* $\lambda(n)$ is defined to be the smallest positive integer k such that $a^k \equiv 1 \bmod n$ for all integers a with $\gcd(a, n) = 1$. It turns out, for example, that $\lambda(1729) = 36$, and so every integer a with $\gcd(a, 1729) = 1$ satisfies $a^{36} \equiv 1 \bmod 1729$.

    (a) Prove that n is Carmichael number if and only if $\lambda(n) \mid n - 1$.

    (b) Compute $\lambda(3), \lambda(11)$ and $\lambda(17)$.

    (c) Compute $\lambda(561)$.

20. The goal of this problem is to give an alternate proof of Euler's theorem.

    (a) Prove that for any integer k with $1 \leq k \leq p - 1$, that $\binom{p}{k} \equiv 0 \bmod p$.

    (b) For any integer $a \geq 0$, prove by induction on a that $a^p \equiv a \bmod p$. Deduce that for $\gcd(a, p) = 1$, $a^{\varphi(p)} \equiv 1 \bmod p$.

    (c) Prove that for any integers a, k with $k \geq 1$ and $\gcd(a, p) = 1$, that $a^{\varphi(p^k)} \equiv 1 \bmod p^k$.

    (d) Prove that for integers a, n with $\gcd(a, n) = 1$ that $a^{\varphi(n)} \equiv 1 \bmod n$.

21. Let p be an odd prime and let $k \geq 0$ be an integer. Prove that

$$1^k + 2^k + \ldots + (p-1)^k \equiv \begin{cases} 0 \bmod p & p - 1 \nmid k \\ -1 \bmod p & p - 1 \mid k \end{cases}$$

22. Prove the following generalization of Wilson's theorem: if $(\mathbb{Z}/n\mathbb{Z})^\times$ has a generator, then the product of all units mod n is congruent to $-1 \bmod n$.

23. The goal of this problem is to count the number of solutions to the congruence $x^2 \equiv 1 \bmod n$ for arbitrary $n > 1$.

    (a) Let $p > 2$ be prime and let $k \geq 1$. Prove that $x^2 \equiv 1 \bmod p^k$ if and only if $x \equiv \pm 1 \bmod p^k$.

    (b) Prove by induction that for $k \geq 3$, $x^2 \equiv 1 \bmod 2^k$ has exactly four solutions: $x \equiv \pm 1 \bmod 2^k$ and $x \equiv \pm 1 + 2^{k-1} \bmod 2^k$.

    (c) Let $f(x) \in \mathbb{Z}[x]$, and let $\gcd(m, n) = 1$. Let $N(k)$ denote the number of solutions to the congruence $f(x) \equiv 0 \bmod k$. Prove that $N(mn) = N(m)N(n)$. For $f(x) = x^2 - 1$, compute $N(n)$ for arbitrary $n > 1$ in terms of the prime factorization of n.

    (d) Find all solutions to $x^2 \equiv 1 \bmod 1708$.

24. Now, you'll do the same for the congruence $x^2 \equiv -1 \bmod n$.

(a) Let p be an *odd* prime and $k \geq 1$. Prove $x^2 \equiv -1 \bmod p^k$ has either 0 or 2 solutions, depending on the congruence class of p mod 4.

(b) Prove that $x^2 \equiv -1 \bmod 2^k$ is solvable if and only if $k = 0$ or $k = 1$.

(c) Give a formula that counts the number of solutions to $x^2 \equiv -1 \bmod n$ in terms of the prime factorization of n. Use your formula to count the number of solutions to $x^2 \equiv -1 \bmod 2 \cdot 5^3 \cdot 13$.

25. Let p be a prime. By Hensel's lemma, for any $n \geq 2$ there are exactly $p-1$ roots of $x^{p-1} - 1 \bmod p^n$, coming from the lifts of $1, 2, \ldots, p-1 \bmod p$.

(a) Let $n > 0$ and suppose that $x \equiv y \bmod p^n$. Prove that $x^p \equiv y^p \bmod p^{n+1}$.

(b) Define a map $\omega : \mathbb{F}_p^\times \to (\mathbb{Z}/p^n\mathbb{Z})^\times$ by $x \bmod p \to x^{p^{n-1}} \bmod p^n$. Show that ω is well-defined, injective, and satisfies $\omega(x)^{p-1} \equiv 1 \bmod p^n$.

(c) Use the previous part to compute the solutions to $x^4 \equiv 1 \bmod 6125$.

26. Let $v_p(n)$ denote the highest power of p that divides n in its prime factorization, e.g. $v_3(7) = 0$ and $v_3(9) = 2$. In this problem, you'll see how you can use lifting to compute orders mod n.

(a) (Let p be an odd prime and suppose that $a \equiv 1 \bmod p$. Prove that $v_p(a^n - 1) = v_p(n) + v_p(a-1)$. *(Hint: consider separately the cases $p \nmid n$ and $n = p$ before handling the general case.)*

(b) Let p be an odd prime and let $d = \mathrm{ord}_p(a)$. Prove that

$$\mathrm{ord}_{p^k}(a) = \begin{cases} d & m \geq k \\ d \cdot p^{k-m} & m < k \end{cases}$$

where $m = v_p(a^d - 1)$.

(c) Compute $\mathrm{ord}_{847847}(3)$ using the previous part and properties of orders. (Note: $847847 = 7^2 \cdot 11^3 \cdot 13$).

27. ASCII is the standard way of converting symbols into numbers. Let $A = 65, B = 66, \ldots, Z = 90$, and let a space correspond to 32. (This is not a particularly *secure* method for converting text into numbers, but this is a simple illustration of how your computer might do so without any other algorithm.)

(a) Let $(m, e) = (4951760154835678088235319297, 185056762330061596630395 4877)$ be the public key. Convert the message "HELLO WORLD" into a number and using WolframAlpha, use the RSA method to encrypt your message.

(b) Suppose that you are Eve the spy. Using WolframAlpha, compute the factorization of m and then compute $\varphi(m)$. Then, compute the decryption key d. Use this to decode the intercepted message $(55799119760817384352725395, 31323399839857355784726744 02)$ that Bob sent to Alice, and convert it back to plain text (using a decimal to ASCII converter, perhaps).

As you can see, the numbers used in this example are not even remotely difficult for a computer to break RSA with. The actual recommendation for primes used for serious applications of RSA are between 309 to 617 digits long!

Part III

Quadratic Number Theory

Chapter 5

Quadratic congruences

5.1 Quadratic residues

Now that we have a good understanding of the structure of $(\mathbb{Z}/n\mathbb{Z})^\times$, we focus our attention on quadratic congruences of the form $x^2 \equiv a \bmod n$.

Definition 5.1.1. We call a a **quadratic residue** mod n if the congruence $x^2 \equiv a \bmod n$. Otherwise, we call a a **quadratic non-residue** mod n. Notationally, we will write $a \equiv \square \bmod n$ to mean that a is a quadratic residue mod n.

In the previous chapter, we saw that when $n = p$ is prime, we have the following criteria for determining when a is a quadratic residue mod n or not.

Proposition 40 (Euler's criterion)**.** *Let p be a prime and $p \nmid a$. Then*

$$a^{\frac{p-1}{2}} \equiv \begin{cases} 1 & a \equiv \square \bmod p \\ -1 & a \not\equiv \square \bmod p \end{cases}$$

Proof. Note that $a^{p-1} \equiv 1 \bmod p$ means that $(a^{\frac{p-1}{2}} - 1)(a^{\frac{p-1}{2}} + 1) \equiv 0 \bmod p$. Therefore, $a^{\frac{p-1}{2}} \equiv \pm 1 \bmod p$. We saw before that $a^{\frac{p-1}{2}} \equiv 1 \bmod p$ if and only if $a \equiv \square \bmod p$, which means that if $a \not\equiv \square \bmod p$ we must have $a^{\frac{p-1}{2}} \equiv -1 \bmod p$. □

Corollary 5.1.2. *For p odd, there are exactly $\frac{p-1}{2}$ non-zero quadratic residues mod p.*

Proof. There are $p-1$ non-zero elements mod p, each of which is either a root of $x^{\frac{p-1}{2}} - 1$ or $x^{\frac{p-1}{2}} + 1 \bmod p$ by the above. This means each polynomial has exactly $\frac{p-1}{2}$ roots because they have at most $\frac{p-1}{2}$ roots each. We're done once we combine with the above proposition. □

Definition 5.1.3. Let p be an odd prime and $p \nmid a$. The **Legendre symbol** $(\frac{a}{p})$ is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & a \equiv \square \bmod p \\ -1 & a \not\equiv \square \bmod p \end{cases}$$

Proposition 41. *Let p be an odd prime and $p \nmid a, b$.*

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

2. $(\frac{ab}{p}) \equiv (\frac{a}{p})(\frac{b}{p}) \bmod p$.

3. *If* $a \equiv b \bmod p$ *then* $(\frac{a}{p}) = (\frac{b}{p})$.

Proof.

1. This follows immediately from Euler's criterion.

2. By 1., $(ab)^{\frac{p-1}{2}} \equiv (\frac{ab}{p}) \bmod p$, and $(ab)^{\frac{p-1}{2}} \equiv a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} \equiv (\frac{a}{p})(\frac{b}{p}) \bmod p$.

3. Trivial.

□

Corollary 5.1.4. *Let p be an odd prime. Then $(\frac{-1}{p}) = (-1)^{\frac{p-1}{2}}$.*

Proof. The above proposition tells us that $(\frac{-1}{p}) \equiv (-1)^{\frac{p-1}{2}} \bmod p$. Since both sides of the congruence are ± 1, because p is odd, this is only possible if they are actually equal. □

In particular, casing on when $\frac{p-1}{2}$ is even or not, this tells us that $-1 \equiv \square \bmod p$ if and only if $p \equiv 1 \bmod 4$.

As the Legendre symbol is multiplicative, this means that if $a = p_1^{e_1} \cdots p_k^{e_k}$ is the prime factorization of a, we have $(\frac{a}{p}) = (\frac{p_1}{p})^{e_1} \cdots (\frac{p_k}{p})^{e_k}$. Therefore, understanding how to compute Legendre symbols reduces to understanding how to compute Legendre symbols of the form $(\frac{q}{p})$ for prime q.

It's also worth pointing out that the Legendre symbol can be viewed as a group homomorphism $(\frac{\cdot}{p}) : \mathbb{F}_p^\times \to \mathbb{Z}/2\mathbb{Z} \cong \langle -1 \rangle \leq \mathbb{C}^\times$. In other words, the Legendre symbol is a *character* of the group $\mathbb{F}_p^\times$. The benefit of this viewpoint is that it allows one to deduce properties of the Legendre symbols via *Fourier analysis*. The main result of the next section, the quadratic reciprocity law, is frequently proven using this view point in many number theory textbooks by examining *Gauss sums*, which are, essentially, just the Fourier transform of the Legendre symbol $(\frac{\cdot}{p})$.

5.2 Quadratic reciprocity

Numerically, if one tabulates whether p is a square mod q and q is a square mod p for various primes p and q, there is a (non-obvious!) pattern:

- If one of $p, q \equiv 1 \bmod 4$, then $p \equiv \square \bmod q$ if and only if $q \equiv \square \bmod p$.
- If both $p, q \equiv 3 \bmod 4$, then $p \equiv \square \bmod q$ if and only if $q \not\equiv \square \bmod p$.

This is the so called *law of quadratic reciprocity*, and was first proven by Gauss. He gave *8* different proofs of this result! Today, there are over 250 different proofs of quadratic reciprocity. It's a very central result in number theory!

Theorem 5.2.1 (Quadratic reciprocity)**.** *Let p and q be odd primes.*

1. $(\frac{-1}{p}) = (-1)^{\frac{p-1}{2}}$
2. $(\frac{2}{p}) = (-1)^{\frac{p^2-1}{8}}$
3. $(\frac{p}{q})(\frac{q}{p}) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$

The first two parts of the theorem are often known as the *supplementary laws*, with the third part often being synonymous with the phrase "quadratic reciprocity". We are going to give Eisenstein's proof of quadratic reciprocity, which is a modification of Gauss's third proof. We've already proven the first supplementary law in the previous section, so we will proceed by proving the second.

Proposition 42. $(\frac{2}{p}) = (-1)^{\frac{p^2-1}{8}}$

Proof. The idea is a rather clever counting argument. Set $s = \frac{p-1}{2}$, and consider the s equations:

$$\begin{aligned} 1 &= (-1)\cdot(-1) \\ 2 &= 2\cdot(-1)^2 \\ 3 &= (-3)\cdot(-1)^3 \\ &\vdots \\ s &= (\pm s)\cdot(-1)^s \end{aligned}$$

where the sign in the last equation is chosen so that the equality is correct. Multiply all these equations together. The left hand side is just $s!$. What happens on the right hand side? There's clearly a factor of $(-1)^{1+2+\ldots+s} = (-1)^{\frac{s(s+1)}{2}}$, all the even integers $2, 4, \ldots,$ up until s, and some negative odd numbers. Note that $2s \equiv -1 \bmod p$, $2(s-1) \equiv -3 \bmod p$, etc. so that modulo p, these negative odd numbers are really just even numbers in disguise! Therefore, we find that $s! \equiv (-1)^{\frac{s(s+1)}{2}} s! 2^s \bmod p$, where each factor of 2 gets paired up with one of the integers in $s!$ to produce the alternating sign pattern that we have. Since $s!$ is a unit mod p, we find that $2^s \equiv (-1)^{\frac{s(s+1)}{2}} \bmod p$. Using that $s = \frac{p-1}{2}$, we find that $(\frac{2}{p}) \equiv 2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}} \bmod p$. Since both sides are $\pm 1 \bmod p$, this forces actual equality which is we wanted. □

As part of our proof, we need the following technical lemmas:

Proposition 43 (Gauss's lemma)**.** *Suppose that p is an odd prime and $p \nmid a$. For $1 \leq k \leq \frac{p-1}{2}$, write $a \cdot k \equiv \varepsilon_k r_k \bmod p$ with $\varepsilon_k = \pm 1$ and $0 < r_k < \frac{p}{2}$. Then $(\frac{a}{p}) = (-1)^\mu$, where $\mu = \#\{k : \varepsilon_k = -1\}$.*

Proof. First, we show that all such r_k are distinct. Suppose that $r_i \equiv r_j \bmod p$ for some i, j. Then $a \cdot i \equiv \varepsilon_i r_i \equiv \varepsilon_i r_j \bmod p$, and $a \cdot j \equiv \varepsilon_j r_j \equiv \varepsilon_j r_i \bmod p$. Therefore, $a \cdot i \cdot \varepsilon_j \equiv a \cdot j \cdot \varepsilon_j \bmod p$, so that $i \equiv \pm j \bmod p$, as $\varepsilon_i, \varepsilon_j$ are either ± 1. As $1 \leq i, j \leq \frac{p-1}{2}$, the only way this is possible is if $i = j$. Therefore, for all k in the range of interest, we must have r_k are distinct. $1 \leq r_k \leq \frac{p-1}{2}$, this then says that $\{r_k\} = \{1, 2, \ldots, \frac{p-1}{2}\}$. Taking the product of the $\frac{p-1}{2}$ congruences $a \cdot k \equiv \varepsilon_k r_k \bmod p$, we find that $a^{\frac{p-1}{2}}(\frac{p-1}{2})! \equiv (\frac{p-1}{2})! \prod_{k=1}^{\frac{p-1}{2}} \varepsilon_k \bmod p$. Since $(\frac{p-1}{2})!$ is invertible mod p, and the value of $\prod_{i=1}^{\frac{p-1}{2}} \varepsilon_k$ only depends on the number of negative 1's in the product, this yields $(\frac{a}{p}) \equiv a^{\frac{p-1}{2}} \equiv (-1)^{\mu} \bmod p$, and then actual equality follows since both sides are $\pm 1 \bmod p$. □

Proposition 44 (Eisenstein's lemma)**.** *Let p be an odd prime and $p \nmid a$. Then $(\frac{a}{p}) = (-1)^m$, where $m = \sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{2ak}{p} \rfloor$.*

Proof. Write $2ak = pq_k + r'_k$ by the division algorithm for $1 \leq k \leq \frac{p-1}{2}$ and $0 \leq r'_k \leq p-1$. Note that $q_k = \lfloor \frac{2ak}{p} \rfloor$, and that because p is odd, $q_k \equiv r'_k \bmod 2$. Therefore, $m = \sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{2ak}{p} \rfloor \equiv \sum_{k=1}^{\frac{p-1}{2}} q_k \equiv \sum_{k=1}^{\frac{p-1}{2}} r'_k \bmod 2$. Thus, $(-1)^m = (-1)^{\sum_{k=1}^{\frac{p-1}{2}} r'_k} = \prod_{k=1}^{\frac{p-1}{2}} (-1)^{r'_k}$. Using the notation of the previous proposition, write $a \cdot k \equiv \varepsilon_k r_k \bmod p$. Then this says $2ak \equiv r'_k \equiv 2\varepsilon_k r_k \bmod p$. In the proof of Gauss's lemma, we showed that $\{r_k\}$ is a rearrangement of $\{1, 2, \ldots, \frac{p-1}{2}\}$, and therefore this means that $\{2r_k\}$ consists of the even integers $\{2, 4, \ldots, p-1\}$. Note that if $\varepsilon_k = 1$, then r'_k must be even. Otherwise, if $\varepsilon_k = -1$, then r'_k is congruent to a negative even integer mod p, which is congruent to a positive *odd* integer because p is odd! Therefore, $\prod_{k=1}^{\frac{p-1}{2}} (-1)^{r'_k} = (-1)^{\mu}$. This means that $m \equiv \mu \bmod 2$, so we're done by Gauss's lemma. □

The above lemmas are not particularly enlightening, and are not seemingly very useful. The brilliance in Eisenstein's proof is that the sum in the previous proposition has a very simple geometric interpretation, which leads to a very clean geometric proof of quadratic reciprocity.

Theorem 5.2.2 (Quadratic Reciprocity)**.** *Let p, q be distinct odd primes. Then $(\frac{p}{q})(\frac{q}{p}) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}$.*

Proof. The key observation is that the sum $\sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{2q}{p} k \rfloor$ counts the number of lattice points in the interior of the triangle with vertices $(0, 0)$, $(p, 0)$ and (p, q) that have *even* x-coordinate (this is an exercise!).

Consider the rectangle with vertices $(0, 0), (p, 0), (0, q), (p, q)$. There are $p - 1$ columns, each with $q - 1$ lattice points, and these correspond to all the lattice points inside the rectangle. Consider the diagonal from $(0, 0)$ to (p, q). Since p and q are relatively prime, there are no lattice points on this diagonal. Therefore, for each column, the number of lattice points above the diagonal has the same parity as the number of lattice points below the diagonal, because each column has an *even* number of points.

Consider the triangle with vertices $(\frac{p}{2}, \frac{q}{2}), (\frac{p}{2}, q), (p, q)$. There is a bijection between the lattice points in the interior of this triangle with *even* x-coordinate and lattice points in the interior of the triangle whose vertices are $(0,0), (\frac{p}{2}, 0), (\frac{p}{2}, \frac{q}{2})$ that have *odd* x-coordinate, given by reflecting across the line $y = \frac{q}{2}$ and then across the line $x = \frac{p}{2}$. Putting this together, the number of even x-coordinate lattice points inside the large triangle that are in the right half of the rectangle has the same parity as the number of even x-coordinate lattice points inside the upper triangle with even x-coordinate, which is the same as the number of lattice points in the lower left triangle with odd x-coordinate, so that $\sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{2q}{p} k \rfloor$ counts the *total* number of lattice points inside the lower left triangle.

By symmetry, $\sum_{k=1}^{\frac{q-1}{2}} \lfloor \frac{2p}{q} k \rfloor$ counts the total number of lattice points inside the triangle with vertices $(0,0), (0, \frac{q}{2}), (\frac{p}{2}, \frac{q}{2})$, so $(\frac{p}{q})(\frac{q}{p}) = (-1)^{\sum_{k=1}^{\frac{p-1}{2}} \lfloor \frac{2q}{p} k \rfloor + \sum_{k=1}^{\frac{q-1}{2}} \lfloor \frac{2p}{q} k \rfloor}$.

This exponent is the number of lattice points inside the rectangle whose vertices are $(0,0), (\frac{p}{2}, 0), (0, \frac{q}{2}), (\frac{p}{2}, \frac{q}{2})$, which is simply $\frac{p-1}{2} \cdot \frac{q-1}{2}$. This proves $(\frac{p}{q})(\frac{q}{p}) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}$, which is what we wanted. □

Armed with quadratic reciprocity, we can now see some applications.

Example 5.2.3. The quadratic congruence $2x^2 + 5x - 9 \equiv 0 \bmod 101$ has a solution if and only if the expression $\sqrt{97}$ makes sense in $\mathbb{F}_{101}$ by the quadratic formula. This is equivalent to computing $(\frac{97}{101})$. Since $101 \equiv 1 \bmod 4$, by quadratic reciprocity, $(\frac{97}{101}) = (\frac{101}{97}) = (\frac{4}{97}) = 1$, so the congruence is solvable. The solutions are given by $x \equiv \frac{-5 \pm \sqrt{97}}{2} \bmod 101$, which after doing the computation yields $x \equiv 19, 29 \bmod 101$.

Example 5.2.4. To compute $(\frac{79}{101})$, we repeatedly apply quadratic reciprocity over and over until we arrive at a symbol we can easily calculate. We have $(\frac{79}{101}) = (\frac{101}{79}) = (\frac{22}{79}) = (\frac{2}{79})(\frac{11}{79}) = (\frac{11}{79}) = -(\frac{79}{11}) = -(\frac{2}{11}) = 1$.

Example 5.2.5. For which prime p is $-5 \equiv \square \bmod p$? Equivalently, for which prime p is $(\frac{-5}{p}) = 1$? Factoring the Legendre symbol, we want to find all prime p such that $(\frac{-1}{p})(\frac{5}{p}) = 1$, so we need $(\frac{-1}{p}) = (\frac{5}{p}) = 1$ or $(\frac{-1}{p}) = (\frac{5}{p}) = -1$. By quadratic reciprocity, $(\frac{5}{p}) = (\frac{5}{p}) \equiv p^2 \bmod 5$. Therefore, we see that $(\frac{5}{p}) = 1$ for $p \equiv \pm \bmod 5$ and $(\frac{5}{p}) = -1$ for $p \equiv \pm 2 \bmod 5$. Since $(\frac{-1}{p}) = 1$ for $p \equiv 1 \bmod 4$ and $(\frac{-1}{p}) = -1$ for $p \equiv 3 \bmod 4$, we get the following systems

$$\begin{cases} p \equiv 1 \bmod 4 \\ p \equiv \pm 1 \bmod 5 \end{cases} \qquad \begin{cases} p \equiv 3 \bmod 4 \\ p \equiv \pm 2 \bmod 5 \end{cases}$$

resulting in $p \equiv 1, 3, 7, 9 \bmod 20$.

5.3 Jacobi reciprocity

Definition 5.3.1. Let $n > 1$ be odd, and write $n = p_1^{e_1} \cdots p_k^{e_k}$. For any a with $\gcd(a, n) = 1$, the **Jacobi symbol** $(\frac{a}{n})$ is defined by

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{e_1} \cdots \left(\frac{a}{p_k}\right)^{e_k}$$

The Jacobi symbol is then an extension of the Legendre symbol to non-prime denominator. It should be clear why such a definition would be interesting: the Jacobi symbol allows factorization of both the numerator *and* the denominator, so the computations are potentially easier.

Proposition 45. *Let $n > 1$ be odd and a, b with $\gcd(a, n) = \gcd(b, n) = 1$. Then the following properties hold:*

1. $(\frac{ab}{n}) = (\frac{a}{n})(\frac{b}{n})$
2. $(\frac{a}{mn}) = (\frac{a}{m})(\frac{a}{n})$

Proof. This is immediate from the definition of the Jacobi symbol. □

Unlike the Legendre symbol, the Jacobi symbol does *not* detect if $a \equiv \square \bmod n$. For example, $(\frac{2}{15}) = (\frac{2}{3})(\frac{2}{5}) = (-1) \cdot (-1) = 1$ by the above proposition, but obviously $x^2 \equiv 2 \bmod 15$ is not solvable because it's not solvable mod 3 or mod 5.

However, it *is* true that if $(\frac{a}{n}) = -1$, that $a \not\equiv \square \bmod n$, simply because this means $x^2 \equiv a \bmod p_i$ is not solvable for some prime factor p_i of n.

Example 5.3.2. $(\frac{28}{45}) = (\frac{2}{45})^2(\frac{7}{3})^2(\frac{7}{5}) = -1$, so $x^2 \equiv 28 \bmod 45$ is not solvable. The obstruction comes from $x^2 \equiv 28 \equiv 3 \bmod 5$ not being solvable.

Much like Legendre symbols, Jacobi symbols have a version of reciprocity.

Theorem 5.3.3 (Jacobi Reciprocity)**.** *Let $n, m > 1$ be odd integers that are co-prime. Then the following hold:*

1. $(\frac{-1}{n}) = (-1)^{\frac{n-1}{2}}$
2. $(\frac{2}{n}) = (-1)^{\frac{n^2-1}{8}}$
3. $(\frac{n}{m})(\frac{m}{n}) = (-1)^{\frac{m-1}{2} \cdot \frac{n-1}{2}}$

We postpone the proof until after the following remark. The main application of Jacobi symbols is that they make calculating Legendre symbols easier. The Legendre symbol *requires* one the factor the numerator before being able to apply quadratic reciprocity, whereas Jacobi symbols do *not* require factoring to apply Jacobi reciprocity unless the numerator is even! It's easy to compute the exponent of 2 in the factorization of any integer, so this is not computationally an issue. As there are not efficient factoring algorithms, this means using Jacobi reciprocity is in general, going to be superior.

Example 5.3.4. Suppose we wanted to compute $(\frac{1001}{9907})$ using Jacobi reciprocity. We just continually flip the symbol, and factor if even integers appear. We have the following sequence of steps: $(\frac{1001}{9907}) = (\frac{898}{1001}) = (\frac{2}{1001})(\frac{449}{1001}) = (\frac{449}{1001}) = (\frac{103}{449}) = (\frac{37}{103}) = (\frac{29}{37}) = (\frac{8}{29}) = (\frac{2}{29})^3 = -1$.

We now proceed to the proof of Jacobi reciprocity. Since the Jacobi symbol is defined in terms of the Legendre symbol, the proof will not be very hard.

Lemma 5.3.5. *For any $a, b > 1$ odd and $e \geq 1$ we have:*

1. $\frac{ab-1}{2} \equiv \frac{a-1}{2} + \frac{b-1}{2} \bmod 2$
2. $\frac{a^2b^2-1}{8} \equiv \frac{a^2-1}{8} + \frac{b^2-1}{8} \bmod 2$
3. $\frac{a^e-1}{2} \equiv e\frac{a-1}{2} \bmod 2$
4. $\frac{a^{2e}-1}{8} \equiv e\frac{a^2-1}{8} \bmod 2$

Proof.

1. Since a, b are odd we have $(a-1)(b-1) \equiv 0 \bmod 4$. This means $ab - 1 \equiv (a-1) + (b-1) \bmod 4$, so dividing by 2 yields the result.
2. The proof is very similar, note that $(a^2-1)(b^2-1) \equiv 0 \bmod 16$ so that $a^2b^2 - 1 = (a^2-1) + (b^2-1) \bmod 16$. The result follows upon dividing by 8.
3. We have $(a^e - 1) \equiv (a-1)(1 + a + \ldots + a^{e-1}) \bmod 4$. Divide by 2 and use that a is odd to get $\frac{a^e-1}{2} \equiv e\frac{a-1}{2} \bmod 2$.
4. Identical to 3., just work mod 16 and divide by 8 instead.

□

Corollary 5.3.6. *Let $a_1, \ldots, a_k$ be odd positive integers and $e_1, \ldots, e_k \geq 1$.*

1. $\sum_{i=1}^k e_i \frac{a_i-1}{2} \equiv \frac{(a_1^{e_1} \cdots a_k^{e_k})-1}{2} \bmod 2$
2. $\sum_{i=1}^k e_i \frac{a_i^2-1}{8} \equiv \frac{(a_1^{e_1} \cdots a_k^{e_k})^2-1}{8} \bmod 2$

Proof. Both statements easily follow by induction and using the previous lemma. □

Proof. (of Jacobi reciprocity)

1. Write $n = p_1^{e_1} \cdots p_k^{e_k}$. By definition, we have $(\frac{-1}{n}) = (\frac{-1}{p_1})^{e_1} \cdots (\frac{-1}{p_k})^{e_k} = (-1)^{\sum_{i=1}^k e_i \frac{p_i-1}{2}} = (-1)^{\frac{n-1}{2}}$ by the above corollary.
2. Same as the above, except using the other part of the corollary.
3. Let $m = p_1^{e_1} \cdots p_k^{e_k}$ and $n = q_1^{f_1} \cdots q_k^{f_\ell}$ be the prime factorizations of m and n respectively. Factoring and using quadratic reciprocity, we have $(\frac{n}{m})(\frac{m}{n}) = (-1)^{\sum_{i=1}^k \sum_{j=1}^\ell e_i f_j \frac{p_i-1}{2} \cdot \frac{q_j-1}{2}} = (-1)^{\frac{m-1}{2} \cdot \frac{n-1}{2}}$ by applying the corollary twice.

□

5.4 Squares mod n

To wrap up the discussion of quadratic congruences, we'll talk briefly about solutions to quadratic congruences mod n. For a general quadratic congruence $ax^2 + bx + c \equiv 0 \bmod n$, if $\gcd(n, 2a) = 1$ then by the quadratic formula, the congruence will have a solution if and only if $b^2 - 4ac$ is a square mod n. Under these assumptions, understanding when quadratic congruences are solvables reduces to understanding when quadratic congruences of the form $x^2 \equiv a \bmod n$ are solvable.

For the sake of keeping the analysis simple, we'll investigate what happens when a is *odd*, so that the headaches involved with trying to solve quadratic equations mod powers of two are kept to a minimum.

Proposition 46. *Let $e \geq 3$. Then $x^2 \equiv a \bmod 2^e$ is solvable if and only if $a \equiv 1 \bmod 8$.*

Proof. The forward direction is clear, as if $x^2 \equiv a \bmod 2^e$ then $x^2 \equiv a \bmod 8$ and all odd integers square to 1 mod 8. Conversely, if $e = 3$ and $a \equiv 1 \bmod 8$, then obviously $x^2 \equiv 1 \bmod 8$ is solvable. Now, assume that $e > 3$ and that we know the statement "$a \equiv 1 \bmod 8 \implies x^2 \equiv a \bmod 2^k$ is solvable" holds for all $3, \ldots, k$. We need to then construct a solution mod 2^{k+1}. If $x^2 \equiv a \bmod 2^{k+1}$ were a solution, then necessarily we must have $x^2 \equiv a \bmod 2^k$. By induction hypothesis, this has a solution, say $c^2 \equiv a \bmod 2^k$. This means that $a = c^2 + 2^k\ell$ for some ℓ. The claim is that $c' = c + 2^{k-1}\ell$ is a a solution to the desired congruence. Indeed, squaring we find that $(c')^2 = c^2 + 2^k c\ell + 2^{2k-2}\ell^2 \equiv a + 2^k\ell(c-1) \bmod 2^{k+1}$ as $2k - 2 \geq k + 1$ for $k \geq 3$. Note that c must be *odd*, as a is odd, which means that $2 \mid c - 1$. This then shows that $(c')^2 \equiv a \bmod 2^{k+1}$, as desired. Therefore by induction, if $a \equiv 1 \bmod 8$ then $x^2 \equiv a \bmod 2^e$ is solvable for all $e \geq 3$. □

Theorem 5.4.1. *Let $n = 2^e p_1^{e_1} \cdots p_k^{e_k}$ be the prime factorization of n, and suppose that $\gcd(a, n) = 1$. Then the congruence $x^2 \equiv a \bmod n$ is solvable if and only if $(\frac{a}{p_i}) = 1$ for all i, $a \equiv 1 \bmod 4$ if $e = 2$, and $a \equiv 1 \bmod 8$ if $e \geq 3$.*

Proof. If $x^2 \equiv a \bmod n$ is solvable then clearly $x^2 \equiv a \bmod p_i$ holds for all i. If $e = 2$ then $a \equiv 1 \bmod 4$ because all odd integers square to 1 mod 4, and if $e \geq 3$ then $a \equiv 1 \bmod 8$ because all odd integers square to 1 mod 8.

Conversely, suppose that $(\frac{a}{p_i}) = 1$ for all i. Then by Hensel's lemma, for each i there are exactly two solutions to $x^2 \equiv a \bmod p_i^{e_i}$, given by lifting up the two solutions to the congruence $x^2 \equiv a \bmod p$. If $e = 2$ and $a \equiv 1 \bmod 4$, then clearly $x^2 \equiv a \bmod 4$ is solvable. If $a \equiv 1 \bmod 8$ then by the above lemma, $x^2 \equiv a \bmod 2^e$ is solvable. Therefore, we have $x^2 \equiv a \bmod 2^e$ is solvable and $x^2 \equiv a \bmod p_i^{e_i}$ are all solvable, so by the Chinese remainder theorem, we must have $x^2 \equiv a \bmod n$ is solvable by gluing together a solution to each congruence. □

5.5 Exercises

1. Determine if the congruence $3x^2 + 6x + 5 \equiv 0 \bmod 89$ is solvable.

2. Use Jacobi reciprocity to compute the Legendre symbols $\left(\frac{113}{997}\right)$, $\left(\frac{514}{1093}\right)$, and $\left(\frac{401}{757}\right)$.

3. Use quadratic reciprocity to find all primes p such that $\left(\frac{7}{p}\right) = 1$. Do the same for $\left(\frac{15}{p}\right) = 1$.

4. Find a prime number p that can be simultaneously expressed in all three of the following forms: $x^2 + y^2$, $x^2 + 2y^2$, $x^2 + 3y^2$.

5. Show that for any prime p, the polynomial $f(x) = x^4 - 10x^2 + 1 \in \mathbb{Z}[x]$ is reducible in $\mathbb{F}_p[x]$.

6. Recall that a *Mersenne prime* is a prime of the form $2^p - 1$, and a *Fermat prime* is a prime of the form $2^{2^n} + 1$. (Refer to exercise 4.7.16).

   (a) Let p be a prime dividing $2^{2^n} + 1$. Use Legendre symbols to prove that $p \equiv 1 \bmod 2^{n+2}$ (*Hint:* show that $2^{n+1} \mid \frac{p-1}{2}$).

   (b) Show that if p is an odd prime, then every prime divisor q of $2^p - 1$ must satisfy $q \equiv \pm 1 \bmod 8$.

   (c) Prove by hand that $2^{17} - 1$ is not prime.

7. (a) Let p be an odd prime. Prove that $\sum_{a=1}^{p-1} \left(\frac{a}{p}\right) = 0$.

   (b) Let p be a prime with $p \equiv 1 \bmod 4$. Prove that $\sum_{a=1}^{p-1} a\left(\frac{a}{p}\right) = 0$.

8. Let $N = p_1 \cdots p_k$ be a product of distinct odd primes. Prove that there exists an integer a such that $\left(\frac{a}{N}\right) = -1$.

9. In this exercise, you'll see how to prove special cases of Dirichlet's theorem using your knowledge of Legendre symbols!

   (a) Prove there are infinitely many primes of the form $4k + 1$. (*Hint*: consider $N = (2p_1 \cdots p_n)^2 + 1$.)

   (b) Prove there are infinitely many primes of the form $8k + 7$. (*Hint*: Construct a similar choice of N which has a prime factor that is 7 mod 8.)

Chapter 6

Conic sections

6.1 Rational points on conics

Now having a solid understanding of when quadratic equations have solutions mod n, we turn our attention back to the integers. It's trivial to determine when $ax^2 + bx + c = 0$ has solutions in $\mathbb{Z}$, so our next goal will be trying to understand when the curve $C : ax^2+bxy+cy^2+dx+ey+f = 0$ has solutions in $\mathbb{Z}^2$. Such curves C are called *conic sections*.

There are many interesting problems in number theory that essentially boil down to finding integer or rational solutions on curves. The classical example that we'll start with is the problem of trying to classify all *Pythagorean triples*, which are the integer points on the surface $x^2 + y^2 = z^2$. We start with the following simple observation: if (x, y, z) is a solution in $\mathbb{Z}^3$ to $x^2 + y^2 = z^2$, then $(x/z, y/z)$ is a solution to $x^2 + y^2 = 1$. In other words, $(x/z, y/z)$ is a *rational* point on the unit circle S^1. Conversely, any rational point (r, s) on S^1 can be written so that both coordinates have a common denominator, say $r = \frac{a}{c}$ and $s = \frac{b}{c}$, from which $r^2 + s^2 = 1$ results in $a^2 + b^2 = c^2$. Therefore, there is a correspondence

$$\{\text{Pythagorean triples}\} \longleftrightarrow \{\text{Rational point of } S^1\}$$

As it turns out, there is a rather easy geometric way to find rational points on the unit circle! First, we note that $(0, 1)$ lives on S^1. Consider the line $L_m : y - 1 = mx$ that passes through $(0, 1)$ and has slope m. For each value of m, L_m intersects the circle at a second point (where we count intersection with multiplicity). To find the intersection point, we must solve the quadratic equation $x^2+(1+mx)^2 = 1$. Expanding and solving for x yields $x = \frac{-2m}{1+m^2}$, and plugging back into the equation of L_m says that $y = \frac{1-m^2}{1+m^2}$. If $m \in \mathbb{Q}$, then the coordinates of this intersection point are *rational*, and so we've found a new rational point on S^1! On the other hand, starting with any rational point $(r, s) \neq (0, -1) \in S^1$ (this would correspond to an *infinite* slope!), the line passing through $(0, 1)$ and (r, s) has slope $\frac{s-1}{r}$, which is rational. This tells us the set of rational points on S^1 is $\{(\frac{-2m}{1+m^2}, \frac{1-m^2}{1+m^2}) : m \in \mathbb{Q}\} \cup \{(0, -1)\}$, and that there is a bijection

$$\{\text{Rational points of } S^1\} \setminus \{(0, -1)\} \longleftrightarrow \mathbb{Q}$$

Witing $m = \frac{u}{v}$ for $\gcd(u, v) = 1$ as a fraction in lowest terms, the rational point corresponding to m is $(-\frac{2uv}{u^2+v^2}, \frac{v^2-u^2}{u^2+v^2})$. From our initial observation, this then corresponds to the

Pythagorean triple $(-2uv, v^2-u^2, u^2+v^2)$.

This idea can be adapted to find rational points on *any* conic section, provided we have found one point to start with!

Example 6.1.1. Suppose we wanted to find all rational points on the hyperbola $x^2-y^2=1$. We note that $(1,0)$ is such a point, and that for $m \in \mathbb{Q}$, the line $L_m : y = m(x-1)$ passing through $(1,0)$ has at most one other intersection point with the hyperbola. The coordinates of this intersection come from solving the equation $x^2-m^2(x-1)^2=1$, which is the quadratic equation $(1-m^2)x^2+2m^2x-(m^2+1)=0$. For $m=\pm 1$, we just recover the point $(1,0)$ (which is because the hyperbola has asymptotes of $y=\pm x$!). For $m \neq \pm 1$, we may divide to write this as $x^2+\frac{2m^2}{1-m^2}x-\frac{m^2+1}{1-m^2}=0$.

There's a trick to avoid messy algebra from trying to use the quadratic formula. For a general quadratic equation x^2+ax+b, there's a relation between the roots and the coefficients of the polynomial: for two roots $r_1, r_2 \in \mathbb{C}$, the factorization $x^2+ax+b=(x-r_1)(x-r_2)$ tells us that $r_1r_2=b$ and $r_1+r_2=-a$. We know that one of the roots must be $r_1=1$. Why? Solving the quadratic equation tells us the x-coordinates of the points on the hyperbola that is passes through. By construction, one of those points is $(1,0)$! Therefore, we find that the roots are $r-1=1$ and $r_2=\frac{m^2+1}{m^2-1}$. Therefore, the intersection point is $(\frac{m^2+1}{m^2-1}, \frac{2m}{m^2-1})$. Once more, we note that rational choices of m produce rational points, and any rational point must come from a rational slope. This tells us that the set of rational points is $\{(\frac{m^2+1}{m^2-1}, \frac{2m}{m^2-1}) : m \neq \pm 1 \in \mathbb{Q}\} \cup \{(1,0)\}$.

That this procedure works is not entirely obvious. For the conic $C : ax^2+bxy+cy^2+dx+ey+f=0$, one may prove that there is a new coordinate system given by an invertible affine change of variables such that, with respect to these new coordinates X and Y, C takes on one of three "reduced" forms: $X^2+BY^2=D$, $X^2=Y$, or $X^2=E$ for some $B, D, E \in \mathbb{Z}$ with $B \neq 0$. The proof of this fact is rather technical and not enlightening, so we just take it for granted. From there, however, we may observe that the procedure works for any of these three reduced forms by using relation between roots and coefficients of a quadratic to see that because the coefficients of the quadratic resulting from trying to find the of intersection point of the conic C and L_m has *rational* coefficients, and one root is, by assumption, *rational*, the other must be as well.

There is *another* method for finding Pythagorean triples, using algebra instead of geometry. The idea comes from the observation that if $x^2+y^2=z^2$ in $\mathbb{Z}^3$, then this means $(x+iy)(x-iy)=z^2$ in $\mathbb{Z}[i]$, and so this is now a question about factorization in $\mathbb{Z}[i]$.

Theorem 6.1.2. *Let (x,y,z) a primitive Pythagorean triple. Then $x=m^2-n^2$, $y=2mn$ and $z=m^2+n^2$ for some integers m,n. For any choice of $m>n>0$, and $\gcd(m,n)=1$, the resulting values of x,y,z form a primitive Pythagorean triple, which is uniquely determined by (m,n).*

Proof. Suppose that (x,y,z) is a primitive Pythagorean triple, so that $x^2+y^2=z^2$. First, we note that z must be odd. If it was even, then working mod 4 we must have $x^2+y^2 \equiv 0 \bmod 4$.

The only way this is possible is if $x, y \equiv 0 \bmod 4$, which would mean both x, y are even, contradicting that (x, y, z) is primitive. As a consequence, exactly one of x, y can be even. Without loss of generality, we will assume that x is odd and y is even.

In $\mathbb{Z}[i]$, this means that $(x + iy)(x - iy) = \alpha\bar{\alpha} = z^2$. First, observe that α and $\bar{\alpha}$ are relatively prime in $\mathbb{Z}[i]$. This is because if δ is a greatest common divisor of α and $\bar{\alpha}$, then $\delta \mid \alpha \pm \bar{\alpha}$, so $\delta \mid 2\gcd(x, y) = 2$. The factorization of 2 in $\mathbb{Z}[i]$ is $2 = -i(1 + i)^2$, so the only non-trivial divisor is (up to unit multiple) $1 + i$. Could $1 + i$ be a common divisor? The answer is no. If $1 + i \mid \alpha$ then taking norms, $2 \mid N(\alpha) = z^2$ which is not possible, as z^2 is odd. Therefore, $\gcd(\alpha, \bar{\alpha})$ must be a unit. Since $\alpha\bar{\alpha} = z^2$ and $\alpha, \bar{\alpha}$ share no prime factors in common, by unique factorization the only way this is possible is if both $\alpha, \bar{\alpha}$ are perfect squares in $\mathbb{Z}[i]$. Therefore, we may write $\alpha = (m+ni)^2$ for some integers m, n. Equating real and imaginary parts yields $x = m^2 - n^2$ and $y = 2mn$. This says $z^2 = x^2 + y^2 = (m^2 + n^2)^2$. As $z > 0$, this means $z = m^2 + n^2$.

That all such choices $x = m^2-n^2, y = 2mn, z = m^2+n^2$ for $m > n > 0$ and $\gcd(m, n) = 1$ actually are solutions to $x^2 + y^2 = z^2$ is easy to verify. Must (x, y, z) be primitive? If it were not, suppose that p is a common prime factor. As z is odd, we must have $p \neq 2$. Therefore, from $p \mid mn$ we must have $p \mid m$ or $p \mid n$. Either way, $m^2 \equiv n^2 \bmod p$ would mean both $m, n \equiv 0 \bmod p$, which is a contradiction. Therefore, (x, y, z) is indeed a primitive triple. Finally, from the above paragraph we see that (m, n) correspond to one of the square roots of $x + yi \in \mathbb{Z}[i]$. The other square root is $-m - ni$, which is not included as a possibility. □

6.2 Sums of squares

What about finding *integer* points on circles? This is harder! We'll start with a famous question of Fermat: For integer $n > 0$, when does the equation $x^2 + y^2 = n$ have integer solutions? Equivalently, when is n a sum of two squares in $\mathbb{Z}$?

The first observation is the following:

Theorem 6.2.1 (Fermat)**.** *Let $p \in \mathbb{Z}$ be an odd prime. Then $x^2 + y^2 = p$ has integer solutions if and only if $p \equiv 1 \bmod 4$.*

Proof. Firstly, suppose that $x^2 + y^2 = p$ has solutions. Then working mod p, this means that $x^2 + y^2 \equiv 0 \bmod p$. Since neither $x, y \equiv 0 \bmod p$, this means that $(x/y) \equiv -1 \bmod p$, so $(\frac{-1}{p}) = 1$, which means that $p \equiv 1 \bmod 4$. On the other hand, suppose that $p \equiv 1 \bmod 4$, so $x^2 \equiv -1 \bmod p$ is solvable. This means there is some integer x with $p \mid x^2 + 1$ in $\mathbb{Z}$, or in $\mathbb{Z}[i]$, $p \mid (x + i)(x - i)$. If p was prime, this would then mean that $p \mid x \pm i$, which is clearly impossible. Therefore, p is composite, so we may write $p = \alpha\beta$ as a non-trivial factorization in $\mathbb{Z}[i]$. Taking norms, this means $p^2 = N(\alpha)N(\beta)$ in $\mathbb{Z}$. Since the factorization is non-trivial, this forces $N(\alpha) = N(\beta) = p$. Writing $\alpha = x + yi$, this says $x^2 + y^2 = p$ as desired. □

The proof above gives us a complete classification of how primes split in $\mathbb{Z}[i]$:

Corollary 6.2.2. *Let $p \in \mathbb{Z}$ be a prime. The factorization of p in $\mathbb{Z}[i]$ is:*

- $p = (1+i)(1-i) = -(1+i)^2$ *if* $p = 2$.
- $p = \pi\bar{\pi}$ *for a Gaussian prime* π *if* $p \equiv 1 \bmod 4$.
- p *remains prime in* $\mathbb{Z}[i]$ *if* $p \equiv 3 \bmod 4$.

Proof.

- Obvious.
- The above proof show that if $p \equiv 1 \bmod 4$ then $p = N(\pi) = \pi\bar{\pi}$ for some Gaussian integer π, which is then prime because its norm is prime.
- Once more, the above proof shows that if p factored it would have to be a sum of two squares.

□

The splitting of integer primes p in $\mathbb{Z}[i]$ then tells us a complete classification of *all* primes in $\mathbb{Z}[i]$, after we observe the following:

Proposition 47. *Let $\pi \in \mathbb{Z}[i]$ be a prime. Then $\pi \mid p$ in $\mathbb{Z}[i]$ for some prime $p \in \mathbb{Z}$.*

Proof. $N(\pi)$ is an integer, and therefore is a product of primes in $\mathbb{Z}$. Since π is prime, in $\mathbb{Z}[i]$, this means π divides one of these primes by Euclid's lemma. □

Corollary 6.2.3. *Let $\pi \in \mathbb{Z}[i]$ be prime. Then up to unit multiple, π is one of the following:*

- $1 + i$.
- α *or* $\bar{\alpha}$, *where* $\alpha \in \mathbb{Z}[i]$ *satisfies* $N(\alpha) = p$ *for some prime* $p \in \mathbb{Z}$ *with* $p \equiv 1 \bmod 4$.
- p *for* $p \in \mathbb{Z}$ *primes and* $p \equiv 3 \bmod 4$.

Proof. Follows immediately from the above proposition + corollary. □

Returning back to the geometry, we're ready to prove the following:

Theorem 6.2.4 (Fermat)**.** *Let $n > 1$ be an integer. Then $x^2 + y^2 = n$ has integer solutions if and only if for every prime $p \mid n$ with $p \equiv 3 \bmod 4$, we have $v_p(n)$ is even.*

Proof. Write $n = 2^e p_1^{e_1} \cdots p_k^{e_k} \cdots q_1^{f_1} \cdots q_\ell^{f_\ell}$, where $p_i \equiv 1 \bmod 4$ and $q_j \equiv 3 \bmod 4$. By the previous theorem, we know that $p_i = \pi_i \bar{\pi}_i$ for some Gaussian prime π_i. Therefore, we see that $2^e p_1^{e_1} \cdots p_k^{e_k} = N((1+i)^e \alpha_i^{e_1} \cdots \alpha_k^{e_k})$, which is then a sum of two squares. So it's sufficient to only consider the part of the factorization that are coming from q_j.

If all f_j are even, then $q_j^{f_j} = N(q_j^{j/2})$, or equivalently, $q_j^{f_j} = (q_j^{f_j/2})^2 + 0^2$. Therefore, $n = N((1+i)^e \alpha_i^{e_1} \cdots \alpha_k^{e_k} q_1^{f_1/2} \cdots q_j^{f_j/2})$ so that n is a sum of two squares. It remains to show

that if e_j is odd for some j, then n cannot be written as a sum of two squares.

To handle this, we proceed by induction. If $n = 2$, then n has no prime factors that are 3 mod 4 and so there's nothing to prove. Suppose we know for all integers $2 \leq k < n$ that if n is the sum of two squares, the multiplicity of all prime factors that are 3 mod 4 is even. If n has no prime factors that are 3 mod 4, there's nothing to prove and we're done. Otherwise, assume that n has some prime factor p with $p \equiv 3 \bmod 4$. As $n = x^2 + y^2$ and $p \mid n$, this means $p \mid (x + iy)(x - iy)$ in $\mathbb{Z}[i]$. Since $p \equiv 3 \bmod 4$ we've seen that p is prime in $\mathbb{Z}[i]$, so this means $p \mid x \pm iy$ by Euclid's lemma, which then says that $p \mid x, y$. If $n = p^2$ we're done, so assume that $n > p^2$. We may write $n = p^2(x/p)^2 + (y/p)^2$, so $n/p^2 = (x/p)^2 + (y/p)^2$. As $2 \leq n/p^2 < n$, by induction hypothesis , $v_p(n/p^2)$ is even, and therefore $v_p(n)$ is even as well. By induction, we're done. □

Example 6.2.5. Let $n = 45 = 3^2 \cdot 5$. By the above theorem, n can be written as a sum of two squares. The proof tells us exactly how to do this: in $\mathbb{Z}[i]$, we have $5 = N(1 + 2i)$ and $3^2 = N(3)$, so we may write $45 = N(3 + 6i) = (3 + 6i)(3 - 6i) = 3^2 + 6^2$.

Example 6.2.6. What are *all* integer points on $x^2 + y^2 = 45$? From the above example, each different prime $x + yi$ of norm 5 produces the point $(3x, 3y)$ on the circle. There are precisely 8 different primes of norm 5: the unit multiples of the conjugate primes $1 + 2i$ and $1 - 2i$. This results in the 8 integer points $(3, \pm 6), (-3, \pm 6), (6, \pm 3), (-6, \pm 3)$.

Alternatively, we could have just brute forced our way this solution: if $x^2 + y^2 = 45$ then we clearly see that $|x| \leq 6$. We can then set $y^2 = 45 - x^2$ just plug in $x = \pm 1, \ldots, \pm 6$ to find which resulting y values are perfect square.

As before, once we've found one rational point, we then know all rational points on $x^2 + y^2 = 45$.

Example 6.2.7. We've seen that $(3, 6)$ is an integer point. Consider $L_m : y - 6 = m(x - 3)$ of arbitrary rational slope m. Then as m varies among the rational numbers, the intersection of L_m with $x^2 + y^2 = 45$ will hit all other rational points on the circle, with the exception of $(3, -6)$. Solving for the intersection point, we find that the rational points on this circle are $\{(\frac{3(m^2-4m-1)}{m^2+1}, \frac{-6(m^2+m-1)}{m^2+1}) : m \in \mathbb{Q}\} \cup \{(3, -6)\}$. Writing $m = \frac{u}{v}$ for $\gcd(u, v) = 1$, we get $\{(\frac{3(u^2-4uv-v^2)}{u^2+v^2}, \frac{-6(u^2+uv-v^2)}{u^2+v^2} : \gcd(u, v) = 1\} \cup \{(3, -6)\}$.

How can we find the integer points from the rational points? Firstly, we need a count. There are certainly finitely many integer solutions to this equation. If (x, y) is one solution to $x^2+y^2 = 45$, then so are $(\pm x, y), (x, \pm y), (\pm y, x), (y, \pm x)$. Therefore, the number of solutions is divisible by 8. Also, if $x^2 + y^2 = 45$ then $x^2 + y^2 \equiv 0 \bmod 3$. If one of $x, y \not\equiv 0 \bmod 3$, this means either $(x/y)^2 \equiv -1 \bmod 3$ or $(y/x)^2 \equiv -1 \bmod 3$, which is impossible as $(\frac{-1}{3}) = -1$. Therefore, both $x, y \equiv 0 \bmod 3$, and so $(x/3)^2 + (y/3)^2 = 5$. Therefore, it's sufficient to count the number of solutions to $x^2 + y^2 = 5$. Clearly, $|x|, |y| \leq 2$, so there are at most 16 solutions, and it's also obvious that there are < 16 solutions, so there are exactly 8 integer points on $x^2 + y^2 = 45$. We know $(3, 6)$ is one of them, so therefore all of them are given by

$(3, \pm 6), (-3, \pm 6), (6, \pm 3), (-6, \pm 3)$.

How could we find values of (u, v) that produce these points? Taking $(u, v) = (0, \pm 1)$ results in $(-3, 6)$, so assume $u^2 + v^2 > 1$. Necessarily, we need that $u^2 + v^2 \mid 3(u^2 - 4uv - v^2)$ and $u^2 + v^2 \mid -6(u^2 + uv - v^2)$. Let p be a prime factor of $u^2 + v^2$. We must have either $p = 2$, or $p \equiv 1 \bmod 4$, and $u^2 + v^2 \equiv 0 \bmod p$. We'll ignore the possibility of $p = 2$ for the moment, and therefore, we require that $3(u^2 - 4uv - v^2) \equiv 0 \bmod p$, so that $u^2 - v^2 \equiv 4uv \bmod p$. Adding with $u^2 + v^2 \equiv 0 \bmod p$ says $2u^2 \equiv 4uv \bmod p$, and subtracting says $2v^2 \equiv -4uv \bmod p$. If both $u, v \not\equiv 0 \bmod p$, then we would have $u \equiv 2v \bmod p$ and $v \equiv -2u \bmod p$, so that $5u \equiv 0 \bmod p$, which tells us that $p = 5$. If one of $u, v \equiv 0 \bmod p$, then the other must be as well, which is not possible since $\gcd(u, v) = 1$, so we may just assume that $u, v \not\equiv 0 \bmod p$. From $u^2 + v^2 \equiv 0 \bmod 5$, we have $(v/u)^2 \equiv -1 \equiv 4 \bmod 5$, so this means $v \equiv 2u \bmod 5$ or $v \equiv 3u \bmod 5$. We also require that $u^2 + uv - v^2 \equiv 0 \bmod 5$, but note $u^2 + uv - v^2 \equiv u^2 - 4uv - v^2 \bmod 5$ so this says nothing new.

There are 8 potential points in $(\mathbb{F}_5^\times)^2$ that satisfy $v \equiv 2u \bmod 5$ and $v \equiv 3u \bmod 5$: these are $(1, 3), (2, 1), (3, 4), (4, 2), (1, 2), (2, 4), (3, 1), (4, 3)$. Only three of them could actually yield integer points: $(1, 3), (2, 1), (4, 2)$, because these are the only three which satisfy $u^2 - 4uv - v^2 \equiv 0 \bmod 5$. We can then search through (u, v) with $(u, v) \equiv (1, 3), (2, 1), (4, 2) \bmod 5$ to find values that work. For example, the 7 integer points produced by the line procedure can be found by taking $(u, v) = (0, 1), (1, -2), (1, 3), (2, 1), (-3, 1), (1, 1), (1, -1)$.

This might seem rather useless, however there is reason to investigate this solution more closely: it reveals a connection between *integer points* on the circle $x^2 + y^2 = 45$ and *lattice points* of certain lattices.

6.3 Sums of squares and lattices

We start with some definitions.

Definition 6.3.1. A **lattice** of $\mathbb{R}^n$ is $L = \mathrm{Span}_{\mathbb{Z}}\{v_1, \ldots, v_n\}$ for $\mathbb{R}$-linearly independent vectors $v_1, \ldots, v_n \in \mathbb{R}^n$. The **fundamental parallelogram** T of the lattice L is $T = \{c_1 v_1 + \ldots + c_n v_n : 0 \leq c_i \leq 1\}$.

Definition 6.3.2. A set $S \subset \mathbb{R}^n$ is called **convex** if for any $x, y \in S$ and any $0 \leq t \leq 1 \in \mathbb{R}$, the line $tx + (1 - t)y \in S$. We say that S is **symmetric around the origin** if for any $x \in S$, we also have $-x \in S$.

Our interest in lattices is the following theorem of Minkowski:

Theorem 6.3.3 (Minkowski)**.** *Let $S \subset \mathbb{R}^n$ be a convex set that is symmetric around the origin, and let L be a lattice in $\mathbb{R}^n$. If $Vol(S) > 2^n Vol(T)$, then S contains a non-zero lattice point of L.*

Proof. First, we start with the case $L = \mathbb{Z}^n$. Consider the map $f : S \to \mathbb{R}^n/2L$ by $x \mapsto x \bmod 2L$. We claim that f cannot be an injection. If it were, we would be able

to fit S inside a hyper-cube of volume 2^n, which is not possible because $\text{Vol}(S) > 2^n$. Therfore, there are two distinct points x, y such that $f(x) = f(y)$. This means that $x \equiv y \bmod 2L$, so x and y differ by a non-zero element of $2L$, i.e. $x = y + 2p$ for some non-zero lattice point p. Since S is symmetric around the origin, $-x \in S$ and because S is convex, the line between $-x$ and y must be contained in S. Therefore, the midpoint $\frac{1}{2}(-x + y) = p$ is contained in S, which is what we wanted.

For the general case, let $L = \text{Span}_{\mathbb{Z}}\{v_1, \ldots, v_n\}$, and let $A = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$. Then A maps $\mathbb{Z}^n$ to L, and because $\text{Vol}(S) > 2^n \text{Vol}(T) = 2^n \det(A)$, this means that $\text{Vol}(A^{-1}S) > 2^n$. By the above case, there is a non-zero lattice point x of $\mathbb{Z}^n$ contained in $A^{-1}S$, and mapping back says Ax is a non-zero lattice point of L contained in S. □

Using Minkowski's theorem, we can give another proof that primes congruent to 1 mod 4 are sums of squares.

Theorem 6.3.4 (Fermat)**.** *Let $p \equiv 1 \bmod 4$ be a prime. Then there are integers x, y such that $p = x^2 + y^2$.*

Proof. Since $p \equiv 1 \bmod 4$, we know that $(\frac{-1}{p}) = 1$. Choose $1 < k < p - 1$ such that $k^2 \equiv -1 \bmod p$, and consider the lattice $L = \text{Span}_{\mathbb{Z}}\left\{ \begin{pmatrix} 1 \\ k \end{pmatrix}, \begin{pmatrix} 0 \\ p \end{pmatrix} \right\}$. Note that any $\begin{pmatrix} x \\ y \end{pmatrix} \in L$ satisfies $y \equiv kx \bmod p$, which means that $x^2 + y^2 \equiv 0 \bmod p$. Therefore, every lattice point of L lies on the circle $x^2 + y^2 = kp$ for some integer k. Let S be the open disk $x^2 + y^2 < 2p$. Note that S is convex, symmetric around the origin, and has area $2\pi p > 4p$. Therefore, by Minkowski's theorem, S contains a non-zero lattice point of L. The only circle of the form $x^2 + y^2 = kp$ contained in S is $x^2 + y^2 = p$, so we're done. □

In our previous proof using unique factorization, we saw that $p = \pi\bar{\pi}$ for some Gaussian prime $\pi = x + yi$. This corresponds to the 8 distinct integer points on the circle $(x, \pm y), (-x, \pm y), (y, \pm x), (-y, \pm x)$, by writing $p = N(\alpha)$ for any of the unit multiples of $\pi, \bar{\pi}$.

In the spirit of the proof using Minkowski's theorem, is it possible to give a proof of this using a geometric argument? The answer is yes! To do so, we'll need another theorem about lattices:

Theorem 6.3.5 (Pick)**.** *Let L be a lattice in $\mathbb{R}^2$ and let P be a convex polygon with vertices on L. Then $Area(P) = (I + \frac{1}{2}B - 1)Area(T)$, where I is the number of lattice points of L inside P, and B is the number of lattice points of L on the boundary of P.*

We'll take Pick's theorem for granted for the moment, and return to our question. Fix a choice of k with $k^2 \equiv -1 \bmod p$. Take L to be the lattice in the above proof, $L = \text{Span}_{\mathbb{Z}}\left\{ \begin{pmatrix} 1 \\ k \end{pmatrix}, \begin{pmatrix} 0 \\ p \end{pmatrix} \right\}$ and let $C : x^2 + y^2 = p$ be the circle of interest. Note that for any $(x, y) \in C$, exactly one of (x, y) or $(x, -y)$ must live in $C \cap L$: this is because $x^2 + y^2 \equiv 0 \bmod p$ means that $x \equiv \pm ky \bmod p$, and if both (x, y) and $(x, -y) \in L$ then this would mean $y \equiv -y \equiv kx \bmod p$, so $y \equiv 0 \bmod p$, and this is impossible. Therefore,

showing that there are exactly 8 integer points is equivalent to showing that $C \cap L$ contains exactly 4 points.

If $(x, y) \in C \cap L$, this tells us that $(-x, -y), (-y, x), (y, -x)$ are all in $C \cap L$ as well, which means that $|C \cap L$ is divisible by 4. Now, construct the convex polygon P having its vertices the points of $C \cap L$. The polygon P is contained entirely on and within the circle. By Pick's theorem, $\text{Area}(P) = (I + \frac{B}{2} - 1) \cdot p$. As all non-zero lattice points of L lie on a circle of non-zero radius, there is a single interior point of L contained in P, namely, the origin. Therefore, $I = 1$, and all other lattice points are on the boundary. Thus, $|C \cap L| = \frac{2\text{Area}(P)}{p} < \frac{2 \cdot \pi p}{p} = 2\pi$. This forces $|C \cap L| = 4$, which is what we wanted!

Pick's theorem has a surprising number of number theoretic applications. Here's another one, proving a remark we made earlier in chapter 3:

Proposition 48. *Let $\alpha \in \mathbb{Z}[i]$. Then $|\mathbb{Z}[i]/(\alpha)| = N(\alpha)$.*

Proof. Identify $\mathbb{C}$ with $\mathbb{R}^2$ in the usual way, which identifies $\mathbb{Z}[i]$ with the lattice $\mathbb{Z}^2$. Consider the lattice $L = \text{Span}_{\mathbb{Z}}\{\alpha, i\alpha\}$. As we saw before, saying that $\beta \equiv \beta' \bmod \alpha$ is the same thing as saying that β, β' differ by a lattice point of L, and therefore $\mathbb{Z}[i]/(\alpha) \cong \mathbb{R}^2/L$. The different equivalence classes of $\mathbb{Z}[i]/(\alpha)$ come from the lattice points of $\mathbb{Z}^2$ inside the fundamental parallelogram T of L, and *half* the number of points on the boundary, because the points "above" the diagonal of the parallelogram can be obtained by moving along the lattice from the lower half. However, this counts the origin *twice*, so we need to subtract 1. By Pick's theorem, this count is precisely $\text{Area}(T)$. This is given by $|(x, y, 0) \times (-y, x, 0)| = x^2 + y^2 = N(\alpha)$ as desired. □

Example 6.3.6. With the new results about $\mathbb{Z}[i]$ from this chapter, we can understand quotients of $\mathbb{Z}[i]$ better.

- Let $\alpha = 3 - i \in \mathbb{Z}[i]$. Note that $\alpha = (1 - i)(2 + i)$ is a prime factorization of α. By the Chinese remainder theorem, $\mathbb{Z}[i]/(\alpha) \cong \mathbb{Z}[i]/(1 - i) \times \mathbb{Z}[i]/(2 + i)$. Each of these rings are fields of size 2 and 5 respectively, and so are isomorphic to $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$. Therefore, $\mathbb{Z}[i]/(\alpha) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \cong \mathbb{Z}/10\mathbb{Z}$.

- With $\alpha = 5$, then $5 = (1 + 2i)(1 - 2i)$ is a prime factorization in $\mathbb{Z}[i]$. Therefore, $\mathbb{Z}[i]/(5) \cong (\mathbb{Z}/5\mathbb{Z})^2$ by the reasoning above.

- If $\alpha = 2$, then $\mathbb{Z}[i]/(2) = \mathbb{Z}[i]/(1 + i)^2$, which is a ring of size 4, which is *not* a field, because $(1 + i)^2 \equiv 0 \bmod (1 + i)^2$. Up to isomorphism, there are two such possibilities for the isomorphism class of this ring: $(\mathbb{Z}/2\mathbb{Z})^2$ and $\mathbb{Z}/4\mathbb{Z}$. The former has no non-zero element with the property that $x^2 = 0$, and therefore, this forces $\mathbb{Z}[i]/(2) \cong \mathbb{Z}/4\mathbb{Z}$.

- With $\alpha = 7$, we know that α is prime in $\mathbb{Z}[i]$. Therefore, $\mathbb{Z}[i]/(7)$ is a field of size 49. As $\mathbb{Z}[i] \cong \mathbb{Z}[x]/(x^2 + 1)$, this means $\mathbb{Z}[i]/(7) \cong (\mathbb{Z}[x]/(x^2 + 1))/(7) \cong \mathbb{Z}[x]/(7, x^2 + 1) \cong (\mathbb{Z}[x]/(7))/(x^2 + 1) \cong \mathbb{F}_7[x]/(x^2 + 1)$. This latter ring is a field extension of $\mathbb{F}_7$ where $x^2 + 1$ has a root, which we may as well call i, so we may write $\mathbb{F}_7[x]/(x^2 + 1) \cong \mathbb{F}_7(i)$. Therefore, $\mathbb{Z}[i]/(7) \cong \mathbb{F}_7(i)$.

Now, we return to the proof of Pick's theorem.

Proof. (of Pick's theorem) Let $L = \text{Span}_{\mathbb{Z}}\{v_1, v_2\}$ and let $A = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$. Then A maps $\mathbb{Z}^2$ to L. As the image of a convex set under a linear transformation is convex, if we know the result for $L = \mathbb{Z}^2$, then we get the result for a general L by linear algebra.

The proof of Pick's theorem goes as follows:

1. Triangulating the polygon P, we may reduce to the case of triangles.
2. Pick's theorem holds for rectangles.
3. Cutting the rectangle in half diagonally, Pick's theorem holds for right triangles with sides parallel to the axes.
4. Any triangle can be embedded into a rectangle in such a way that it divides it into four triangles, three of which are right with sides parallel to the axes.

We leave the proof of each step as an exercise for the interested reader. □

6.4 Pell's equation and quadratic rings

Now, we turn our attention to hyperbolas. The general approach of section 1 tells us how to find rational points on hyperbolas, so we would like to focus on how to find *integer* points on hyperbolas. Any hyperbola can be transformed into a standard equation $x^2 - Dy^2 = N$ for some integers, D, N. When D is a perfect square, this turns into the equation $(x - Dy)(x + Dy) = N$, from which we can determine solutions from the factors of N. Therefore, the first "interesting" case comes from studying equations of the form $x^2 - Dy^2 = N$, for integers D, N with D square-free.

Definition 6.4.1. Let D be a square-free integer. The equation $x^2 - Dy^2 = 1$ is called a **Pell equation**.

Hyperbolas of this form are named after the English mathematician John Pell, although their study long predates him, appearing in texts of the Indian mathematician Brahmagupta in the 7th century.

This hyperbola always has the trivial points $(\pm 1, 0)$. A point (x, y) is called *positive* if $x, y > 0$. Here are two reasons why Pell's equation is interesting:

- Suppose that (a, b) is an integer solution to $x^2 - Dy^2 = 1$. This means that $a^2 - Db^2 = 1$, so $\sqrt{(\frac{a}{b})^2 - \frac{1}{b^2}} = \sqrt{D}$. For large denominators b, this means that $\frac{a}{b}$ is a "good" rational approximation to $\sqrt{D}$.
- An integer solution (a, b) to the Pell equation $x^2 - Dy^2 = 1$ is a unit in the ring $\mathbb{Z}[\sqrt{D}]$. Therefore, if one is interested in understanding units of quadratic rings, Pell's equation naturally arises.

Example 6.4.2. The point $(19601, 13860)$ lies on the hyperbola $x^2 - 2y^2 = 1$. Indeed, $\frac{19601}{13860} \approx 1.41421356$ agrees with $\sqrt{2}$ to 8 decimal places.

Both of these viewpoints are useful for understanding Pell equations. First, we'll show that any Pell equation $x^2 - Dy^2 = 1$ has a non-trivial solution. To do so, we need the following approximation result of Dirichlet:

Lemma 6.4.3 (Dirichlet)**.** *Let α be an irrational number. There are infinitely many integers x, y with $\gcd(x, y) = 1$ such that $|\alpha - \frac{x}{y}| < \frac{1}{y^2}$.*

Proof. Choose $n \geq 2$ and partition the half open interval $[0, 1)$ as $[0, 1) = [0, \frac{1}{n}) \cup \ldots \cup [\frac{n-1}{n}, 1)$. Consider $\{\alpha\} = \alpha - \lfloor \alpha \rfloor$, the fractional part of α. Then $\{\alpha\}$ lies in a unique one of these sets. Next, consider $\{0\}, \{\alpha\}, \{2\alpha\}, \ldots, \{n\alpha\}$. By the Pidgeonhole principle, two of these must lie in the same set. That is, there exist j, k with $0 \leq j, k \leq n$ and $k < j$ such that $|j\alpha - \lfloor j\alpha \rfloor - (k\alpha - \lfloor k\alpha \rfloor)| < \frac{1}{n}$. Set $y = j - k$ and $x = \lfloor j\alpha \rfloor - \lfloor k\alpha \rfloor$, so we have $|x - y\alpha| < \frac{1}{n}$. We may assume that $\gcd(x, y) = 1$, as otherwise we just divide by an integer, which strengthens the inequality. Next, because $0 \leq j, k \leq n$ we see that $0 < y < n$, which means $|x - y\alpha| = |\frac{x}{y} - \alpha| < \frac{1}{ny} < \frac{1}{y^2}$. This shows the existence of such a rational number $\frac{x}{y}$. It remains to show we can find infinitely many with this property.

With x, y as above, note that $|\frac{x}{y} - \alpha| \neq 0$ as α is irrational. Therefore, choose $m > \frac{1}{|\frac{x}{y} - \alpha|}$. This procedure produces x_1, y_1 such that $|\frac{x_1}{y_1} - \alpha| < \frac{1}{my_1} < |\frac{x}{y} - \alpha| < \frac{1}{y^2}$ and $0 < y_1 < m$. Repeating this process produces infinitely many such rational numbers. □

Lemma 6.4.4. *Let D be a positive, square-free integer. Then there is a constant M such that $|x^2 - Dy^2| < M$ has infinitely many integral solutions.*

Proof. In $\mathbb{R}$, we write $x^2 - Dy^2 = (x + \sqrt{D}y)(x - \sqrt{D}y)$. By the previous lemma, there are infinitely many pairs (x, y) with $\gcd(x, y) = 1$ with $|x - \sqrt{D}y| < \frac{1}{y}$. By the triangle inequality, $|x + \sqrt{D}y| = |x - \sqrt{D}y + 2\sqrt{D}y| \leq |x - \sqrt{D}y| + 2\sqrt{D}|y| < \frac{1}{y} + 2\sqrt{D}y$. Therefore, $|x^2 - Dy^2| = |x - \sqrt{D}y||x + \sqrt{D}y| < \frac{1}{y}(\frac{1}{y} + 2\sqrt{D}y) \leq 2\sqrt{D} + 1$. Taking $M = 2\sqrt{D} + 1$ finishes the proof. □

We're now ready to prove that Pell equations always have solutions.

Theorem 6.4.5 (Lagrange)**.** *Let $D > 0$ be a square-free integer. Then $x^2 - Dy^2 = 1$ has a non-trivial solution.*

Proof. By the previous lemma, there is an integer M such that $x^2 - Dy^2 = M$ for infinitely many integral solutions (x, y), and we may assume that $x, y > 0$ are distinct. Why? There are finitely many integers $-(\lfloor 2\sqrt{D} + 1 \rfloor), \ldots, \lfloor 2\sqrt{D} + 1 \rfloor$, and infinitely many solutions with $|x^2 - Dy^2| < 2\sqrt{D} + 1$, so by the pidgeonhole principle, one choice of M has infinitely many solutions. Now, we work modulo $|M|$. Since there are finitely many residue classes mod $|M|$, applying the pidgeonhole principle a second time, we can find two pairs $(x_1, y_1) \neq (x_2, y_2)$ such that $x_1 \equiv x_2 \bmod |M|$ and and $y_1 \equiv y_2 \bmod |M|$. Therefore, there are integers k, ℓ such that $x_2 = x_1 + Mk$ and $y_2 = y_1 + M\ell$, so that $x_2 + y_2\sqrt{D} = (x_1 + y_1\sqrt{D}) + M(k + \ell\sqrt{D})$. We may then compute that $(x_1 - y_1\sqrt{D})(x_2 + y_2\sqrt{D}) = M(a + b\sqrt{D})$ for some integers

a, b. Similarly, $(x_1 + y_1\sqrt{D})(x_2 - y_2\sqrt{D}) = M(a - b\sqrt{D})$, and so multiplying together yields $M^2 = M^2(a^2 - Db^2)$. This says that $a^2 - Db^2 = 1$. Finally, we note that $(a, b) \neq (\pm 1, 0)$ because if so, this would mean that $(x_1 - y_1\sqrt{D})(x_2 + y_2\sqrt{D}) = \pm M$, and multiplying through by $x_2 - y_2\sqrt{D}$ would say $M(x_1 - y_1\sqrt{D}) = \pm M(x_2 - y_2\sqrt{D})$. This would then imply that $x_1 = \pm x_2$, which is not possible because we assumed that $x_1, x_2 > 0$ and $x_1 \neq x_2$. Therefore, (a, b) is a non-trivial solution, which is what we wanted. □

The above theorem tells us that Pell equations always have a solution, but the proof does not help us *find* a solution. In order to understand this, we have to look through an algebraic lens.

Definition 6.4.6. Let $D \neq$ be a square-free integer. Set $\mathbb{Z}[\sqrt{D}] = \{a + b\sqrt{D} : a, b \in \mathbb{Z}\}$, which is a subring of either $\mathbb{R}$ or $\mathbb{C}$ depending on if $D > 0$ or $D < 0$. Addition is defined by $(a + b\sqrt{D}) + (c + d\sqrt{D}) = (a + c) + (b + d)\sqrt{D}$ and multiplication is defined by $(a + b\sqrt{D})(c + d\sqrt{D}) = (bdD + ac) + (ad + bc)\sqrt{D}$.

The **norm** of $\alpha = x + y\sqrt{D} \in \mathbb{Z}[\sqrt{D}]$ is defined by $N(\alpha) = x^2 - Dy^2$. We usually write $\overline{\alpha}$ to mean the conjugate expression $x - y\sqrt{D}$, but beware that this is *not* equal to the complex conjugate of α, unless $D < 0$!

We're already familiar with a special case of these quadratic rings: when $D = -1$ then we recover $\mathbb{Z}[i]$. In general, the arithmetic of these rings varies wildly depending on the value of D. For example, $\mathbb{Z}[\sqrt{3}]$ has unique factorization, but $\mathbb{Z}[\sqrt{-3}]$ does not. Neither $\mathbb{Z}[\sqrt{5}]$ or $\mathbb{Z}[\sqrt{-5}]$ have unique factorization, either! In general, we won't be able to say too much, but because all these rings have norms, they all share some similar properties.

Proposition 49. *For $\alpha, \beta \in \mathbb{Z}[\sqrt{D}]$, $N(\alpha\beta) = N(\alpha)N(\beta)$.*

Proof. Same proof as $\mathbb{Z}[i]$ but with i replaced by $\sqrt{D}$. □

Definition 6.4.7. We say that $\alpha \in \mathbb{Z}[\sqrt{D}]$ is a **unit** if there is some $\beta \in \mathbb{Z}[\sqrt{D}]$ such that $\alpha\beta = 1$. If $\alpha = x + y\sqrt{D}$ is a unit, we call α **positive** if $x, y > 0$.

Proposition 50. *$\alpha \in \mathbb{Z}[\sqrt{D}]$ is a unit if and only if $N(\alpha) = \pm 1$.*

Proof. If $\alpha\beta = 1$ then by multiplicativity, $N(\alpha)N(\beta) = 1$ in $\mathbb{Z}$, so $N(\alpha) = \pm 1$. Conversely, if $N(\alpha) = \pm 1$ then either $\overline{\alpha}$ or $-\overline{\alpha}$ is a multiplicative inverse of α, so it's a unit. □

The connection between quadratic rings and Pell's equation comes from units: saying $\alpha \in \mathbb{Z}[\sqrt{D}]$ is a unit with $N(\alpha) = 1$ is equivalent to saying there are integers x, y such that $x^2 - Dy^2 = 1$. In particular, the above theorem tells us there is always a non-trivial unit in $\mathbb{Z}[\sqrt{D}]$. From the multiplicativity of the norm, if $N(\alpha) = 1$ then $N(\alpha^k) = 1$ for any $k \geq 1$. Since $\alpha^k \in \mathbb{Z}[\sqrt{D}]$, we can find x_k, y_k such that $\alpha^k = x_k + y_k\sqrt{D}$, so that (x_k, y_k) is a solution to $x^2 - Dy^2 = 1$. Therefore, if we can find *one* non-trivial solution to a Pell equation, we therefore know infinitely many solutions. Also note that $\alpha^{-k} = \frac{1}{\alpha^k} = \overline{\alpha}^k$ by multiplying numerator/denominator by $\overline{\alpha}^k$ and using that $N(\alpha) = 1$. Since the coefficients of $\overline{\alpha}^k$ are *also* solutions $x^2 - Dy^2 = 1$, it's therefore true for all $k \in \mathbb{Z}$.

Example 6.4.8. Observe that $3+2\sqrt{2}$ is a unit in $\mathbb{Z}[\sqrt{2}]$ because $N(3+2\sqrt{2}) = 9-8 = 1$. Therefore, $(3,2)$ is a solution to $x^2 - 2y^2 = 1$. Any integer power of $3+2\sqrt{2}$ therefore produces another solution. For example, $(3+2\sqrt{2})^5 = 3363 + 2378\sqrt{2}$ produces the solution $(3363, 2378)$ and $(3+2\sqrt{2})^{-3} = 99 - 70\sqrt{2}$ produces the solution $(99, -70)$.

Next, our goal is to describe all possible solutions to Pell's equation.

Lemma 6.4.9. *Let (x,y) be a positive solution to $x^2 - Dy^2 = 1$, and let (x',y') be any other solution. Then $x + y\sqrt{D} < x' + y'\sqrt{D}$ if and only if $x < x'$ and $y < y'$.*

Proof. The backwards direction is obvious, so we need only to prove the forward direction. Suppose that (x,y) is a positive solution and (x',y') is any solution with $x+y\sqrt{D} < x'+y'\sqrt{D}$. Since (x,y) is positive, we have $x, y \geq 1$ so $x' + y'\sqrt{D} > x + y\sqrt{D} > 1$. Inverting says $0 < x' - y'\sqrt{D} < 1$, so $x' + y'\sqrt{D} > x' - y'\sqrt{D}$. This says $2y'\sqrt{D} > 0$, so $y' > 0$. Since $x' - y'\sqrt{D} > 0$, this says $x' > y'\sqrt{D} \geq \sqrt{D} > 1$, so (x',y') is a positive solution. Now from $x + y\sqrt{D} < x' + y'\sqrt{D}$, inverting says $x - y\sqrt{D} > x' - y'\sqrt{D}$, so $(x+x') + (y-y')\sqrt{D} < (x+x') + (y'-y)\sqrt{D}$, which yields $y' > y$. This then give $x^2 = 1 + Dy^2 < 1 + Dy'^2 = x'^2$, and since $x, x' > 0$ taking a square root tells us $x < x'$. □

If we find a positive solution (x,y) with y minimal, then in fact, x is minimal as well. This is because if (x',y') is any other solution, we have $x^2 = 1 + Dy^2 < 1 + Dy'^2 = x'^2$, and so once more, because (x,y) is positive this means (x',y') is positive, so $x < x'$. We will call a positive solution to $x^2 - Dy^2 = 1$ with x, y minimal the **fundamental solution** to $x^2 - Dy^2 = 1$.

Theorem 6.4.10. *Let (x_1, y_1) be the fundamental solution to $x^2 - Dy^2 = 1$. If (x,y) is any other positive solution to $x^2 - Dy^2 = 1$, then $x = x_n$ and $y = y_n$ for some $n \geq 1$, where $(x_1 + y_1\sqrt{D})^n = x_n + y_n\sqrt{D}$.*

Proof. Since (x_1, y_1) is a positive solution, we have $x_1 + y_1\sqrt{D} > 1$, so $(x_1 + y_1\sqrt{D})^n \to \infty$ as $n \to \infty$. Therefore we can find N such that $(x_1 + y_1\sqrt{D})^{N+1} > x + y\sqrt{D} \geq (x_1 + y_1\sqrt{D})^N$. Dividing through says $1 \leq (x + y\sqrt{D})(x_1 + y_1\sqrt{D})^{-N} < x_1 + y_1\sqrt{D}$. Write $(x + y\sqrt{D})(x_1 + y_1\sqrt{D})^{-N} = a + b\sqrt{D}$ for some a, b. The above results tell us that (a,b) is a solution to Pell's equation, and since $1 \leq a + b\sqrt{D}$, we have that (a,b) is a *positive* solution. Therefore, from $a + b\sqrt{D} < x_1 + y_1\sqrt{D}$, this means that $a < x_1$ and $b < y_1$, which contradicts that (x_1, y_1) is the fundamental solution. Therefore, $x + y\sqrt{D} = (x_1 + y_1\sqrt{D})^N$ for some N. □

Corollary 6.4.11. *Let (x,y) be the fundamental solution to $x^2 - Dy^2 = 1$. Then all solutions to $x^2 - Dy^2 = 1$ are of the form $\pm(x_n, y_n)$, where $(x + y\sqrt{D})^n = x_n + y_n\sqrt{D}$.*

Proof. In the above theorem, we proved that if (x',y') is a positive solution, then $(x',y') = (x_n, y_n)$ for some n, so it remains to handle the case that (x',y') is not positive. The number $x' + y'\sqrt{D}$ lies in one of the intervals $(0,1), (-1,0), (-\infty, -1)$, so exactly one of the numbers $\frac{1}{x'+y'\sqrt{D}} = x' - y'\sqrt{D}, -\frac{1}{x'+y'\sqrt{D}} = -(x' - y'\sqrt{D}), -(x' + y'\sqrt{D})$ lives in $(1, \infty)$. Each of these numbers still have norm 1 in $\mathbb{Z}[\sqrt{D}]$, and so their coefficients form solutions to Pell's equation. Therefore, $\pm(x' + y'\sqrt{D})^{\pm 1} = (x + y\sqrt{D})^n$ for some $n \geq 1$, which is what we wanted. □

Example 6.4.12. To solve the Pell equation $x^2 - 2y^2 = 1$, we need to find a fundamental solution, which comes from a solution (x, y) with y minimal. Setting $x^2 = 1 + 2y^2$, and plugging in values of y, we see that $(2, 3)$ is the fundamental solution. Therefore, any other positive solution (x_k, y_k) come from $x_k + y_k\sqrt{D} = (2 + 3\sqrt{2})^k$. To find explicit formulas for x_k and y_k, note that $x_k + y_k\sqrt{D} = (2 + 3\sqrt{2})^k$ and $x_k - y_k\sqrt{2} = (2 - 3\sqrt{2})^k$ by inverting. Therefore, $x_k = \frac{(3+2\sqrt{2})^k + (3-2\sqrt{2})^k}{2}$ and $y_k = \frac{(3+2\sqrt{2})^k - (3-2\sqrt{2})^k}{2\sqrt{2}}$.

Next, we turn our attention to the negative Pell equation $x^2 - Dy^2 = -1$. Since units in $\mathbb{Z}[\sqrt{D}]$ can have norm ± 1, this is the other equation we need to understand if we would like to understand the structure of the units. If $x+y\sqrt{D}$ has norm -1, then $(x+y\sqrt{D})^2$ has norm 1, because the norm is multiplicative. This means that solutions to $x^2 - Dy^2 = -1$ generate solutions to $x^2 - Dy^2 = 1$! Similarly to before, the fundamental solution of $x^2 - Dy^2 = -1$ will generate *all* solutions to either Pell equation.

Lemma 6.4.13. *Suppose that $x^2 - Dy^2 = -1$ and $x + y\sqrt{D} > 1$. Then $x, y \geq 1$.*

Proof. We have $\frac{1}{x+y\sqrt{D}} = -(x - y\sqrt{D}) = -x + y\sqrt{D}$, so $x + y\sqrt{D} > 1 > -x + y\sqrt{D} > 0$. Subtracting says that $2x > 0$, so $x \geq 1$. Therefore, $y\sqrt{D} > x$, so $y \geq 1$. □

Once more, we can order *positive* solutions to the negative Pell equation.

Lemma 6.4.14. *Suppose that $x^2 - Dy^2 = -1$ and (x', y') is any other positive solution. Then $x + y\sqrt{D} < x' + y'\sqrt{D}$ if and only if $x < x'$ and $y < y'$.*

Proof. Once again, the backwards direction is clear. Suppose that $x+y\sqrt{D} < x'+y'\sqrt{D}$. This means $x'+y'\sqrt{D} > x+y\sqrt{D} \geq 1+\sqrt{D} > 1$, so inverting tells us that $-x+y\sqrt{D} > -x'+y'\sqrt{D}$. Adding yields $(x' - x) + (y + y')\sqrt{D} > (x - x') + (y + y')\sqrt{D}$. This means that $2x' > 2x$, so $x' > x$. From this, $Dy^2 = x^2 + 1 < (x')^2 + 1 = D(y')^2$, so $y < y'$ since y, y' are both positive. □

Theorem 6.4.15. *Let (x_1, y_1) be the fundamental solution to $x^2 - Dy^2 = -1$. Then all solutions to $x^2 - Dy^2 = \pm 1$ are of the form $x + y\sqrt{D} = \pm(x_1 + y_1\sqrt{D})^k$ for some $j \in \mathbb{Z}$. The solutions to $x^2 - Dy^2 = -1$ have k odd and the solutions to $x^2 - Dy^2 = 1$ have k even.*

Proof. The idea is as follows: first, we show that $(x_1 + y_1\sqrt{D})^2$ yields the fundamental solution to the positive Pell equation $x^2 - Dy^2 = 1$. This will then tell us that positive solutions to $x^2 - Dy^2 = 1$ come from *even* powers of $(x_1 + y_1\sqrt{D})^2$, and then we just need to show that odd powers produce all positive solutions to $x^2 - Dy^2 = -1$. Once we know the result for positive solutions, we'll know it for all solutions using a similar argument to before.

Let (X_1, Y_1) denote the fundamental solution to $x^2 - Dy^2 = 1$, with associated unit $X_1 + Y_1\sqrt{D}$. Since $(x_1 + y_1\sqrt{D})^2$ has norm 1, we can write $(x_1 + y_1\sqrt{D})^2 = (X_1 + Y_1\sqrt{D})^k$ for some $k \geq 1$. If k is even, say, $k = 2\ell$, this would mean $x_1 + y_1\sqrt{D} = (X_1 + Y_1\sqrt{D})^\ell$. The right hand side has norm 1, while the left hand side has norm -1, which is a contradiction. Therefore, $k = 2\ell + 1$ must be odd. This means that $X_1 + Y_1\sqrt{D} = (a + b\sqrt{D})^2$, where $a + b\sqrt{D} = (x_1 + y_1\sqrt{D})(X_1 - Y_1\sqrt{D})^\ell$. Comparing coefficients, we see that $2ab = Y_1 > 0$, which means that a, b have the same sign. Therefore, without loss of generality, we may

assume that $a, b > 0$. If $\ell > 0$, this would mean that $(x_1 + y_1\sqrt{D})^2 = (X_1 + Y_1\sqrt{D})^{2\ell+1} > X_1 + Y_1\sqrt{D} = (a + b\sqrt{D})^2$, so taking square roots would say $x_1 + y_1\sqrt{D} > a + b\sqrt{D}$. Note that $a + b\sqrt{D}$ has norm -1, so by the previous lemma, this would mean $y_1 > b$. However, this would contradict the minimality of y_1. Thus, $\ell = 0$ and $X_1 + Y_1\sqrt{D} = (x_1 + y_1\sqrt{D})^2$ as desired.

Now, we show that positive solutions come from powers of $x_1 + y_1\sqrt{D}$. If (x, y) is a positive solution to $x^2 - Dy^2 = 1$, then we may write $x + y\sqrt{D} = (X_1 + Y_1\sqrt{D})^k = (x_1 + y_1\sqrt{D})^{2k}$ for some $k \geq 1$. This shows that solutions to the positive Pell equation come from even powers of the fundamental solution. Now, if (x, y) is a positive solution to $x^2 - Dy^2 = -1$, we can write $(x + y\sqrt{D})^2 = (X_1 + Y_1\sqrt{D})^k = (x_1 + y_1\sqrt{D})^{2k}$ for some $k \geq 1$. Taking square roots, $x + y\sqrt{D} = (x_1 + y_1\sqrt{D})^k$. Taking norms says $-1 = (-1)^k$, and therefore k must be odd.

This shows that *all* positive solutions to $x^2 - Dy^2 = \pm 1$ come from powers of $x_1 + y_1\sqrt{D}$. Now, suppose that we have a solution that is not positive. This means that $\alpha = x + y\sqrt{D}$ is not contained in $(1, \infty)$. However, one of $\frac{1}{\alpha}, -\frac{1}{\alpha}, -\alpha$ must be, and therefore we recover $x + y\sqrt{D} = \pm(x_1 + y_1\sqrt{D})^k$ for some $k \in \mathbb{Z}$, which is what we wanted. □

The important implication of the above result is that it tells us the structure of the unit group of $\mathbb{Z}[\sqrt{D}]$!

Corollary 6.4.16. *Let D be a square free integer. Then*

$$\mathbb{Z}[\sqrt{D}]^\times \cong \begin{cases} \mathbb{Z}/2\mathbb{Z} & D < -1 \\ \mathbb{Z}/4\mathbb{Z} & D = -1 \\ \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z} & D > 0 \end{cases}$$

Proof. If $D < -1$ the only solutions to $x^2 + Dy^2 = \pm 1$ are $(\pm 1, 0)$ corresponding to the units ± 1, and $\langle -1 \rangle \cong \mathbb{Z}/2\mathbb{Z}$. If $D = -1$ then $\mathbb{Z}[i]^\times = \langle i \rangle \cong \mathbb{Z}/4\mathbb{Z}$. If $D > 0$, we may write any unit in the form $x + y\sqrt{D} = \pm(x_1 + y_1\sqrt{D})^k$ for some $k \in \mathbb{Z}$, and fundamental unit $x_1 + y_1\sqrt{D}$ (which is a fundamental solution for the negative Pell equation if it's solvable, and the positive one if it's not). The isomorphism is then given by $x + y\sqrt{D} \to (\varepsilon(x + y\sqrt{D}), k)$ where $\varepsilon(x + y\sqrt{D})$ denotes the sign in this decomposition. That this is a group isomorphism is quite clear. □

When $D > 0$, the unit used to generate all other units is called the *fundamental unit* of $\mathbb{Z}[\sqrt{D}]$. For example, $1 + \sqrt{2}$ has norm -1 in $\mathbb{Z}[\sqrt{2}]$, and is clearly the fundamental unit because the coefficient on $\sqrt{2}$ is minimal among units larger than 1.

To wrap up our discussion on Pell's equation, we'll briefly talk about the connection with *continued fractions*.

Definition 6.4.17. Let $\alpha \in \mathbb{R}$. The **continued fraction expansion** of α is a sequence of integers $\{a_0, a_1, \ldots\}$ such that $\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}}$. Notationally, we denote this continued fraction as $[a_0; a_1, a_2, \ldots]$. The k-th **convergent** of α is given by $[a_0; a_1, \ldots, a_k]$.

Example 6.4.18. We have $2 = [1; 1]$. Let $\varphi = \frac{1+\sqrt{5}}{2}$. Then $\varphi = 1 + \frac{1}{\varphi}$, from which we see that $\varphi = [1; 1, 1, \ldots]$.

That a continued fraction expansion of a real number exists and is unique is a fact we'll take for granted. The idea is rather simple: any *rational* number has a finite continued fraction expansion. Any irrational number can be approximated by rationals, and so one shows that the sequence of continued fractions converges, and so are the convergents of α. However, it's rather technical and not enlightening.

The computation of a continued fraction expansion can be done using an analogous process to the Euclidean algorithm. Since it's not terribly important to us, we won't be very formal, and will describe it through an example.

Example 6.4.19. Suppose that we wished to compute the continued fraction expansion of e. We can do this as follows. Firstly, $e \approx 2.718$, so we can write $e = 2+(e-2) = 2+\frac{1}{1/(e-2)}$. Take $a_0 = 2$ and $\alpha_1 = \frac{1}{e-2}$. Then $\alpha_1 \approx 1.39$, so we can write $\alpha_1 = 1+\frac{1}{1/(1-\alpha_1)}$. We take $a_1 = 1$ and set $\alpha_2 = \frac{1}{1-\alpha_1}$. We have $\alpha_2 \approx 2.54$, so we can write $\alpha_2 = 2+\frac{1}{1/(2-\alpha_2)}$ and we take $\alpha_3 = \frac{1}{2-\alpha_2}$. The first few terms of the continued fraction expansion are $[2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$.

The connection between Pell's equation and continued fractions is as follows. If (x, y) is a solution to $x^2 - Dy^2 = 1$, then $\frac{x}{y}$ is a convergent of $\sqrt{D}$. It's a fact that $\sqrt{D}$ has a *periodic* continued fraction expansion. The period tells us how to find the fundamental solution to the Pell equation.

Theorem 6.4.20. *Let $D > 0$ be square-free, and let m be the period of the continued fraction expansion of $\sqrt{D}$. The fundamental solution to $x^2 - Dy^2 = 1$ is given by*

- (p_{m-1}, q_{m-1}) *where* $\frac{p_{m-1}}{q_{m-1}}$ *is the $m-1$-st convergent of $\sqrt{D}$, if m is even.*
- (p_{2m-1}, q_{2m-1}) *if m is odd.*

Example 6.4.21. One may compute that $\sqrt{13} = [3; \overline{1, 1, 1, 1, 6}]$. The length of this period is 5, which is odd, so the theorem says the fundamental solution to $x^2 - 13y^2 = 1$ is given by $(x, y) = (p_9, q_9) = (649, 180)$. Note that $\sqrt{649 + 180\sqrt{13}} = 18 + 5\sqrt{13}$, so that $(18, 5)$ is the fundamental solution to the *negative* Pell equation $x^2 - 13y^2 = -1$. Observe that this is the 5th convergent!

6.5 Quadratic fields

We'll begin with a general discussion about algebraic numbers. Let $K \subset \mathbb{C}$ be a sub-field.

Definition 6.5.1. A number $\alpha \in \mathbb{C}$ is called **algebraic** if there is a polynomial $p(x) \in \mathbb{Q}[x]$ such that $p(\alpha) = 0$. We call an algebraic number α an **algebraic integer** if there is a monic polynomial $p(x) \in \mathbb{Z}[x]$ with $p(\alpha) = 0$. Notationally, we use $\overline{\mathbb{Q}}$ to denote the set of algebraic numbers, $\overline{K} = \overline{\mathbb{Q}} \cap K$, and $\mathcal{O}_K$ to denote the set of algebraic integers in K.

Definition 6.5.2. The **minimal polynomial** of $\alpha \in \overline{\mathbb{Q}}$, $m_\alpha(x)$, is the monic polynomial of smallest degree such that $m_\alpha(x) \in \mathbb{Q}[x]$ and $m_\alpha(\alpha) = 0$.

It makes sense to speak of *a* minimal polynomial, because by definition, any algebraic number α is killed by *some* polynomial with rational coefficients. By the well-ordering principle, there is then a polynomial that kills α of minimal degree, and rescaling to be monic forces uniqueness. We begin by proving some basic properties of the minimal polynomial.

Proposition 51. *Let $\alpha \in \overline{\mathbb{Q}}$.*

1. *For $f(x) \in \mathbb{Q}[x]$, $m_\alpha(x) \mid f(x)$ in $\mathbb{Q}[x]$ if and only if $f(\alpha) = 0$.*
2. *$m_\alpha(x)$ is irreducible in $\mathbb{Q}[x]$.*
3. *α is an algebraic integer if and only if $m_\alpha(x) \in \mathbb{Z}[x]$.*

Proof.

1. The forward direction is obvious. For the backwards direction, write $f(x) = m_\alpha(x)q(x) + r(x)$ with $r(x) = 0$ or $\deg(r(x)) < \deg(m_\alpha(x))$ by the division algorithm in $\mathbb{Q}[x]$. Plugging in α, we see that $r(\alpha) = 0$. By minimality, this forces $r(x) = 0$, so $m_\alpha(x) \mid f(x)$.
2. For contradiction, suppose that $m_\alpha(x) = g(x)h(x) \in \mathbb{Q}[x]$ is a non-trivial factorization. Without loss of generality, we may assume that $f(x), g(x)$ are monic by rescaling. Plugging in α means that either $f(\alpha) = 0$ or $g(\alpha) = 0$. But this would mean that α is a root of a monic polynomial with rational coefficients of strictly smaller degree, which contradicts the definition of $m_\alpha(x)$. Therefore, $m_\alpha(x)$ is irreducible.
3. The backwards direction is true by definition, so we only prove the forward direction. Suppose that α is an algebraic integer. Then there is some monic $f(x) \in \mathbb{Z}[x]$ such that $f(\alpha) = 0$. By the first part, we can write $f(x) = m_\alpha(x)g(x)$ as a factorization in $\mathbb{Q}[x]$. Since $f(x)$ is *monic*, by Gauss's lemma, this means the factorization actually happens in $\mathbb{Z}[x]$, so $m_\alpha(x) \in \mathbb{Z}[x]$.

□

Example 6.5.3.

- $\sqrt{2}$ is an algebraic integer, because it's minimal polynomial is $x^2 - 2 \in \mathbb{Z}[x]$.
- $\frac{1}{2}$ is an algebraic number, as it's a root of $x - \frac{1}{2} \in \mathbb{Q}[x]$. However, it's *not* an algebraic integer, because it's minimal polynomial is $x - \frac{1}{2} \notin \mathbb{Z}[x]$.
- $\alpha = \sqrt[3]{2} + 1$ is an algebraic integer. We see that $(\alpha - 1)^3 = 2$, and so α is a root of $(x-1)^3 - 2 = x^3 - 3x^2 + 3x - 3 \in \mathbb{Z}[x]$. In fact, $m_\alpha(x) = x^3 - 3x^2 + 3x - 3$: by the rational root theorem, the only possible roots are ± 3, neither of which are roots. If a cubic polynomial was reducible, it would have to have a linear factor, and therefore a root, which means this polynomial is indeed irreducible.

Definition 6.5.4. Let $\alpha \in \overline{\mathbb{Q}}$. We define $\mathbb{Q}(\alpha)$ to be the smallest subfield of $\mathbb{C}$ containing both α and $\mathbb{Q}$.

Proposition 52. *There is an isomorphism of fields* $\mathbb{Q}(\alpha) \cong \mathbb{Q}[x]/(m_\alpha(x))$.

Proof. Define a map $\varphi : \mathbb{Q}[x]/(m_\alpha(x)) \to \mathbb{Q}(\alpha)$ by $p(x) \bmod m_\alpha(x) \mapsto p(\alpha)$. Firstly, we check that this map is well-defined, since it depends on a choice of representative for the equivalence class. If $p(x) \equiv q(x) \bmod m_\alpha(x)$, we may write $q(x) = p(x) + f(x)m_\alpha(x)$ for some $f(x) \in \mathbb{Q}[x]$. Therefore, $\varphi(q(x) \bmod m_\alpha(x)) = q(\alpha) = p(\alpha) + f(\alpha)m_\alpha(\alpha) = p(\alpha) = \varphi(p(x) \bmod m_\alpha(x))$. Next, we observe that φ actually is a homomorphism. This is easy to see, and leave the verification to the reader. Finally, we prove that φ is an isomorphism. If $\varphi(p(x) \bmod m_\alpha(x)) = \varphi(q(x) \bmod m_\alpha(x))$, this would mean $p(\alpha) = q(\alpha)$, so $(p-q)(\alpha) = 0$. This says $m_\alpha(x) \mid (p(x) - q(x))$, which means that $p(x) \equiv q(x) \bmod m_\alpha(x)$, so φ is injective. Finally, we note that $x \bmod m_\alpha(x) \mapsto \alpha$ and $\frac{r}{s} \bmod m_\alpha(x) \mapsto \frac{r}{s}$. Since $\mathbb{Q}[x]/(m_\alpha(x))$ is a field because $m_\alpha(x)$ is irreducible in $\mathbb{Q}[x]$, this means its image under φ is isomorphic to a subfield of $\mathbb{Q}(\alpha)$ containing $\mathbb{Q}$ and α, and therefore *equals* $\mathbb{Q}(\alpha)$ by definition. □

Corollary 6.5.5. *As a set,* $\mathbb{Q}(\alpha) = \{a_0 + \ldots + a_{d-1}\alpha^{d-1} : a_i \in \mathbb{Q}\}$, *where* $d = \deg(m_\alpha(x))$.

Proof. This immediately follows from the above isomorphism: any element of $\mathbb{Q}[x]/(m_\alpha(x))$ has a representative that is a polynomial with rational coefficients of degree strictly smaller than d. □

For some concrete examples, any element of $\mathbb{Q}(i)$ is of the form $a + bi$ for some $a, b \in \mathbb{Q}$. Any element of $\mathbb{Q}(\sqrt[3]{2})$ is of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, for some $a, b, c \in \mathbb{Q}$. Another way of thinking about this result is that the powers $1, \alpha, \ldots, \alpha^{d-1}$ of α form a $\mathbb{Q}$-basis of $\mathbb{Q}(\alpha)$ as a $\mathbb{Q}$-vector space.

Fields of the form $\mathbb{Q}(\alpha)$ for some α algebraic are called *number fields*. They're the main object of study in *algebraic number theory*. A common misconception is that this branch of number theory is "using algebra to study number theory". However, as not all number fields are *real* subfields of $\mathbb{C}$, one naturally needs to bring in tools from complex analysis to understand the full picture. Although not yet clear, the algebraic integers of a number field play a role analogous to the role that $\mathbb{Z}$ plays inside of $\mathbb{Q}$.

Proposition 53. $\mathcal{O}_\mathbb{Q} = \mathbb{Z}$.

Proof. Suppose that $\alpha \in \mathbb{Q}$ is an algebraic integer. Write $\alpha = \frac{r}{s}$ with $\gcd(r, s) = 1$. By definition, there is some monic $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0 \in \mathbb{Z}[x]$ with $f(\alpha) = 0$. Plugging in, this means $(\frac{r}{s})^n + \ldots + a_0 = 0$, so clearing denominators yields $r^n + a_{n-1}r^{n-1}s + \ldots + s^n = 0$. This means $r^n = -s(a_{n-1}r^{n-1} + \ldots + s^{n-1})$, so $s \mid r^n$. Since $\gcd(r, s) = 1$, this means $s = \pm 1$, so $\alpha \in \mathbb{Z}$. Obviously any integer is an algebraic integer, so we're done. □

This result may seem silly, but it's actually very important. It allows us to adjust the way we think about the subject! So far, we've really been considering the *integers* to be our main object of study. Instead, we can now think of the *rationals* as what we're really trying to understand. In particular, that means understanding the special subring $\mathcal{O}_\mathbb{Q} = \mathbb{Z}$! The

main power of this viewpoint is that by studying number *fields*, powerful tools from field theory and Galois theory can be used.

For our purposes though, we're just going to focus on *quadratic fields*, fields of the form $K = \mathbb{Q}(\sqrt{D})$ for D a square-free integer. For any $\alpha = a + b\sqrt{D} \in K$, we note that the minimal polynomial of α is $m_\alpha(x) = x^2 - 2ax + (a^2 - Db^2) \in \mathbb{Q}[x]$, because its roots α and $\bar{\alpha}$ are not rational (assuming $b \neq 0$). If $\alpha \in \mathbb{Z}[\sqrt{D}]$, then both coefficients are integers, and therefore the earlier proposition tells us that $\mathbb{Z}[\sqrt{D}] \subset \mathcal{O}_K$. In general though, $\mathcal{O}_K$ is larger than $\mathbb{Z}[\sqrt{D}]$. To see this, note that $\frac{1+\sqrt{5}}{2} \in \mathbb{Q}(\sqrt{5})$ has minimal polynomial $x^2 - x - 1 \in \mathbb{Z}[x]$, and therefore $\mathbb{Z}[\sqrt{5}] \subsetneq \mathcal{O}_K$.

Theorem 6.5.6. *Let D be a square-free integer, and let $K = \mathbb{Q}(\sqrt{D})$. Then*

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\frac{1+\sqrt{D}}{2}] & D \equiv 1 \bmod 4 \\ \mathbb{Z}[\sqrt{D}] & D \equiv 2, 3 \bmod 4 \end{cases}$$

Proof. Let $\alpha \in \mathcal{O}_K$. Write $\alpha = a + b\sqrt{D}$ for some $a, b \in \mathbb{Q}$. As observed above, $m_\alpha(x) = x^2 - 2ax + (a^2 - Db^2)$. Therefore, we require $2a, a^2 - Db^2 \in \mathbb{Z}$. This means that $a \in \frac{1}{2}\mathbb{Z}$, so write $a = \frac{c}{2}$ for some c.

- If c is even, then $a \in \mathbb{Z}$ so from $a^2 - Db^2 \in \mathbb{Z}$, this means that $Db^2 \in \mathbb{Z}$. Writing $b = \frac{r}{s}$ for $\gcd(r, s) = 1$, this means $Dr^2 = s^2 k$ for some k. This means $s^2 \mid D$ because $\gcd(s^2, r^2) = 1$, and because D is square-free, $s^2 = 1$ means $s = \pm 1$. Therefore, $b \in \mathbb{Z}$, so $\alpha \in \mathbb{Z}[\sqrt{D}]$ and therefore $\mathcal{O}_K = \mathbb{Z}[\sqrt{D}]$.
- If c is odd, then $\frac{c^2 - 4Db^2}{4} \in \mathbb{Z}$, so we can write $c^2 - 4Db^2 = 4k$ for some k. This means that $d(2b)^2 \in \mathbb{Z}$, so by the above argument this means $2b \in \mathbb{Z}$. Therefore, we can write $2b = \ell$ for some ℓ. As $c^2 - D\ell^2 = 4k$, because c is odd, working mod 4 means $1 \equiv D\ell^2 \bmod 4$. This tells us that $D \equiv 1 \bmod 4$, and ℓ is odd. Returning to α, we have $\alpha = a + b\sqrt{D} = \frac{c}{2} + \frac{\ell}{2}\sqrt{D} = \frac{c-\ell}{2} + \ell(\frac{1+\sqrt{D}}{2}) \in \mathbb{Z}[\frac{1+\sqrt{D}}{2}]$.

Therefore, we've shown that if $D \equiv 2, 3 \bmod 4$ then c is even, which means $\mathcal{O}_K = \mathbb{Z}[\sqrt{D}]$. If $D \equiv 1 \bmod 4$, we show that $\mathbb{Z}[\frac{1+\sqrt{D}}{2}] \subset \mathcal{O}_K$, which then would mean c is odd, and give us the other inclusion. To see this, let $\alpha = a + b(\frac{1+\sqrt{D}}{2}) \in \mathbb{Z}[\frac{1+\sqrt{D}}{2}]$. Then writing $\alpha = (a + \frac{b}{2}) + \frac{b}{2}\sqrt{D}$, we may compute that the minimal polynomial of α is $m_\alpha(x) = x^2 - (2a + b)x + (a^2 + ab + b^2(\frac{1-D}{4}) \in \mathbb{Z}[x]$, so we're done. □

In particular, we see that $\mathcal{O}_K$ for $K = \mathbb{Q}(\sqrt{D})$ is a *ring*. It turns out this is true for any number field, although the proof is more complicated. The ring $\mathcal{O}_K$ is supposed to be an object inside of K that plays a role similar to that of $\mathbb{Z}$ inside of $\mathbb{Q}$. For example, the subring $\mathbb{Z}[\sqrt{5}]$ of $\mathbb{Q}(\sqrt{5})$ does not have unique factorization, because $(3 + \sqrt{5})(3 - \sqrt{5}) = 4 = 2 \cdot 2$ are two genuinely different irreducible factorizations of 4. However, it turns out that the ring of integers $\mathbb{Z}[\frac{1+\sqrt{5}}{2}]$ *does* have unique factorization! This example is a bit disingenuous, though, as unfortunately $\mathcal{O}_K$ still does *not* have unique factorization. For example, $\mathbb{Z}[\sqrt{-5}]$ is the ring of integers in $\mathbb{Q}(\sqrt{-5})$, but $\mathbb{Z}[\sqrt{-5}]$ does not have unique factorization (see exercise

2.4.9).

For quadratic rings, we know the following: if $D < 0$, there precisely 9 values of D for which $\mathcal{O}_K$ has unique factorization: these are $D = -1, -2, -3, -7, -11, -19, -43, -67, -163$. This is known as the *Heegner-Stark theorem*. For $D > 0$, we know know even less! It was conjectured by Gauss that there are infinitely many values of D for which $\mathcal{O}_K$ has unique factorization. Even today, this is still an open problem. In fact, we expect this to hold $\approx 76\%$ of the time for prime values of D!

It's often times easy to show that a quadratic ring does *not* have unique factorization.

Proposition 54. *Let D be squarefree and suppose that $\mathbb{Z}[\sqrt{D}]$ has unique factorization. Then for prime $p \in \mathbb{Z}$, the following are equivalent:*

1. *p factors in $\mathbb{Z}[\sqrt{D}]$.*
2. *$\pm p = x^2 - Dy^2$ for some $x, y \in \mathbb{Z}$.*
3. *$D \equiv \square \bmod p$.*

Proof. If $p = \alpha\beta$ in $\mathbb{Z}[\sqrt{D}]$, then taking norms means $p^2 = N(\alpha)N(\beta)$. If the factorization is non-trivial, this means either $p = N(\alpha)$ or $-p = N(\alpha)$, which means $\pm p = x^2 - Dy^2$. If $\pm p = x^2 - Dy^2$, then working mod p says $x^2 - Dy^2 \equiv 0 \bmod p$. Since $x, y \not\equiv 0 \bmod p$, this means $(x/y)^2 \equiv D \bmod p$. Finally, If $D \equiv \square \bmod p$, there is some x such that $p \mid x^2 - D$ in $\mathbb{Z}$. In $\mathbb{Z}[\sqrt{D}]$, this means $p \mid (x - \sqrt{D})(x + \sqrt{D})$. Note that p is not prime in $\mathbb{Z}[\sqrt{D}]$, because if it were, then this would mean $p \mid x \pm \sqrt{D}$ so that $p \mid \pm 1$, which is impossible. Therefore, as $\mathbb{Z}[\sqrt{D}]$ has unique factorization, this means that p is reducible, so there is non-trivial factorization $p = \alpha\beta$ in $\mathbb{Z}[\sqrt{D}]$. □

Example 6.5.7. The ring $\mathbb{Z}[\sqrt{10}]$ does not have unique factorization. Note that $10 \equiv 0 \bmod 2$, so by the above result, this would mean that either $2 = x^2 - 10y^2$ or $-2 = x^2 - 10y^2$ is solvable. However, this would mean that $x^2 \equiv \pm 2 \bmod 5$ is solvable, which is a contradiction. We are able to show this *abstractly*, without exhibiting any different factorizations!

Although $\mathcal{O}_K$ does not generally have unique factorization of *elements*, it remarkably turns out that it has unique factorization of *ideals*! In fact, historically the term *ideal* comes from number theory, not algebra. Kummer observed that $\mathcal{O}_K$ does not always have unique factorization. In an attempt to explain (and fix) this failure, Kummer's idea was that there were so called "ideal numbers" for which unique factorization would hold. Dedekind later fleshed out this idea and it led to the definition of ideals for rings.

Here are some facts about ideals in $\mathcal{O}_K$ which we'll use to understand how this works.

- Any ideal $I \subset \mathcal{O}_K$ is finitely generated, with at most two generators. Concretely, $I = (\alpha)$ or $I = (\alpha, \beta)$ for some $\alpha, \beta \in \mathcal{O}_K$.

- For two ideals $I, J \subset \mathcal{O}_K$, the product IJ is defined by $IJ = \{\sum_i a_i b_i : a_i \in I, b_i \in J\}$, the set of finite sums of products of elements of I and J. If we write $I = (\alpha_1, \alpha_2)$ and $J = (\beta_1, \beta_2)$, then $IJ = (\alpha_1\beta_1, \alpha_1\beta_2, \alpha_2\beta_1, \alpha_2\beta_2)$, which by fact 1 can be written as an ideal generated by two elements.

Example 6.5.8. For $K = \mathbb{Q}(\sqrt{-14})$, we have $\mathcal{O}_K = \mathbb{Z}[\sqrt{-14}]$ because $-14 \equiv 2 \bmod 4$. Let $I = (5 + \sqrt{-14}, 2 + \sqrt{-14})$ and $J = (4 + \sqrt{-14}, 2 - \sqrt{-14})$. Then $IJ = (6 + 9\sqrt{-14}, -6 + 6\sqrt{-14}, 24 - 3\sqrt{-14}, 18)$. Observe that $IJ = (6, 3\sqrt{-14})$: that $IJ \subset (6, 3\sqrt{-14})$ is clear, and to see the other containment, note that adding the first two generators shows that $15\sqrt{-14} \in IJ$, and multiplying the third by 3 and adding 18 shows that $18\sqrt{-14} \in IJ$, so $3\sqrt{-14} \in IJ$. Finally, note adding the second generator to twice the third says $42 \in IJ$, so $\gcd(18, 42) = 6 \in IJ$, which says $IJ = (6, \sqrt{-14})$.

Definition 6.5.9. Let I, J be ideals of $\mathcal{O}_K$. We say that I **divides** J if there is another ideal A such that $J = IA$. We say that an ideal $\mathfrak{p}$ is **prime** if $\mathfrak{p} \mid IJ$ means $\mathfrak{p} \mid I$ or $\mathfrak{p} \mid J$.

To the reader familiar with abstract algebra, this might seen different than the "usual" definition of a prime ideal. Most algebra textbooks would call $\mathfrak{p}$ prime if $ab \in \mathfrak{p}$ means either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. For the ring of integers $\mathcal{O}_K$, these definitions are *equivalent*. This relies on two facts:

- For any commutative ring R, an ideal P is prime if and only if for any ideals $I, J \subset R$ with $IJ \subset P$, either $I \subset P$ or $J \subset P$.
- In $\mathcal{O}_K$, $I \mid J$ if and only if $I \supset J$.

This latter property is rather special: this is *not* true for most rings! Now that we have a definition of divisibility of ideals and a notion of what a prime ideal is, we can state what we mean precisely by unique factorization.

Theorem 6.5.10. *Let I be an ideal of $\mathcal{O}_K$. Then there exist unique prime ideals $\mathfrak{p}_1, \dots \mathfrak{p}_k$ and integers $e_1, \dots, e_k \geq 1$ such that $I = \mathfrak{p}^{e_1} \cdots \mathfrak{p}_k^{e_k}$.*

We won't prove this theorem, as we would need to develop some theory in order to do so. The idea, though, is define a notion of a *norm* of an ideal, as a way of measuring its size. The proof then proceeds analogously to the proof in theorem 2.3.8 by induction on the norm of an ideal.

Example 6.5.11. Let $K = \mathbb{Q}(\sqrt{-5})$, which has $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$. The ring $\mathcal{O}_K$ does not have unique factorization: $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ are two genuinely different factorizations of 6 that do not differ by a unit. As *ideals*, though, we have the unique factorization $(6) = (2, 1+\sqrt{-5})^2(3, 1+\sqrt{-5})(3, 1-\sqrt{-5})$. That these ideals are indeed prime can be verified by using the result from abstract algebra than an ideal P of a ring R is prime if and only if the quotient ring R/P is a domain. For example, to compute $\mathbb{Z}[\sqrt{-5}]/(2, 1 + \sqrt{-5})$, in this quotient we have $\sqrt{-5} = -1$ and $2 = 0$, so there's an isomorphism of rings $\mathbb{Z}[\sqrt{-5}]/(2, 1 + \sqrt{-5}) \cong \mathbb{Z}/2\mathbb{Z}$ given by $a + b\sqrt{-5} \bmod (2, 1 + \sqrt{-5}) \mapsto a - b \bmod 2$.

The obstruction of when $\mathcal{O}_K$ has unique factorization is the existence of *non-principal* ideals:

Theorem 6.5.12. *$\mathcal{O}_K$ has unique factorization if and only if every ideal in $\mathcal{O}_K$ is principal.*

We'll use this theorem to give another proof that $\mathbb{Z}[\sqrt{-5}]$ does not have unique factorization.

Example 6.5.13. The ideal $(2, \sqrt{-5})$ of $\mathbb{Z}[\sqrt{-5}]$ is not principal. Suppose it were, then we could write $(2, \sqrt{-5}) = (\alpha)$ for some $\alpha \in \mathbb{Z}[\sqrt{-5}]$. Since $2 \in (2, \sqrt{-5})$, this means $2 \in (\alpha)$, so $2 \mid \alpha$ in $\mathbb{Z}[\sqrt{-5}]$. Write $2 = \alpha\beta$ for some β. Taking norms would mean $4 = N(\alpha)N(\beta)$, so $N(\alpha) \mid 4$. Similarly, $\sqrt{-5} \in (\alpha)$ so $N(\alpha) \mid 5$. Therefore, $N(\alpha) \mid \gcd(4, 5) = 1$, so $N(\alpha) = 1$ and therefore α is a unit. This would mean that $\mathbb{Z}[\sqrt{-5}] = (2, \sqrt{-5})$. If this were true, then $\mathbb{Z}[\sqrt{-5}]/(2, \sqrt{-5})$ would be the 0 ring, but $\mathbb{Z}[\sqrt{-5}]/(2, \sqrt{-5}) \cong \mathbb{Z}/2\mathbb{Z}$ as we saw before, which yields a contradiction.

Finally, we close with the factorization of primes in $\mathcal{O}_K$ for $K = \mathbb{Q}(\sqrt{D})$.

Theorem 6.5.14. *Let $p \in \mathbb{Z}$ be a prime. Set $d = D$ if $D \equiv 1 \bmod 4$, or $d = 4D$ if $D \equiv 2, 3 \bmod 4$. Then the way (p) factors in $\mathcal{O}_K$ for $K = \mathbb{Q}(\sqrt{D})$ is:*

- *$(p) = \mathfrak{p}^2$ for some prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$ if and only if $p \mid d$. In this case, we say that (p) is ramified.*
- *$(p) = \mathfrak{p}\overline{\mathfrak{p}}$ for some prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$ if and only if $(\frac{d}{p}) = 1$. In this case, we say that (p) splits.*
- *(p) is prime in $\mathcal{O}_K$ if and only if $(\frac{d}{p}) = -1$. In this case, we say that (p) is intert.*

Here, by $\overline{\mathfrak{p}}$ we mean the *conjugate* ideal of $\mathfrak{p}$. If $\mathfrak{p} = (\alpha, \beta)$ for $\alpha, \beta \in \mathcal{O}_K$, we define $\mathfrak{p} = (\overline{\alpha}, \overline{\beta})$. The number d in the theorem is called the *discriminant* of K.

Note that $(\frac{d}{p}) = (\frac{D}{p})$ because $(\frac{4}{p}) = 1$ is always true. One may use quadratic reciprocity to aid in the computation of the Legendre symbol, and ultimately, will conclude that the factorization type of (p) is governed by *congruence classes* of $p \bmod d$.

Example 6.5.15. Let $K = \mathbb{Q}(\sqrt{3})$, so the discriminant is $d = 12$. The primes which ramify are $p = 2, 3$. The primes which split are those for which $(\frac{12}{p}) = (\frac{3}{p}) = 1$, which is given by $p \equiv \pm 1 \bmod 12$. The primes which are inert are those for which $(\frac{3}{p}) = -1$, which is given by $p \equiv \pm 5 \bmod 12$.

In the above example, for a random prime p it should be "equally likely" that (p) splits in $\mathbb{Z}[\sqrt{3}]$ and (p) is inert in $\mathbb{Z}[\sqrt{3}]$, because the probability that p lies in the congruence classes $\pm 1 \bmod 12$ should be "the same" as p lying in the congruence classes $p \equiv \pm 5 \bmod 12$. Since any prime $p \neq 2, 3$ must satisfy $p \equiv \pm 1, \pm 5 \bmod 12$, this means about half the primes split in $\mathbb{Z}[\sqrt{3}]$ and half the primes stay inert in $\mathbb{Z}[\sqrt{3}]$.

This is a consequence of the *prime number theorem for arithmetic progressions*. For any integer n, all but finitely many primes must lie in some congruence class $a \bmod n$ with $\gcd(a, n) = 1$. Dirichlet's theorem says that each of these congruence classes $a \bmod n$ contains infinitely many primes, and the prime number theorem for arithmetic progressions says

that the primes are "equidistributed" among the $\varphi(n)$ congruence classes of interest.

For a general quadratic field $K = \mathbb{Q}(\sqrt{D})$, the same story is going to be true. For a general number field K, there are finitely many "splitting types" of the ideal (p) in $\mathcal{O}_K$. The proportion of primes that fall into a certain splitting type is completely understood, and described in the *Chebotarev density theorem*, which is, essentially, a generalization of Dirichlet's theorem. The Chebotarev density theorem is one of the groundbreaking results of the early 1900s. The idea that *reciprocity laws* correspond to *splitting of primes* was one of the major ideas that went into the development of *class field theory*.

6.6 Exercises

1. (a) Find all rational points on the ellipse $x^2 - xy + y^2 = 1$.
   (b) Find all *integer* points on the ellipse $x^2 - xy + y^2 = 1$.

2. Can two (distinct) perfect squares average to be a perfect square? In other words, are there integers $0 < x < y < z$ such that $\frac{x^2+z^2}{2} = y^2$? If no, provide a proof. If so, find all such solutions.

3. Let p be an odd prime.
   (a) Prove that for any $a \in \mathbb{F}_p$, $x^2 + y^2 = a$ is solvable in $(\mathbb{F}_p)^2$. (*One approach: show the sets* $\{x^2 : x \in \mathbb{F}_p\}$ *and* $\{a - x^2 : x \in \mathbb{F}_p\}$ *overlap*).
   (b) How many solutions in $(\mathbb{F}_p)^2$ are there to $x^2 + y^2 = 1$? *(Does the usual approach still work?)*

4. Let p be an odd prime. Use Minkowski's theorem to prove that $x^2 + 2y^2 = p$ has solutions if and only if $p \equiv 1, 3 \bmod 8$.

5. For prime p, the solvability of $x^2+y^2 = p$ in $\mathbb{Z}^2$ was related to the arithmetic of the ring $\mathbb{Z}[i]$: integer points on this circle correspond to factorizations $p = \alpha\bar{\alpha}$ for $\alpha = x + yi$ in $\mathbb{Z}[i]$. In this problem, you'll see how the solvability of $x^2 + 2y^2 = p$ is related to the arithmetic of $\mathbb{Z}[\sqrt{-2}]$.
   (a) Modifying the argument that was used to show $\mathbb{Z}[i]$ has a division algorithm, state and prove a division algorithm for $\mathbb{Z}[\sqrt{-2}]$.
   (b) Once you know that $\mathbb{Z}[\sqrt{-2}]$ has a division algorithm, the usual chain of reasoning will show that it has unique factorization. Modify the proof of when p is a sum of two squares to show that $x^2 + 2y^2 = p$ has solutions if and only if $p = 2$ or $p \equiv 1, 3 \bmod 8$.
   (c) For prime $p \in \mathbb{Z}$, list (with proof!) the possible factorization types of $p \in \mathbb{Z}[\sqrt{-2}]$.

6. Find the fundamental unit of $\mathbb{Z}[\sqrt{6}]$, and then give a description of all units in $\mathbb{Z}[\sqrt{6}]$.

7. Find all integer solutions to $x^2 - 10y^2 = -1$.

8. Observe that $(6, 1)$ is a solution to $x^2 - 10y^2 = 26$. Find two *other* positive solutions.

9. (a) Suppose that (x, y) is a solution to $x^2 - Dy^2 = 1$. Show that $|\frac{x}{y} - \sqrt{D}| < \frac{1}{y^2}$.

   (b) Find the fundamental solution to $x^2 - 101y^2 = 1$.

   (c) Use the previous part to find a rational number $\frac{x}{y}$ such that $|\sqrt{101} - \frac{x}{y}| < 10^{-12}$. Use this to write down an approximation to $\sqrt{101}$.

10. (a) Observe that $x^2 - 80y^2 = 1$ has a positive solution by inspection. Hence, prove that $x^2 - 80y^2 = -1$ has no integer solutions. Generalize this argument to show that $x^2 - (n^2 - 1)y^2 = -1$ has no integer solutions.

    (b) Let $n \geq 2$ be an integer. Prove that the fundamental unit of $\mathbb{Z}[\sqrt{n^2 - 1}]$ is $n + \sqrt{n^2 - 1}$.

11. (a) Let $p \equiv 3 \bmod 4$ be a prime. Prove that $x^2 - py^2 = -1$ is not solvable.

    (b) Let $p \equiv 1 \bmod 4$ be a prime, and let $a + b\sqrt{p} > 1$ correspond to the fundamental solution to $x^2 - py^2 = 1$. Prove that a is odd, b is even, and $\gcd(a + 1, a - 1) = 2$.

    (c) From $pb^2 = a^2 - 1 = (a-1)(a+1)$, deduce that either $a+1 = 2pu^2$ and $a-1 = 2v^2$ or $a + 1 = 2u^2$ and $a - 1 = 2pv^2$ for some integers (u, v).

    (d) Show the latter case cannot actually happen, so that $a+1 = 2pu^2$ and $a-1 = 2v^2$. Deduce that $v^2 - pu^2 = -1$, so that the equation $x^2 - py^2 = -1$ *does* have a solution.

12. The n-th *triangular number* T_n is defined by $T_n = 1+2+\ldots+n = \frac{n(n+1)}{2}$. The numbers T_n count the number of dots in the triangular arrangement with n dots on each side. Prove there are infinitely many pairs of integers (n, k) such that $T_n = k^2$. (*Hint: write down an equation describing such pairs and transform it into a Pell equation*).

13. Find (with proof) the minimal polynomial of $\sqrt{2} + \sqrt{3}$.

Appendix A

Induction

The most common statement of mathematical induction may be stated as follows:

Theorem A.0.1. *(Principle of Mathematical Induction) For $n \in \mathbb{N}$, let $P(n)$ be a statement such that*

1. *$P(n_0)$ is true for some n_0*
2. *$P(k)$ is true implies $P(k+1)$ is true for all $k \geq n_0$.*

Then $P(n)$ is true for all $n \geq n_0$.

There is also a "stronger" version of induction:

Theorem A.0.2. *(Principle of Strong Induction) For $n \in \mathbb{N}$, let $P(n)$ be a statement such that*

1. *$P(n_0)$ is true for some n_0*
2. *$P(n_0), \ldots, P(k)$ is true implies $P(k+1)$ is true for all $k \geq n_0$.*

Then $P(n)$ is true for all $n \geq n_0$.

It's not terribly hard to show that these two forms of induction are equivalent to each other. More interestingly, is that both forms of induction are equivalent to the well-ordering principle!

Theorem A.0.3. *The principle of strong induction is equivalent to the well-ordering principle.*

Proof. Suppose that the principle of strong induction holds. Let $S \subset \mathbb{Z}^+$ be a non-empty subset. We wish to show that S has a least positive element. For sake of contradiction, suppose that S does not have a least positive element. Then $1 \notin S$, because 1 is the smallest positive integer. From this we see that $2 \notin S$, because $1 \notin S$ and 2 is the next positive integer after 1. Continuing this train of thought, we see that if $1, 2, \ldots, k \notin S$ for some k, then we must have $k+1 \notin S$. By induction, we must then have $n \notin S$ for all$n \geq 1$, which says that S is empty, a contradiction. Therefore, if we assume strong induction holds, then

the well-ordering principle holds.

Now suppose that the well-ordering principle holds, and let P be a statement about integers such that $P(n_0)$ is true for some n_0 and $P(n_0), \ldots, P(k)$ true implies that $P(k+1)$ is true for all $k \geq n_0$. We wish to show that $P(n)$ is true for all $n \geq n_0$. Suppose otherwise, that there is some $m \geq n_0$ such that $P(m)$ is false. Let $S = \{n \in \mathbb{Z}^+ : P(n) \text{ is false}\}$. By assumption S is non-empty, so by the well-ordering principle, S has a smallest positive element, say k. Since $P(n_0)$ is true, we must have that $k > n_0$. Now by definition of k, $P(k-1)$ must be true. Similarly, $P(n_0), P(n_0+1), \ldots, P(k-1)$ must all be true. By strong induction, this then says that $P(k)$ is true, which is a contradiction. Therefore if the well-ordering principle holds, then strong induction holds, so we are done. $\square$

We have shown that induction and the well-ordering principle are equivalent, but we haven't shown that either one of these statements are actually true. In fact, we can't! Any construction of the integers (e.g. the Peano construction) must take either the well-ordering principle or mathematical induction as an axiom.

Appendix B

Algebraic Structures

B.1 Groups

Definition B.1.1. A **group** is a pair $(G, \cdot)$ for a non-empty set G and binary operation $\cdot$ that satisfy the following properties:

1. For any $a, b, c \in G$, we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
2. There exists $e \in G$ such that for any $a \in G$, $a \cdot e = e \cdot a = a$.
3. For all $a \in G$, there exists $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$.

The element e is called the **identity element** of the group G, and a^{-1} is called the **inverse** of the element a. Furthermore, if G is a group and satisfies the additional property that for all $a, b \in G$ we have $a \cdot b = b \cdot a$, we say that G is **abelian**. When G is abelian, it is customary to write the operation on G as "+" instead of "$\cdot$" (unless working with a specific example of a group).

In other words, a group is a set where you can "multiply" elements together in some way. The axioms for a group are rather loose, and so the notion of "multiplication" in a group can be rather abstract.

Example B.1.2. The trivial group $\{0\}$ defined by the operation $0 \cdot 0 = 0$. This is an abelian group.

Example B.1.3. The pair $(\mathbb{Z}, +)$ with the usual addition of integers form an abelian group. The identity element is 0, and the inverse of a number $a \in \mathbb{Z}$ is the negative number $-a$ because $a + (-a) = 0$. However, $(\mathbb{Z}, \cdot)$ with the operation of multiplication does *not* form a group. Although there is an identity element with respect to this operation (the number 1), not all integers have an inverse with respect to $\cdot$. For example, there is no integer a such that $2a = 1$, so 2 has no inverse with respect to $\cdot$.

Example B.1.4. The pairs $(\mathbb{Q}, +)$ and $(\mathbb{Q} \setminus \{0\}, \cdot)$ are with the usual operations of addition and multiplication respectively are abelain groups.

Example B.1.5. The pair $(M_2(\mathbb{R}), +)$ of 2×2 matrices with real number entries is an abelain group. The pair $(\mathrm{SL}_2(\mathbb{R}), \cdot)$ of 2×2 matrices of determinant 1 is a group under multiplication of matrices. The identity element is the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and the fact that this forms a group is because from linear algebra, $\det(AB) = \det(A)\det(B)$, and therefore the product of two matrices of determinant 1 has determinant 1. However, this group is *not* an abelian, because matrix multiplication is generally not commutative. As an explicit counterexample, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, while $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

Example B.1.6. Let X be a non-empty set, and let $\mathrm{Sym}(X)$ denote the set of bijections on X. That is, an element of $\mathrm{Sym}(X)$ is a bijection $f : X \to X$. Then $(\mathrm{Sym}(X), \circ)$ is a group under composition of functions. The identity element of $\mathrm{Sym}(X)$ is the identity map id_X defined by $\mathrm{id}_X(x) = x$ for all $x \in X$. The inverse of $f \in \mathrm{Sym}(X)$ is the inverse map $f^{-1} : X \to X$, which necessarily exists as f is a bijection. The group $\mathrm{Sym}(X)$ is called the **symmetric group on** X.

Definition B.1.7. Let $(G, \cdot_G)$ and $(H, \cdot_H)$ be groups. The **product** of G and H, $G \times H$ is the set $G \times H = \{(g, h) : g \in G, H \in H\}$ with group operation $\cdot$ given by $(g, h) \cdot (g', h') = (g \cdot_G g', h \cdot_H h')$ for (g, h) and $(g', h') \in G \times H$. The identity element is given by (e_G, e_H).

Definition B.1.8. A **subgroup** of a group $(G, \cdot)$ is a subset $H \subset G$ such that $(H, \cdot)$ is also a group, and the identity element of H is the same as the identity element of G.

Example B.1.9. The set of even integers $2\mathbb{Z}$ is a subgroup of $(\mathbb{Z}, +)$. The integers $\mathbb{Z}$ are a subgroup of $(\mathbb{Q}, +)$ and also of $(\mathbb{R}, +)$. The group $\mathrm{SL}_2(\mathbb{R})$ is a subgroup of $(M_2(\mathbb{R}), \cdot)$.

B.2 Rings and fields

Definition B.2.1. A **ring** is a pair $(R, +, \cdot)$ consisting of a non-empty set R with binary operations $+, \cdot$ called addition and multiplication that satisfy the following axioms for all $a, b, c \in R$:

1. $a + b = b + a$.
2. $a + (b + c) = (a + b) + c$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
3. $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$.
4. There is an element $0 \in R$ with the property $a + 0 = a$.
5. For any a, there is an element $-a \in R$ such that $a + (-a) = 0$.
6. There is an element $1 \in R$ such that $a \cdot 1 = a$.

Note that the definition of a ring does *not* require that multiplication be commutative. A ring that satisfies $ab = ba$ for all $a, b \in R$ is called a **commutative ring**.

Some algebra textbooks do not require that a ring have a multiplicative identity, and instead call our definition a "ring with identity". This is very bad – for various reasons, it ends up being better to think of not having an identity element as something *missing* from a ring instead of something *added* to a ring. There are a few arguments for not including a multiplicative identity as part of the definition of a ring, but at the end of the day, objects that behave like rings without identity are much better labeled under different terms.

Example B.2.2. The zero ring is a set $\{0\}$ with operations defined by $0+0=0$ and $0\cdot 0=0$. This satisfies the axioms for a commutative ring, with additive and multiplicative identity both given by 0.

Example B.2.3. The integers $\mathbb{Z}$ are the prototypical example of a ring. Other familiar examples: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$. All these rings are commutative.

Example B.2.4. $M_2(\mathbb{R})$ is a ring with operations of matrix addition and matrix multiplication. This ring is not commutative, as can be seen by the example in the previous section.

Example B.2.5. Let R be a ring, and let $R[x]$ denote the set of polynomials with coefficients in R in the indeterminate variable x. An element $p(x) \in R[x]$ looks like $p(x) = a_0 + a_1 x + \ldots + a_n x^n$ for some $n \geq 0$ and $a_i \in R$. Then $R[x]$ forms a commutative ring with operations of polynomial addition and polynomial multiplication. These operations are the same as the ones you learn in middle school. For example, in $\mathbb{Z}[x]$, one has $(2x+1)+(x^2+2) = x^2+2x+3$ and $(2x+1)(x^2+2) = 2x^3+x^2+4x+2$. In $R[x]$ the additive identity is the constant polynomial $p(x) = 0$ and the multiplicative identity is the constant polynomial $p(x) = 1$.

For any ring R, one may "forget" the multiplicative structure to produce a group $(R, +)$ with respect to the operation of addition. Can one similarly "forget" the additive structure to be left with a group $(R, \cdot)$? Not quite, because the axioms of a ring do not require that all elements have a multiplicative inverse. However, a special subset of R will have this property.

Definition B.2.6. Let R be a ring. An element $r \in R$ is called a **unit** of R if it has a multiplicative inverse. That is, there exists $s \in R$ such that $rs = sr = 1$. If r is a unit, we denote its inverse as r^{-1}. The set of units of R is denoted $R^\times$, and $(R^\times, \cdot)$ forms a group called the **unit group** of R.

Example B.2.7. $\mathbb{Z}^\times = \{\pm 1\}$, because the only integer solutions to the equation $xy = 1$ are $x = y = 1$ or $x = y = -1$.

Example B.2.8. $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\}$, because all non-zero rational numbers have a multiplicative inverse: if $\frac{a}{b} \neq 0$, then $(\frac{a}{b})^{-1} = \frac{b}{a}$ because $\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = 1$.

Definition B.2.9. Let $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$ be rings. The **product ring** $R \times S$ is the set $R \times S = \{(r, s) : r \in R, s \in S\}$ equipped with a ring structure given by $(r, s) + (r', s') = (r +_R r', s +_S s')$ and $(r, s) \cdot (r', s') = (r \cdot_R r', s \cdot_S s')$ for (r, s) and $(r', s') \in R \times S$. The identity elements for $+$ and $\cdot$ are $(0_R, 0_S)$ and $(1_R, 1_S)$ respectively.

Definition B.2.10. A **subring** of a ring $(R, +, \cdot)$ is a subset $S \subset R$ such that $(S, +, \cdot)$ forms a ring with the same identity elements 0 and 1.

Example B.2.11. $\mathbb{Z}$ is a subring of $\mathbb{Q}$. It's also a subring of $\mathbb{R}$, and $\mathbb{C}$.

Example B.2.12. The set of even integers $2\mathbb{Z}$ is a *not* a subring of $\mathbb{Z}$. Although the sum and product of two even integers is still even, 1 is not an even number, so $2\mathbb{Z}$ does not contain the multiplicative identity element of $\mathbb{Z}$ (and in fact, does not even *have* a multiplicative identity element!)

Definition B.2.13. An **ideal** of a commutative ring $(R, +, \cdot)$ is a subgroup $(I, +)$ of $(R, +)$ such that for any $r \in R$ and $x \in I$, one has $rx \in I$. In other words, an ideal of R is an additive subgroup of R that absorbs multiplication by elements of R.

Note: we restricted the definition of ideals to commutative rings. For a general ring R, one has notions of *left ideals* and *right ideals*, where the multiplication by elements of R happen on the left or right respectively. For a commutative ring, there is no difference between these notions. An *ideal* of a general ring R is a *two-sided* ideal, meaning it is both a left and a right ideal. However, all rings we care about will be commutative, and so we will not bother with these distinctions. Note that being an ideal of R is a *stronger* property than I being a subring of R. A subring must only be closed under multiplication by elements of itself. An ideal is closed under multiplication by *any* element of R. This is part of why we want to require that rings have multiplicative identities. If we didn't then ideals would be subrings. However, they morally *aren't* subrings, because they have stronger properties.

Definition B.2.14. Let $(R, +, \cdot)$ be a commutative ring and let $r \in R$. The **principal ideal** generated by r is $(r) = \{x \cdot r : x \in R\}$, the set of R-multiples of r. An ideal I is called **principal** if $I = (r)$ for some $r \in R$.

Note that (r) really is an ideal: indeed, for any $a, b \in (r)$ we have $a = xr$ and $b = yr$ for some $x, y \in R$, and therefore $a + b = (x + y)r \in (r)$. We have $0 = 0 \cdot r \in (r)$, and clearly $-a = (-x)r \in (r)$ so (r) is an additive subgroup of $(R, +)$. For any $s \in R$ and $a \in (r)$, we have $sa = s(xr) = (sx)r \in (r)$, so (r) absorbs multiplication by elements of R, which means its an ideal.

Example B.2.15. $I = (0)$ is an ideal of any commutative ring R.

Example B.2.16. The even integers $2\mathbb{Z}$ are an ideal of $\mathbb{Z}$. In fact, $2\mathbb{Z} = (2)$ is the principal ideal generated by 2 (hence, the notation!)

Example B.2.17. In $\mathbb{Q}$, one has $(1) = \mathbb{Q}$. This is because any element of $\mathbb{Q}$ is a $\mathbb{Q}$-multiple of 1: in particular, $\frac{a}{b} = \frac{a}{b} \cdot 1 \in (1)$. In general, for any commutative ring R one has $R = (1)$ because for any $r \in R$, one has $r = r \cdot 1 \in (1)$.

Example B.2.18. In $\mathbb{Z}[x]$, the ideal $I = (x^2 + 1)$ consists of all $\mathbb{Z}[x]$-multiples of $x^2 + 1$. For example, $(x^2 + 1) \cdot (x^2 + 2) = x^4 + 3x^2 + 2 \in I$, but $2x \notin I$ because there is no polynomial $p(x) \in \mathbb{Z}[x]$ with $(x^2 + 2)p(x) = 2x$.

Example B.2.19. Consider $S = \mathbb{R}[x^2]$, the set of polynomials with real coefficients in which only even powers of x occur. Then S is a subring of $\mathbb{R}[x^2]$, but S is *not* an ideal of $\mathbb{R}[x^2]$, because $x \cdot (x^2) = x^3 \notin S$.

Definition B.2.20. A **field** is a pair $(F, +, \cdot)$ such that $(F, +, \cdot)$ forms a commutative ring with $0 \neq 1$ and $F^\times = F \setminus \{0\}$. That is, F is a commutative ring where all non-zero have a multiplicative inverse, so one can divide by any non-zero element.

Note that we require $0 \neq 1$ in the definition of a field. This is because we do not want the zero ring to be a field!

Example B.2.21. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all familiar examples of fields. $\mathbb{Z}$ is *not* a field, because one cannot divide by 2 and stay in the integers: $\frac{1}{2} \notin \mathbb{Z}$.

Example B.2.22. Let $\mathbb{Q}(x)$ denote the set of *rational functions* in the indeterminate x with coefficient in $\mathbb{Q}$. That is, a function $f(x) \in \mathbb{Q}(x)$ is of the form $f(x) = \frac{p(x)}{q(x)}$ for some polynomial $p(x), q(x) \in \mathbb{Q}[x]$ with $\mathbb{Q}[x]$. The operations of addition and multiplication are once again, the ones familiar from middle school. With respect to these operations, $\mathbb{Q}(x)$ forms a field.

Example B.2.23. Let F be a field. Then the only ideals of F are (0) and F. Indeed, (0) is always an ideal. Suppose that $I \subset F$ is a non-zero ideal. This means there is some $x \in I$ with $x \neq 0$. Since F is a field, $x^{-1} \in F$ exists, and therefore $x \cdot x^{-1} = 1 \in I$. Therefore, $I = (1) = F$.

Definition B.2.24. A **subfield** of a field $(F, +, \cdot)$ is a subset $K \subset F$ such that $(K, +, \cdot)$ is also a field.

Example B.2.25. $\mathbb{Q}$ is a subfield of $\mathbb{R}$. Its also a subfield of $\mathbb{C}$. $\mathbb{R}$ is also a subfield of $\mathbb{C}$.

B.3 Morphisms

A morphism is a way of talking between two objects of the same type.

Definition B.3.1. Let R, S be two objects of the same type (groups, rings, or fields). A **homomorphism** between R and S is a function $f : R \to S$ that preserves algebraic structure.

- If R, S are groups, a **group homomorphism** is a map $f : R \to S$ such that for any $a, b \in R$ one has $f(a \cdot_R b) = f(a) \cdot_S f(b)$ and $f(e_R) = e_S$ where $\cdot_R$ and $\cdot_S$ are the operations on R, S respectively, and e_R, e_S are the identity elements of R, S.
- If R, S are rings, a **ring homomorphism** is a map $f : R \to S$ such that for any $a, b \in R$ one has $f(a +_R b) = f(a) +_S f(b)$,$f(a \cdot_R b) = f(a) \cdot_S f(b)$, and $f(1_R) = 1_S$.

If a homomorphism f is a bijection, we call f an **isomorphism**. If f is an isomorphism and $R = S$, we call f an **automorphism**. If two objects R and S are isomorphic we write $R \cong S$.

Note that we did not define the notion of a homomorphism of fields. Since fields are *rings*, a field homomorphism is just a ring homomorphism. In particular, one may check that for any homomorphism of ring f, one must also have $f(0_R) = 0_S$, $f(-r) = -f(r)$, and $f(r^{-1}) = f(r)^{-1}$.

Definition B.3.2. Let R, S be two objects of the same type, and suppose that R is a sub-object of S (i.e. a subgroup, subring, subfield, etc.). There is a homomorphism $i : R \to S$ called the **inclusion map**, simply defined by $i(x) = x$ for all $x \in R$. For example, $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{Q}, +)$, so the inclusion map $i : \mathbb{Z} \to \mathbb{Q}$ that sends $x \in \mathbb{Z}$ to $x \in \mathbb{Q}$ is a group homomorphism. We may also view $\mathbb{Z}$ as a subring of $\mathbb{Q}$, and so the inclusion map is also a homomorphism of rings.

Example B.3.3. Consider the two groups $(\mathbb{R}, +)$ and $(\mathbb{R}_{>0}, \cdot)$ with respect to addition and multiplication of real numbers. The exponential map $\exp : \mathbb{R} \to \mathbb{R}_{>0}$ given by $\exp(x) = e^x$ is a group homomorphism. This is because $\exp(0) = e^0 = 1$, and $\exp(x + y) = e^{x+y} = e^x \cdot e^y = \exp(x) \cdot \exp(y)$ by properties of exponential functions. This map is actually a group isomorphism, with the inverse map $\log : \mathbb{R}_{>0} \to \mathbb{R}$, and so $\mathbb{R} \cong \mathbb{R}_{>0}$ as groups.

Example B.3.4. Consider the complex numbers $\mathbb{C}$, and let $\sigma : \mathbb{C} \to \mathbb{C}$ denote the complex conjugation map. That is, $\sigma(a+bi) = a-bi$. Then σ is a field automorphism, and we usually denote $\sigma(a+bi)$ as $\overline{a+bi}$. That σ is an automorphism follows rather easily: for two complex numbers $z = a + bi$ and $w = c + di$, one has $z + w = (a + bi) + (c + di) = (a + c) + (b + d)i$ and $zw = (a + bi)(c + di) = (ac - bd) + (ad + bc)i$ by definition. It's then straightforward to see that $\sigma(z + w) = \sigma(z) + \sigma(w)$, $\sigma(zw) = \sigma(z)\sigma(w)$, $\sigma(0) = 0$, $\sigma(1) = 1$. This means that σ is a field homomorphism, and clearly $\sigma(\sigma(z)) = z$, so the inverse map of σ is itself which means its an automorphism.

Example B.3.5. Let R be a ring and let $r \in R$. The map $\mathrm{ev}_r : R[x] \to R$ defined by $\mathrm{ev}_r(p(x)) = p(r)$ is a ring homomorphism called the **evaluation at r map**. Let $0, 1$ denote the constant polynomials $p(x) = 0$ and $p(x) = 1$ in $R[x]$. Note that $\mathrm{ev}_r(0) = 0(r) = 0$ and $\mathrm{ev}_r(1) = 1(r) = 1$ since 0 and 1 are constant polynomials, and by definition of function addition/multiplication, for two polynomials $p(x), q(x) \in R[x]$ one has $(p + q)(r) = p(r) + q(r)$ and $(p \cdot q)(r) = p(r)q(r)$, which means $\mathrm{ev}_r(p(x) + q(x)) = \mathrm{ev}_r(p(x)) + \mathrm{ev}_r(q(x))$ and $\mathrm{ev}_r(p(x)q(x)) = \mathrm{ev}_r(p(x))\mathrm{ev}_r(q(x))$. Thus, ev_r is a ring homomorphism.

Definition B.3.6. Let $f : R \to S$ be a homomorphism. The **kernel** of f, $\ker(f)$ is defined by:

- $\ker(f) = \{x \in R : f(x) = e_S\}$ if R, S are groups.
- $\ker(f) = \{x \in R : f(x) = 0_S\}$ if R, S are rings.

The **image** of f, $\mathrm{Im}(f)$ is defined by $\mathrm{Im}(f) = \{f(r) : r \in R\}$.

The kernel and image of a morphism f are special subsets of R and S respectively. They measure the failure of a morphism to be injective/surjective.

Proposition 55. *Let $f : R \to S$ be a homomorphism.*

- $\ker(f)$ *is trivial if and only if* f *is injective.*
- $\mathrm{Im}(f) = S$ *if and only if* f *is surjective.*

Proof. We will give the proof for homomorphisms of rings. The proof for groups is left as an exercise.

- First, suppose that $\ker(f)$ is trivial, and that $f(x) = f(y)$ for $x, y \in R$. Then $0_S = f(x) - f(y) = f(x - y)$ because f is a homomorphism, and therefore $x - y \in \ker(f)$. As $\ker(f)$ is trivial, this means $x - y = 0_R$, so $x = y$. This says x is injective. For the other direction, note that $f(0_R) = 0_S$ always, and so if $x \in \ker(f)$, this means $f(x) = 0_S$ and so by injectivity of f, we have $x = 0_R$ which means $\ker(f) = \{0_R\}$ is trivial.
- By definition, f is surjective if and only if for any $s \in S$ we may find $r \in R$ with $f(r) = s$. This is precisely the same thing as saying that $\mathrm{Im}(f) = S$.

□

Proposition 56. *Let* $f : R \to S$ *be a homomorphism.*

- *If* f *is a homomorphism of groups, then* $\ker(f)$ *and* $\mathrm{Im}(f)$ *are subgroups of* R, S *respectively.*
- *If* f *is a homomorphism of rings, then* $\ker(f)$ *is an ideal of* R *and* $\mathrm{Im}(f)$ *is a subring of* S*.*

Proof. Exercise. □

Why is the image of a ring homomorphism not an ideal? Since $f(1_R) = 1_S$, if $\mathrm{Im}(f)$ was an ideal, then necessarily we would have $\mathrm{Im}(f) \supset (1_S) = S$, so $\mathrm{Im}(f) = S$ would mean that f is surjective. However, obviously not all ring homomorphisms are surjective! Similarly, if $1_R \in \ker(f)$ then $R = (1_R) \subset \ker(f)$ so f would be the zero map, and so $\ker(f)$ is almost *never* a subring of R.

Example B.3.7. The inclusion map $i : \mathbb{Z} \to \mathbb{Q}$ as a map of rings is injective, and so $\ker(i) = (0)$ is trivial, and clearly $\mathrm{Im}(f) = \mathbb{Z}$.

Example B.3.8. We showed before that if F is a field, then the only ideals of F are (0) and F. This means that if $f : F \to K$ is a homomorphism of fields, either $\ker(f) = (0)$ so f is injective, or $\ker(f) = F$ so f is the zero map.

Example B.3.9. Let $\mathrm{GL}_2(\mathbb{R})$ be the group of non-zero invertible matrices with real entries with respect to multiplication. Consider the determinant map $\det : \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^\times$ given by $A \to \det(A)$. One may check that det is indeed a group homomorphism. We note that $\ker(\det) = \{A \in \mathrm{GL}_2(\mathbb{R}) : \det(A) = 1\} = \mathrm{SL}_2(\mathbb{R})$, and $\mathrm{Im}(\det) = \mathbb{R}^\times$. That this latter equality is true can be easily seen from $\det(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}) = a$ for any $a \in \mathbb{R}^\times$.

Example B.3.10. Let R, S be rings. The projection map $\pi_R : R \times S \to R$ given by $\pi((r, s)) = r$ is a surjective ring homomorphism with $\mathrm{Im}(\pi_R) = R$ and $\ker(\pi_R) = \{0_R\} \times S \cong S$.