MAU22103/33101 - Introduction to Number Theory

Adam Keilthy

Trinity College Dublin

Contents

0 A brief introduction

Broadly speaking, number theory is a field of mathematics concerned with studying the integers and their properties. This covers questions of algebraic relations, divisibility, the distribution of primes, and the properties of arithmetic functions. More generally, modern number theory also studies the properties of algebraic numbers, considers questions of transcendence, and (rational) point counting on curves. Fields in number theory are more often classified by their techniques than the specific problems they consider, and in this course we will try to touch on three main areas.

Elementary number theory is concerned mostly with the properties of integers, and relies on little to no techniques from analysis or complex numbers. This means that the problems can usually be explained very easily, but the proofs can either be very simple or incredibly involved.

Algebraic number theory considers wider classes of integer-like numbers, such as algebraic integers, and properties of number fields. It makes use of symmetries and group theory to solve problems, considers analogues of prime numbers over the complex numbers, and tries to solve problems modulo various prime powers.

Analytic number theory tries to encode number theoretic problems in terms of analytic functions. This gives a very powerful tool for describing the statistical properties of numbers, such as estimating the number of solutions to a Diophantine equation of a certain size.

One of the most common types of problem in number theory is the question of solving a Diophantine equation.

Definition 0.1. *A Diophantine equation is a polynomial equation*

$$F(x_1, \ldots, x_n) = 0$$

for some polynomial $F \in \mathbb{Z}[x_1, \ldots, x_n]$ with integer coefficients, where we demand that a solution $(x_1, \ldots, x_n)$ be given in integers.

It is easy to determine whether a polynomial equation has solutions over $\mathbb{C}$: pick a value for every variable except one, and apply the fundamental theorem of algebra. Even over $\mathbb{R}$, it is often possible to find a solution relatively easily. For example

$$x^3 + y^3 + z^3 = 29$$

has real solutions given by

$$(x, y, \sqrt[3]{29 - x^3 - y^3})$$

for every real x and y. Finding integer solutions is more difficult.

Example 0.2. *Find all solutions $(x, y, z) \in \mathbb{Z}^3$ to the Diophantine equation*

$$x^3 + y^3 + z^3 = 29.$$

Some solutions include $(1, 1, 3)$ and $(4, -3, -2)$, but it is hard to tell whether these are all solutions, whether there are finitely many solutions, or whether there are infinitely many solutions.

Even very similar Diophantine equations can have very different solutions

Example 0.3. *The Diophantine equation*

$$x^3 + y^3 + z^3 = 30$$

has smallest solution

$$(2,220,422,932; -2,218,888,517; -283,059,965).$$

The Diophantine equations

$$x^3 + y^3 + z^3 = 31,$$
$$x^3 + y^3 + z^3 = 32$$

have no solutions, as can be seen by considering remainders on division by 9. The Diophantine equation

$$x^3 + y^3 + z^3 = 33$$

was only found to have solutions in 2019!

We have many classes of Diophantine equation to study:

- Fermat's Last Theorem says that $x^n + y^n = z^n$ has no positive integer solutions for any $n > 2$,
- The Odd Change Problem (or the Chicken Nugget Problem) asks questions like whether $5x + 7y = 53$ has integer solutions,

- Whether elliptic curves such as $y^2 = x^3 + 7x + 3$ have integer solutions comes up in relation to encryption.

Unfortunately, a result due to Matiyasevich (1970) says that there is no universal algorithm to determine whether a Diophantine equation has solutions, let alone to find them. Each problem has to be considered essentially unique. The upside is that this means we're unlikely to run out of problems to work on!

Even though there is no universal technique to solve every Diophantine equation, there are at least some standard first steps to take. And these begin with understanding the integers.

1 The integers and divisibility

We will denote the set of integers by

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

and the set of positive integers (natural numbers) by

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

Remark 1.1. *Be careful to check how* $\mathbb{N}$ *is defined in any textbooks you might reference. Some authors include* 0 *in the natural numbers!*

One of the most useful basic tools in number theory is the concept of *division with remainder* or *Euclidean division*.

Theorem 1.2. *Let* $a \in \mathbb{Z}$ *and* $b \in \mathbb{N}$*. Then there exists a unique* $q \in \mathbb{Z}$*, called the quotient, and* $r \in \mathbb{Z}$*, called the remainder, such that*

$$a = bq + r \quad \textit{and} \quad 0 \leq r < b.$$

Proof. We will assume, without loss of generality, that $a \geq 0$. Let

$$q := \max\{n \in \mathbb{Z} \mid bn \leq a\}$$

and note that this maximum exists, since this set is bounded above. Indeed, we can easily replace this set with a finite set, if we wanted to, without changing the maximal element. Define $r := a - bq$. Since $bq \leq a$, we have that $r \geq 0$. To see that $r < b$, suppose otherwise. Then we would have

$$a = bq + r \geq bq + b = b(q + 1)$$

contradicting the maximality of q. Hence we must have $r < b$.

To see that q and r are unique, suppose we have q', $r' \in \mathbb{Z}$ such that $0 \leq r' < b$ and $a = bq' + r'$. Since

$$0 \leq r, r' < b$$

we must have that

$$-b < r - r' < b$$

and hence

$$-b < (a - bq) - (a - bq') < b.$$

Simplifying this, we see that

$$-b < b(q' - q) < b$$

and so

$$-1 < q' - q < 1.$$

However, there do not exist distinct integers with a difference of less than 1 in absolute values, and so we must have $q = q'$ and thus $r = r'$. □

Remark 1.3. *We can also define division by $b < 0$ by applying the above theorem to $-b$.*

Definition 1.4. *For $a, b \in \mathbb{Z}$, we say that a divides b, denoted $a|b$, if there exists $q \in \mathbb{Z}$ such that $b = aq$. We call a a divisor of b and b a multiple of a.*

Remark 1.5. *Note that for $a \in \mathbb{N}$, the statement that $a|b$ is equivalent to the remainder of b on division by a is 0.*

Example 1.6. *We have that $2|18$, since $18 = 2(9)$, and $-3|18$, since $18 = -3(-6)$, but $4 \nmid 18$. For any integer n, we have that $\pm 1|n$, but if $n| \pm 1$, then we must have $n = \pm 1$. Note that if $n|m$, and $m \neq 0$, we must have $|n| \leq |m|$. Then for every integer n, we have that $n|0$, since $0 = n(0)$.*

Remark 1.7. *Interestingly, our definition allows us to talk about something being divisible by 0: $0|n$ if and only if $n = 0$. However, even though we can talk about divisibility by 0, we cannot define division by 0. For non-zero numbers a, b, if there exists q such that $b = aq$, that q is unique and can be used as a definition of b divided by a. This uniqueness fails for $a = b = 0$, and so we cannot define 0 divided by 0 in this way.*

Let us quickly discuss some properties of divisibility, and the additive and multiplicative structures on $\mathbb{Z}$.

Proposition 1.8. *Let $a, b, c \in \mathbb{Z}$. If $a|b$ and $a|c$, then $a|(bm + cn)$ for every $m, n \in \mathbb{Z}$.*

Proof. If $a|b$, then $b = ak$ for some $k \in ZZ$. Similarly, $c = a\ell$ for some $\ell \in \mathbb{Z}$. Therefore

$$bm + cn = akm + a\ell n = a(km + \ell n)$$

which means precisely that $a|(bm + cn)$. □

Proposition 1.9. *Let $a, b, c \in \mathbb{Z}$. If $a|b$ or $a|c$, then $a|bc$.*

Proof. We can assume, without loss of generality, that $a|b$ and so $b = ak$ for some $k \in \mathbb{Z}$. Then $bc = (ak)c = a(kc)$, and hence $a|bc$. □

In a much less exciting world, the converse to this theorem would be true: if $a|bc$ then a would divide one of b or c. But while $6|18 = 2(9)$, $6 \nmid 2$ and $6 \nmid 9$. The first goal of this course will be to provide conditions under which a converse does hold.

1.1 Greatest common divisors and Bézout's theorem

Definition 1.10. *Let $a, b \in \mathbb{Z}$, not both zero. We define their greatest common divisor by*

$$\gcd(a, b) := \max\{d \in \mathbb{N} \mid d|a \text{ and } d|b\}.$$

We define $\gcd(0, 0) := 0$. *We define their least common multiple by*

$$\operatorname{lcm}(a, b) := \min\{\ell \in \mathbb{N} \mid a|\ell \text{ and } b|\ell\}.$$

We define $\operatorname{lcm}(0, b) = 0$. *If* $\gcd(a, b) = 1$, *we call a and b coprime.*

Exercise. *Why are the sets involve in the definitions of* gcd *and* lcm *non-empty? Why are the minimum and maximum elements well defined?*

Example 1.11. *By comparing lists of divisors and multiples, it is easy to see that*

$$\gcd(4, 18) = 2\,, \quad \operatorname{lcm}(4, 18) = 36.$$

Example 1.12. *How can we compute the greatest common divisor of n and $n + 1$ for any $n \in \mathbb{Z}$? We know that if $d|n$ and $d|(n + 1)$ then*

$$d|a(n + 1) + bn$$

for any $a, b \in \mathbb{Z}$. In particular, since $\gcd(n, n + 1)|n$ *and* $\gcd(n, n + 1)|(n + 1)$, *we must have*

$$\gcd(n, n + 1)|a(n + 1) + bn.$$

Taking $a = 1$ and $b = -1$, we therefore find

$$\gcd(n, n + 1)|(n + 1 - n) = 1$$

and hence $\gcd(n, n + 1) = 1$.

While in the case of 4 and 18, it was possible to explicitly compare all divisors in order to determine the greatest common divisor, this is impractical for any larger integers. However, Example 1.12 illustrates an alternative approach by considering linear combinations of the integers in question. This leads to the following algorithmic approach for computing the greatest common divisor relatively quickly.

Theorem 1.13 (Euclid's Algorithm)**.** *Let $ab \in \mathbb{N}$. Let $r_{-1} = a$, $r_0 = b$ and for every $n \geq 0$ such that $r_{n-1} \neq 0$, define r_n as the remainder on dividing r_{n-2} by r_{n-1}, so that r_1 is the remainder on dividing a by b, r_2 is the remainder on dividing b by r_1, and so on. Then $\gcd(a, b)$ is equal to the last non-zero remainder.*

Proof. Denote by

$$\mathrm{Div}(a, b) := \{d \in \mathbb{N} \mid d|a \text{ and } d|b\}$$

the set of common divisors of a and b. We first note that, if $a = bq + r$, then

$$\mathrm{Div}(a, b) = \mathrm{Div}(b, r) :$$

Clearly, if $d|a$ and $d|b$, then $d|(a - bq) = r$, and similarly if $d|b$ and $d|r$, then $d|(bq + r) = a$. Therefore

$$\mathrm{Div}(a, b) = \mathrm{Div}(r_{-1}, r_0) = \mathrm{Div}(r_0, r_1) = \mathrm{Div}(r_1, r_2) = \cdots = \mathrm{Div}(r_{n-1}, r_n).$$

As $b = r_0 > r_1 > r_2 > \cdots > r_n \geq 0$, the sequence of remainders is strictly decreasing, so there must exist an n for which $r_n = 0$, but $r_{n-1} \neq 0$. Hence

$$\mathrm{Div}(a, b) = \mathrm{Div}(r_{n-1}, 0) = \mathrm{Div}(r_{n-1}).$$

We have that $\gcd(a, b)$ is the maximum element of the left hand set, and hence the maximum element of the right hand set, which is clearly r_{n-1}. □

Example 1.14. *Let us compute the greatest common divisor of $a = 17$ and $b = 7$. Then $17 = 2(7) + 3$, so $r_1 = 3$, and $7 = 2(3) + 1$, so $r_2 = 1$ and $3 = 3(1)$, so $r_3 = 0$. Hence, we must have $\gcd(17, 7) = 1$, as expected.*

Lets try an example with bigger numbers: $a = 323$ and $b = 102$. Then $323 = 3(102) + 17$, so $r_1 = 17$, and $102 = 6(17)$, so $r_2 = 0$. Hence $\gcd(323, 102) = 17$, which can easily be checked to be true.

This algorithm actually enables us to solve our first family of Diophantine equations: the Odd Change Problem. Specifically, by running this algorithm in reverse, we obtain an expression for $\gcd(a, b)$ in terms of a and b.

Theorem 1.15 (Bézout)**.** *Let $a, b \in \mathbb{Z}$. Then there exist integers u, v such that*

$$au + bv = \gcd(a, b)$$

Proof. The case where one of a or b is zero is easy, so we will assume without loss of generality that $a, b \in \mathbb{N}$. We then compute $\gcd(a, b)$ by Euclid's algorithm to obtain

$$\begin{aligned}
\gcd(a, b) = r_{n-1} &= r_{n-3} - r_{n-2}q_{n-2} \\
&= r_{n-3} - (r_{n-4} - r_{n-3}q_{n-3})\, q_{n-2} \\
&= r_{n-3}\,(1 + q_{n-2}q_{n-3}) + r_{n-4}q_{n-2} \\
&\vdots \\
&= r_{-1}u + r_0 v = au + bv
\end{aligned}$$

where at each stage, we use $r_k = r_{k+1}q_{k+1} + r_{k+2}$ to replace the latest remainder by a combination of earlier ones. □

Example 1.16. *Recall that* $\gcd(17, 7) = 1$. *The algorithm gives us*

$$\begin{aligned} 1 &= 7 - 2(3) \\ &= 7 - 2(17 - 2(7)) \\ &= 5(7) - 2(17). \end{aligned}$$

Corollary 1.17. *Integers a and b are coprime if and only if there exist $u, v \in \mathbb{Z}$ such that $au + bv = 1$.*

Corollary 1.18. *There exists $u, v \in \mathbb{Z}$ such that $au + bv = k$ if and only if $\gcd(a, b)|k$.*

Proof. If $\gcd(a, b)|k$, then there exists $\ell \in \mathbb{Z}$ such that $k = \ell \gcd(a, b)$. Apply Theorem 1.15 to find $u_0, v_0 \in \mathbb{Z}$ such that

$$au_0 + bv_0 = \gcd(a, b).$$

Then

$$a(\ell u_0) + b(\ell v_0) = \ell \gcd(a, b) = k.$$

Conversely, if $au + bv = k$, then $\gcd(a, b)|k$, as $\gcd(a, b)$ divides any integer linear combination of a and b. □

This corollary tells us exactly when we can solve the Odd Change Problem: $ax + by = c$ has a solution with $x, y \in \mathbb{Z}$ if and only if $\gcd(a, b)|c$. By applying Euclid's algorithm, we can even find a solution. But that does not necessarily give us the only solution, or give us a way to find more. Nevertheless, we can explicitly describe all possible solutions. We just need an important lemma first.

Lemma 1.19 (Gauss)**.** *Let $a, b, c \in \mathbb{Z}$, and suppose that $a|bc$ and $\gcd(a, c) = 1$. Then $a|b$.*

Proof. Since $\gcd(a, c) = 1$, there exist $u, v \in \mathbb{Z}$ such that $au + cv = 1$. Hence

$$abu + bcv = b.$$

But $a|a$ and $a|bc$, so $a|(abu + bcv) = b$. □

With this in mind, we can generate all solutions to the odd change problem.

Proposition 1.20. *Let $a, b, c \in \mathbb{Z}$ and suppose $\gcd(a, b)|c$. Suppose we have a particular solution $x_0, y_0 \in \mathbb{Z}$ to the Diophantine equation $ax + by = c$. Then every solution is of the form $(x, y) = (x_0 + bk, y_0 - ak)$ for $k \in \mathbb{Z}$, and every such pair gives a solution.*

Proof. First, we exclude the cases where a or b is zero, as these are straightforward. Next note that, if $ax_0 + by_0 = 0$, then $a(x_0 + bk) + b(y_0 - ak) = c$ for every k. As such, it suffices to show that these are the only possible solutions. Furthermore, by dividing a, b, c by $\gcd(a, b)$ if necessary, we can assume that $\gcd(a, b) = 1$. Now suppose we have two solutions to $ax + by = c$: (x_0, y_0) and (x_1, y_1). Since

$$ax_0 + by_0 = c = ax_1 + by_1,$$

we must have that

$$a(x_0 - x_1) = b(y_1 - y_0)$$

and so $a|b(y_1 - y_0)$ and $b|a(x_0 - x_1)$. But $\gcd(a, b) = 1$ and so by Lemma 1.19, we must have that $a|(y_1 - y_0)$ and $b|(x_0 - x_1)$. Thus $x_1 - x_0 = kb$ and $y_1 - y_0 = \ell a$ for some $k, \ell \in \mathbb{Z}$. Now, since

$$-abk = a(x_0 - x_1) = b(y_1 - y_0) = ab\ell$$

we conclude that $\ell = -k$. The result then follows: $x_1 = x_0 + bk$ and $y_1 = y_0 - ak$ for some integer k. □

Example 1.21. *The Diophantine equation* $3x + 18y = 7$ *has no solutions as* $\gcd(3, 18) = 3 \nmid 7$. *The Diophantine equation* $5x + 7y = 53$ *has infinitely many solutions, as* $\gcd(5, 7) = 1|53$. *Note that*

$$3(5) - 2(7) = 1$$

and so a particular solution is given by

$$159(5) - 106(7) = 53$$

and hence the general solution is given by

$$(x, y) = (159 + 7k, -106 - 5k)$$

for $k \in \mathbb{Z}$. *In particular, taking* $k = -22$, *we get a solution* $(5, 4)$ *in positive integers.*

1.2 Prime numbers

Lemma 1.19 becomes particularly powerful when applied to prime numbers, letting us prove an important theorem about the multiplicative structure of the integers.

Definition 1.22. *A positive integer* $p \in \mathbb{N}$ *is called prime if it has exactly two (distinct) divisors, i.e.* $p \neq 1$ *and if* $d|p$, *then* $d \in \{1, p\}$. *Non-prime integers greater than or equal to* 2 *are called composite.*

An immediate consequence of this definition is that for any integer n, either $p|n$ or $\gcd(p, n) = 1$. This lets us conclude an immediate corollary from Lemma 1.19, originally due to Euclid.

Corollary 1.23 (Euclid)**.** *Let p be prime and $b, c \in \mathbb{Z}$. If $p|bc$, then $p|b$ or $p|c$.*

Proof. If $p|c$, we are done. Otherwise, $\gcd(p, c) = 1$, and so Lemma 1.19 applies and $p|b$. □

Remark 1.24. *Up to positivity and non-oneness, this property uniquely defines the prime numbers. A positive integer greater than one is prime if and only if $p|bc$ imples $p|b$ or $p|c$.*

Theorem 1.25 (The fundamental theorem of arithmetic)**.** *Up to the order of the factors, every $n \in \mathbb{N}$ can be uniquely written as a product of prime numbers (with repetition).*

Proof. If $n = p$ is prime, then $n = p$ is the desired product. If $n = 1$, we take the empty product. Otherwise, n has a divisor $1 < a < n$. Writing $n = ab$, where $1 < b < n$, we can apply induction to obtain the existence of such a product representation: we write a and b as a product of primes, and take their product.

To see that this factorisation is unique, suppose

$$n = p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$$

has two factorisations into primes $p_1, \dots, p_r, q_1, \dots, q_s$. Then we have that

$$p_1 | p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$$

and so, by repeatedly applying Corollary 1.23, we must have that $p_1 | q_k$ for some q_k. But since q_k is prime, and $p_1 \neq 1$, we must have $p_1 = q_k$. Thus, we can divide both factorisations by p_1 to obtain

$$p_2 p_3 \dots p_r = q_1 \dots q_{k-1} q_{k+1} \dots q_s.$$

Repeating this argument, or applying induction, lets us conclude that these two factorisations must be the same, up to order of factors, and hence that the two factorisations of n must be the same, up to order of factors. □

This is a very important result in the structure theory of the integers, and is one of the key reasons to study prime numbers. An obvious, immediate question, is to ask how many primes there are. If there were only finitely many, the integers would be incredibly rigid.

1.2.1 Five proofs of the infinitude of primes.

There are dozens of proofs of there being infinitely many primes, and I encourage you to explore them as many contain some really interesting ideas. Here, I will present five of my favourites.

Theorem 1.26 (Euclid)**.** *There are infinitely many prime numbers*

Euclid. Suppose there are only finitely many prime numbers $p_1, \ldots, p_k$ and consider the integer $N = p_1 p_2 \ldots p_k + 1$. Clearly $p_i \nmid N$, as $p_i | (N-1)$ and $\gcd(N-1, N) = 1$, and $p_i \nmid 1$. But the fundamental theorem of arithmetic tells us that N must have a prime divisor q, and q must be distinct from $p_1, \ldots, p_k$. This gives a contraction, and so there must be infinitely many primes. □

Proof. Call a positive integer $r \in \mathbb{N}$ squarefree if $m^2 | r$ implies that $m = \pm 1$. It is easy to see that every positive integer n can be written uniquely as a product of a square and a squarefree number $n = rs^2$. Suppose there are only finitely many primes $p_1, \ldots, p_k$. Then, every squarefree number r can be written uniquely as a product of these primes, with each prime appearing exactly once. Hence

$$\left(1 + \frac{1}{p_1}\right)\left(1 + 1\frac{1}{p_2}\right) \cdots \left(1 + \frac{1}{p_k}\right) = \sum_{r \text{ squarefree}} \frac{1}{r}$$

and furthermore this sum is finite. Recall from analysis that the sum

$$\sum_{s=1}^{\infty} \frac{1}{s^2}$$

is also finite. Hence, their product

$$\sum_{r \text{ squarefree}} \sum_{s=1}^{\infty} \frac{1}{rs^2}$$

is finite. But every positive integer n will appear as the numerator exactly once in this double sum, and so this product is equal to $\sum_{n=1}^{\infty} \frac{1}{n}$, which we know to be infinite. This gives a contraction, and so there must be infinitely many primes. □

Erdös. Let $\pi(N) = \#\{p \in \mathbb{N} \mid p \leq N \text{ and } p \text{ prime}\}$ be the prime counting function. Since every squarefree number can be uniquely written as a product of distinct primes, there are at most $2^{\pi(N)}$ squarefree positive integers less than or equal to N. There are at most $\sqrt{N}$ perfect squares less than or equal to N. Since every positive integer can be written as a product of a squarefree number and a square, there are at most $2^{\pi(N)}\sqrt{N}$ positive integers less than or equal to N. Since there are exactly N positive integers less than or equal to N, we must have that

$$N \leq 2^{\pi(N)}\sqrt{N}.$$

Rearranging this, we get that $\pi(N) \geq \frac{1}{2} \log N$, which clearly tends to infinity as N does. □

Fürstenberg. Call a subset $U \subset \mathbb{Z}$ *good* if $U = \emptyset$ or U is a union of arithmetic progressions

$$\{x + an \mid n \in \mathbb{Z}\}.$$

Note that every good set is empty or infinite. Good sets can be easily checked to have the following properties

- $\emptyset$ and $\mathbb{Z}$ are good,
- Any union of good sets is good,
- Any finite intersection of good sets is good.

Note that, for fixed m the set $\{n \in \mathbb{Z} \mid m \nmid n\}$ is good:

$$\{n \in \mathbb{Z} \mid m \nmid n\} = \bigcup_{k=0}^{m-1} \{k + mn \mid n \in \mathbb{Z}\}.$$

Hence, for any finite collection of primes $p_1, \ldots, p_r$, the set

$$\{n \in \mathbb{Z} \mid p_1 \nmid n, p_2 \nmid n, \ldots, p_r \nmid n\} = \bigcap_{i=1}^{r} \{n \in \mathbb{Z} \mid p_i \nmid n\}$$

is good. Thus, if there are only finitely many primes $p_1, \ldots, p_r$, the set

$$\{\pm 1\} = \{n \in \mathbb{Z} \mid p_1 \nmid n, p_2 \nmid n, \ldots, p_r \nmid n\}$$

is good. But this set is finite and non-empty, and good sets are empty or infinite. This gives a contraction, and so there must be infinitely many primes. □

Meštrović. Suppose there are only finitely many prime numbers $2, 3, p_3, p_4, \ldots, p_k$ and let $P = 3p_3p_4 \ldots p_k$ be the product all the odd primes. If $3|n$ or $p_i|n$, then we must have $\gcd(P, n) > 1$, and so the set of all numbers coprime to P is $\{1, 2, 2^2, 2^3, \ldots\}$.

Note that $\gcd(P, P-2) | (P - (P-2)) = 2$, and since $2 \nmid P$, we therefore have $\gcd(P, P-2) = 1$. Hence $P - 2 \in \{1, 2, 2^2, \ldots\}$. But $2 \nmid (P-2)$: if it did, then 2 would divide P. Therefore $P - 2 = 1$ is the only possibility. Hence $P = 3$, and 2 and 3 are the only prime numbers. But 5 is prime. This gives a contraction, and so there must be infinitely many primes. □

1.2.2 Factoring with efficiency

In general, it is difficult to determine the factorisation of an integer. It is often difficult to find any non-trivial divisors. This is mostly a good thing, as many encryption algorithms rely on factoring being hard, but does make our computations harder. The best improvement we realistically have is comes from the following lemma.

Lemma 1.27. *Let $n \in \mathbb{N}$, $n \geq 2$ be a composite number. Then there exists a prime $p|n$ such that $p \leq \sqrt{n}$.*

Proof. If n is composite, then we can write $n = ab$ for integers $1 < a, b < n$. If $a, b > \sqrt{n}$, then we would have

$$n = ab > \sqrt{n}\sqrt{n} = n$$

which is absurd. Therefore, at least one of a or b is at most $\sqrt{n}$, without loss of generality a. As any prime factor of a is at most a, picking any prime $p|a$ gives us a $p|n$ with $p \leq \sqrt{n}$. □

Example 1.28. *We can use this lemma to prove that* 23 *is prime. Note that* $\sqrt{23} < \sqrt{25} = 5$*, so if* 23 *is composite, it will have a prime factor less than* 5*. As* 23 *is not divisible by* 2 *or* 3*, it therefore must be prime.*

Example 1.29. *Let us determine the prime factorisation of* 284*. Since* $\sqrt{284} < \sqrt{289} = 17$*, we only need to test prime factors up to* 17*. We quickly find that*

$$284 = 2(142) = 2 \cdot 2 \cdot 71$$

To check if 71 *is prime, we could continue to test primes less than* 17*, but we can use our lemma again to reduce our test space to primes less than* $\sqrt{71} < 9$*. As* 71 *is not divisible by* 2*,* 3*,* 5*, or* 7*, it must be prime. Therefore the prime factorisation of* 284 *is* $2 \cdot 2 \cdot 71$*.*

To determine all divisors of 284*, we note that every divisor of* 284 *must be a product of a subset of its prime divisors. Hence, it suffices to consider all possible subsets of* $\{2, 2, 71\}$*. We therefore find that the set of divisors of* 284 *is*

$$\{1, 2, 4 = 2^2, 71, 142 = 2(71), 284 = 2^2(71)\}$$

1.2.3 p-adic valuations

It is sometimes helpful to consider the exact powers of primes dividing an integer. We sometimes write $p^k || n$ if $p^k | n$, but $p^{k+1} \nmid n$, for a prime p, $k \geq 0$, and $n \in \mathbb{Z}$. We can also use the notation of valuations.

Definition 1.30. *Let* $n \in \mathbb{Z}$*,* $n \neq 0$*, and let* p *be prime. By the fundamental theorem of arithmetic, we can write*

$$n = \pm \prod_{\substack{p|n \\ p \text{ prime}}} p^{a_p}$$

for some $a_p \in \mathbb{N}$*. We define the* p*-adic valuation of* n *to be*

$$v_p(n) = \begin{cases} a_p \text{ if } p|n, \\ 0 \text{ otherwise.} \end{cases}$$

We extend the definition of v_p *to* 0 *by* $v_p(0) = +\infty$ *for every prime* p*.*

Remark 1.31. *It is easy to check that* $p^k || n$ *if and only if* $k = v_p(n)$*. That is to say that the* p*-adic valuation is the biggest/exact power of* p *dividing* n*.*

Remark 1.32. *It is important to note that* $v_p(n) > 0$ *for only finitely many* p*. This lets us cheat a bit notationally and consider infinite products over the primes, in which only finitely many prime contribute. In particular, since* $v_p(n) = 0$ *for all but finitely many* p*, we can make sense of the equality*

$$n = \prod_{p \text{ prime}} p^{v_p(n)}.$$

Proposition 1.33. *The p-adic valuation satisfies the following properties for every $m, n \in \mathbb{Z}$:*

i) $v_p(mn) = v_p(m) + v_p(n)$,

ii) $v_p(m+n) \geq \min\{v_p(m), v_p(n)\}$ *with equality if* $v_p(m) \neq v_p(n)$,

iii) $m|n$ *if and only if* $v_p(m) \leq v_p(n)$ *for every prime* p.

Proof. Exercise! □

Example 1.34. *For* $n = 18$, *we have* $18 = 2 \cdot 3^2$, *so*

$$v_2(18) = 1, \quad v_3(18) = 2, \quad v_p(18) = 0 \text{ for all } p \geq 5.$$

For $n = 14$, *we have* $14 = 2 \cdot 7$, *so*

$$v_2(14) = v_7(14) = 1$$

For $n = 196 = 14^2$, *we have*

$$v_2(196) = v_2(14) + v_2(14) = 2 = v_7(14) + v_7(14) = v_7(196).$$

We also have that

$$1 = v_2(214) = v_2(18 + 196) \geq \min\{v_2(18), v_2(196)\}.$$

Valuations are a very handy tool for computing, and more importantly proving, things to do with greatest common divisors.

Theorem 1.35. *Let* $a, b \in \mathbb{N}$. *Then*

$$\gcd(a,b) = \prod_{p \text{ prime}} p^{\min(v_p(a), v_p(b))},$$
$$\operatorname{lcm}(a,b) = \prod_{p \text{ prime}} p^{\max(v_p(a), v_p(b))}.$$

Proof. If $d|a$ and $d|b$, this is equivalent to $v_p(d) \leq v_p(a)$ and $v_p(d) \leq v_p(b)$ for all primes p. This is in turn equivalent to $v_p(d) \leq \min(v_p(a), v_p(b))$ for all primes p. Clearly d is maximal among divisors when we have equality for every p. Therefore

$$\gcd(a,b) = \prod_{p \text{ prime}} p^{\min(v_p(a), v_p(b))}.$$

The case for the least common multiple is similar. □

Corollary 1.36. *If* $d|a$ *and* $d|b$, *then* $d|\gcd(a,b)$. *If* $a|m$ *and* $b|m$, *then* $\operatorname{lcm}(a,b)|m$.

Proof. Translating the conditions into statements about p-adic valuations immediately gives the result. □

Corollary 1.37. *For $a, b \in \mathbb{N}$, $\gcd(a,b)\operatorname{lcm}(a,b) = ab$.*

Proof.

$$\begin{aligned}\gcd(a,b)\operatorname{lcm}(a,b) &= \prod_{p \text{ prime}} p^{\min(v_p(a),v_p(b))+\max(v_p(a),v_p(b))} \\ &= \prod_{p \text{ prime}} p^{v_p(a)+v_p(b)} \\ &= \left(\prod_p p^{v_p(a)}\right)\left(\prod_p p^{v_p(b)}\right) = ab,\end{aligned}$$

as $\min(x,y) + \max(x,y) = x + y$. □

Corollary 1.37 is quite useful, as it lets us compute least common multiples (which we have computed up until this point by comparing lists of multiples to find a common one - very inefficient) in terms of greatest common divisors (which we can compute reasonably efficiently using Euclid's algorithm)

Example 1.38. *It is quick to compute that* $\gcd(33,12) = 3$*, and so*

$$\operatorname{lcm}(33,12) = \frac{33 \cdot 12}{\gcd(33,12)} = \frac{33 \cdot 12}{3} = 11 \cdot 12 = 132.$$

1.3 Sums of divisors and multiplicative functions

Definition 1.39. *Let $f : \mathbb{N} \to \mathbb{C}$ be a function. We call f strongly multiplicative if*

$$f(mn) = f(m)f(n) \text{ for all } m, n \in \mathbb{N}.$$

We call f (weakly) multiplicative if

$$f(mn) = f(m)f(n) \text{ for all } m, n \in \mathbb{N} \text{ such that } \gcd(m,n) = 1.$$

Strongly multiplicative functions are completely determined by their values on prime numbers, while (weakly) multiplicative functions are completely determined by their values of powers of prime numbers.

Example 1.40. *Some examples of strongly multiplicative functions include*

- $f(n) = 1$,
- $f(n) = n^k$ *for fixed* $k \in \mathbb{N}$,
- $$f(n) = \begin{cases} 1 \text{ if } 2|n, \\ -1 \text{ otherwise}, \end{cases}$$

- $$f(n) = \begin{cases} 0 \text{ if } 2|n, \\ 1 \text{ if } n = 4k+1, \\ -1 \text{ if } n = 4k+3. \end{cases}$$

Strongly multiplicative functions are usually pretty rare.

Example 1.41. *Some examples of (weakly) multiplicative functions include*

- $f(n) = \gcd(n, k)$ *for some fixed* $k \in \mathbb{N}$,
- $f(n) = \sum_{d|n} d^k$ *for some fixed integer* $k \geq 0$,
- $f(n) = \#\{d \in \mathbb{N} \mid d \leq n, \gcd(d, n) = 1\}$.

The case of $f(n) = \sum_{d|n} d^k$ is going to be our next focus, and gets the special notation

$$\sigma_k(n) := \sum_{d|n} d^k$$

for $k \geq 0$.

Example 1.42.

$$\begin{aligned} &\sigma_0(12) = 1 + 2^0 + 3^0 + 4^0 + 6^0 + 12^0 = 6 \quad \textit{(This counts divisors)} \\ &\sigma_1(12) = 1 + 2 + 3 + 4 + 6 + 12 = 28 \quad \textit{(This sums the divisors)} \\ &\sigma_2(12) = 1 + 4 + 9 + 16 + 36 + 144 = 210. \end{aligned}$$

Theorem 1.43. *For* $n \in \mathbb{N}$, *we have the following formulae:*

$$\begin{aligned} \sigma_0(n) &= \prod_{p|n} (v_p(n) + 1)) \\ \sigma_k(n) &= \prod_{p|n} \left(\frac{p^{k(v_p(n)+1)} - 1}{p^k - 1} \right) \text{ for } k > 0 \end{aligned}$$

where the product is taken over prime factors of n.

Remark 1.44. *It is actually fine for use to take a product over all primes in the above formulae, as for* $p \nmid n$, *the corresponding factor will be* 1.

Proof. Lets start with the case of $k = 0$. The sum $\sigma_0(n) = \sum_{d|n} 1$ counts the number of divisors of n. All products will be taken over prime numbers. Writing

$$n = \prod_{p|n} p^{v_p(n)}$$

we must have that the divisors of n are exactly those natural numbers that can be written in the form

$$d = \prod_{p|n} p^{a_p}$$

where $0 \leq a_p \leq v_p(n)$, as $d|n$ if and only if $v_p(d) \leq v_p(n)$ for every prime p. Thus we have $v_p(n)+1$ choices for the value of a_p for each $p|n$. Hence, we have a total of

$$\prod_{p|n} (v_p(n)+1))$$

choices for the tuple $(a_p)_{p|n}$, each of which corresponds bijectively to a divisor of n.

For $k > 0$, we can write the divisor sum as

$$\sigma_k(n) = \sum_{\substack{0 \leq a_1 \leq v_{p_1}(n) \\ 0 \leq a_2 \leq v_{p_2}(n) \\ \vdots \\ 0 \leq a_r \leq v_{p_r}(n)}} p_1^{ka_1} p_2^{ka_2} \dots p_r^{ka_r}$$

where $\{p_1, \dots, p_r\}$ is the set of primes dividing n. We can factorise this sum as

$$\sigma_k(n) = \left(\sum_{a_1=0}^{v_{p_1}(n)} (p_1^k)^{a_1} \right) \left(\sum_{a_2=0}^{v_{p_2}(n)} (p_2^k)^{a_2} \right) \dots \left(\sum_{a_r=0}^{v_{p_r}(n)} (p_r^k)^{a_r} \right),$$

or, more succinctly, as

$$\sigma_k(n) = \prod_{p|n} \left(\sum_{a=0}^{v_p(n)} (p^k)^a \right)$$

Applying the formula for the sum of a geometric series

$$\sum_{i=0}^{N} x^i = \frac{x^{N+1}-1}{x-1}$$

we immediately obtain the claimed result. □

Example 1.45. *Recall that* $12 = 2^2 \cdot 3$. *Hence*

$$\sigma_0(12) = (2+1)(1+1) = 6$$
$$\sigma_1(12) = \left(\frac{2^3-1}{2-1}\right)\left(\frac{3^2-1}{3-1}\right) = 28$$
$$\sigma_2(12) = \left(\frac{2^6-1}{2^2-1}\right)\left(\frac{3^4-1}{3^2-1}\right) = 210.$$

Theorem 1.46. *The divisor sum functions are multiplicative.*

Proof. We leave the case of $k = 0$ as an exercise to the reader, and consider $k > 0$. Let $m, n \in \mathbb{N}$ be natural numbers and write

$$m = \prod_{p|m} p^{v_p(m)}, \quad n = \prod_{q|n} q^{v_q(n)},$$

again taking products over prime divisors. Now suppose $\gcd(m,n) = 1$. This implies that they have no common prime factors, and so the set

$$\{p \text{ prime} \mid p|mn\}$$

is equal to the disjoint union of the set of prime divisors of m and the set of prime divisors of n: if $p|mn$ then p divides exactly one of m or n. This also means that

$$v_p(mn) = v_p(m) + v_p(n) = \begin{cases} v_p(m) \text{ if } p|m, \\ v_p(n) \text{ otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \sigma(mn) &= \prod_{p|mn} \left(\frac{p^{k(v_p(mn)+1)} - 1}{p^k - 1} \right) \\ &= \prod_{p|m} \left(\frac{p^{k(v_p(mn)+1)} - 1}{p^k - 1} \right) \prod_{p|n} \left(\frac{p^{k(v_p(mn)+1)} - 1}{p^k - 1} \right) \\ &= \prod_{p|m} \left(\frac{p^{k(v_p(m)+1)} - 1}{p^k - 1} \right) \prod_{p|n} \left(\frac{p^{k(v_p(n)+1)} - 1}{p^k - 1} \right) \\ &= \sigma_k(m)\sigma_k(n). \end{aligned}$$

□

2 Modular arithmetic

Recall that an equivalence relation on a set S is a subset $R \subset S \times S$ such that

i) For every $x \in S$, $(x,x) \in R$,

ii) For every $x, y \in S$ $(x,y) \in R$ if and only if $(y,x) \in R$,

iii) For every $x, y, z \in S$, if $(x,y) \in R$ and $(y,z) \in R$, then $(x,z) \in R$.

These properties are called reflexivity, symmetry, and transitivity, respectively. We often write $x \sim y$ for $(x,y) \in R$. We may even write $x \sim_R y$ if we need to be extremely clear about what relation we are talking about.

Definition 2.1. *Let $n \in \mathbb{N}$. We define an equivalence relation, called congruence, on $\mathbb{Z}$ by*

$$a \sim b \quad \Leftrightarrow \quad n|(a-b).$$

We say that a is congruent to b modulo (or just mod) n, and often write

$$a \equiv b \ (mod\ n)$$

and call n the modulus.

Lemma 2.2. *Congruence is a well defined equivalence relation.*

Proof. We have that $a \sim a$, as $n|0 = (a - a)$ for every $a \in \mathbb{Z}$, and as $n|(a - b)$ is equivalent to there existing $k \in \mathbb{Z}$ such that $a - b = kn$, then clearly $b - a = (-k)n$, so $n|(b - a)$. Hence $a \sim b$ if and only if $b \sim a$.

Finally, if $a \sim b$ and $b \sim c$ for integers a, b, c, then $n|(a - b)$ and $n|(b - c)$. Thus n divides any integer linear combination of them. In particular

$$n|(a - c) = (a - b) + (b - c).$$

Thus, congruence is a well defined equivalence relation. □

Example 2.3. *For $n = 5$, $22 \sim 17 \sim 2 \sim -3 \sim -8$, or*

$$22 \equiv 17 \equiv 2 \equiv -3 \equiv -8 \ (mod\ 5).$$

Note that $a \equiv b \ (mod\ 1)$ for every $a, b \in \mathbb{Z}$.

We call the set

$$\overline{a} := \{b \in \mathbb{Z} \mid b \sim a\}$$

the equivalence class of a modulo n. By standard results about equivalence relations, the integers $\mathbb{Z}$ are partitioned into a set of equivalence classes. We denote this set of equivalence classes by $\mathbb{Z}/n\mathbb{Z}$.

Example 2.4. *Modulo 2, there are two equivalence classes*

$$\overline{0} = \{\ldots, -4, -2, 0, 2, 4, \ldots\},$$
$$\overline{1} = \{\ldots, -3, -1, 1, 3, 5, \ldots\}.$$

Modulo 5, there are 5 equivalence classes

$$\overline{0} = \{\ldots, -5, 0, 5, 10, \ldots\}$$
$$\overline{1} = \{\ldots, -4, 1, 6, 11, \ldots\}$$
$$\overline{2} = \{\ldots, -3, 2, 7, 12, \ldots\}$$
$$\overline{3} = \{\ldots, -2, 3, 8, 13, \ldots\}$$
$$\overline{4} = \{\ldots, -1, 4, 9, 14, \ldots\}$$

Note that there are many choices of representative for an equivalence class. Modulo 5, we have equality of the equivalence classes

$$\overline{2} = \overline{7} = \overline{-3} = \overline{222}.$$

Theorem 2.5. *The set $\mathbb{Z}/n\mathbb{Z}$ has exactly n elements, with representatives $\overline{0}, \overline{1}, \ldots, \overline{n-1}$.*

Proof. We first note that the equivalence classes

$$\overline{0}, \overline{1}, \ldots, \overline{n-1}$$

give n distinct classes. Suppose $\overline{k} = \overline{\ell}$, with $0 \leq k, l < n$. Then $n|k - \ell$, but since

$$-n < k - \ell < n$$

the only multiple of n in this range is 0. Hence $k = l$. Thus, if $k \neq \ell$ for $0 \leq k, \ell < n$, we must have $\overline{k} \neq \overline{\ell}$ in $\mathbb{Z}/n\mathbb{Z}$.

It remains to show that these are all the equivalence classes modulo n, i.e. that every integer is in one of $\overline{0}, \overline{1}, \ldots, \overline{n-1}$. Suppose $m \in \mathbb{Z}$ and consider Euclidean division by n, so that we write $m = nq + r$ with $0 \leq r < n$. Since $n|m - r$, we must have that $m \sim r$ and hence $m \in \overline{r}$. □

Definition 2.6. *We refer to the equivalence class on k in $\mathbb{Z}/n\mathbb{Z}$ as the congruence class of k mod n, and often refer to the representative of this class in $\{0, 1, \ldots, n-1\}$ as the residue or reduction of k mod n. This is not the unique choice of representative referred to as a residue. Any complete set of choice of representatives is called a residue set, though we usually reserve the term for "small" representatives.*

We actually have that $\mathbb{Z}/n\mathbb{Z}$ is much more than just a set: it is a ring! That means that there it forms a group under addition, and has a well behaved multiplication such that

$$\overline{k} \cdot (\overline{\ell} + \overline{m}) = \overline{k} \cdot \overline{\ell} + \overline{k} \cdot \overline{m}.$$

Definition 2.7. *Given $\overline{k}, \overline{\ell} \in \mathbb{Z}/n\mathbb{Z}$, define*

$$\overline{k} + \overline{\ell} := \overline{k + \ell}, \quad \overline{k} \cdot \overline{\ell} := \overline{k\ell}.$$

Proposition 2.8. *These are well defined operations: they are independent of choice of representative.*

Proof. Suppose $\overline{k} = \overline{k'}$ and $\overline{\ell} = \overline{\ell'}$. To show that these operations are well defined, it suffices to show that

$$\overline{k + \ell} = \overline{k' + \ell'}, \quad \overline{k\ell} = \overline{k'\ell'}.$$

By definition $n|k - k'$ and $n|\ell - \ell'$, and hence

$$n|\left((k - k') + (\ell - \ell')\right) = \left((k + \ell) - (k' + \ell')\right).$$

Therefore $\overline{k + \ell} = \overline{k' + \ell'}$. To see that multiplication is well defined, we need to be a bit more explicit. We can write

$$k = k' + an, \quad \ell = \ell' + bn$$

for some integers a and b. Hence

$$k\ell = (k' + an)(\ell' + bn) = k'\ell' + (a + b + abn)n$$

and so $n|k\ell - k'\ell'$, and so $\overline{k\ell} = \overline{k'\ell'}$. □

Corollary 2.9. *If $k \equiv k' \pmod n$ and $\ell \equiv \ell' \pmod n$, then*

$$k + \ell \equiv k' + \ell' \pmod n$$
$$k\ell \equiv k'\ell' \pmod n$$
$$-k \equiv -k' \pmod n$$

That final congruence is important, as it shows that we have well defined additive inverses in $\mathbb{Z}/n\mathbb{Z}$! In fact, with the exception of the cancellation of factors, we have shown that arithmetic mod n is essentially the same as arithmetic in $\mathbb{Z}$.

Example 2.10. *We have that $7 \equiv 2 \pmod 5$, and so*

$$10 = 7 + 3 \equiv 2 + 3 = 5 \equiv 0 \pmod 5.$$

Similarly, we can reduce 21 modulo 5 by factorisation

$$21 = (7)(3) \equiv 2)(3) \equiv 6 \equiv 1 \pmod 5.$$

Remark 2.11. *While $\overline{0}, \overline{1}, \ldots, \overline{n-1}$ is the standard choice of representatives for elements of $\mathbb{Z}/n\mathbb{Z}$, it can make arithmetic in $\mathbb{Z}/n\mathbb{Z}$ easier if we allow ourselves to choose more symmetric representatives. For example in $\mathbb{Z}/9\mathbb{Z}$, a handy choice of residue set is*

$$\{\overline{-4}, \overline{-3}, \overline{-2}, \overline{-1}, \overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}\}.$$

Modular arithmetic and reduction modulo n are very useful tools for studying the solutions of Diophantine equations, as integer solutions induce solutions in $\mathbb{Z}/n\mathbb{Z}$.

Definition 2.12. *Given $a \in \mathbb{Z}$, $k \in \mathbb{Z}/n\mathbb{Z}$, define $a\overline{k}$ to be the equivalence class given by adding $\overline{k}$ to itself a times:*

$$a\overline{k} = \overline{k} + \cdots + \overline{k} = \overline{ak}.$$

With this definition in mind, we can make sense of the evaluation of a a polynomial $F(x_1, \ldots, x_r)$ with integer coefficients at an element $(\overline{k}_1, \ldots, \overline{k}_r) \in (\mathbb{Z}/n\mathbb{Z})^r$ to obtain $F(\overline{k}_1, \ldots, \overline{k}_r)$, an element of $\mathbb{Z}/n\mathbb{Z}$.

Theorem 2.13. *Let $F(x_1, \ldots, x_r)$ be a polynomial with integer coefficients, and $C \in \mathbb{Z}$, and suppose there exists $k_1, \ldots, k_r \in \mathbb{Z}$ such that $F(k_1, \ldots, k_r) = C$. Then $F(\overline{k}_1, \ldots, \overline{k}_r) = \overline{C}$ in $\mathbb{Z}/n\mathbb{Z}$*

Proof. This is an immediate consequence of the definition of arithmetic operations in $\mathbb{Z}/n\mathbb{Z}$. The fancy way to say this is that the map

$$\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$
$$k \mapsto \overline{k}$$

is a ring homomorphism. □

Corollary 2.14. *If $F(\overline{k}_1, \ldots, \overline{k}_r) = \overline{C}$ has no solutions in $\mathbb{Z}/n\mathbb{Z}$, then*

$$F(x_1, \ldots, x_r) = C$$

has no integer solutions.

This is a very useful result for showing an equation has no solutions, as there are only finitely many possible solutions in $\mathbb{Z}/n\mathbb{Z}$. Finding a solution mod n doesn't strictly help us find a solution in integers, but it does help inform us about the space of possible solutions. Often, an integer solution can be found by finding a solution mod n for large enough n - this is a point we will return to later.

In practical computations, it is usually easier to express this in the $\pmod{n}$ notation: If $F(k_1, \ldots, k_r) = C$, then

$$F(k_1, \ldots, k_r) \equiv C \pmod{n}$$

and if

$$F(k_1, \ldots, k_r) \equiv C \pmod{n}$$

has no solutions, then $F(x_1, \ldots, x_r)$ has no integer solutions.

Example 2.15. *Does $x^2 + y^2 = 2711$ have integer solutions? Lets consider this as an equation in $\mathbb{Z}/4\mathbb{Z}$. In $\mathbb{Z}/4\mathbb{Z}4$ have have that*

x	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
x^2	$\overline{0}$	$\overline{1}$	$\overline{0}$	$\overline{1}$

and hence $x^2 + y^2 \in \{\overline{0}, \overline{1}, \overline{2}\}$. But $2711 = 4(677) + 3$, so $\overline{2711} = \overline{3}$. Thus, we cannot find a solution in $\mathbb{Z}/4\mathbb{Z}$ and therefore there can be no integer solutions.

Let us consider a similar example, using $\pmod{n}$ notation.

Example 2.16. *Does $5x^2 - 7y^2 = 4k + 3$ have integer solutions? Consider this modulo 4. We want to solve*

$$5x^2 - 7y^2 \equiv 4k + 3 \pmod{4}.$$

Reducing both sides modulo 4 this becomes

$$x^2 + y^2 \equiv 3 \pmod{4}$$

but we have seen that, modulo 4, $x^2 + y^2$ can only be congruent to $0, 1, 2$. Hence there can be no solutions modulo 4 and no integer solutions.

Example 2.17. *Does $x^3 + y^3 + z^3 = 31$ have integer solutions? Let us consider this equation modulo 9. Constructing a table similarly to before, we see that the only possible residues of a cube mod 9 are $\{-1, 0, 1\}$, and hence, mod 9, $x^3 + y^3 + z^3$ must be equivalent to one of $\{0, \pm 1, \pm 2, \pm 3\}$. But*

$$31 \equiv 4 \pmod{9}$$

and hence there can be no integer solutions.

A similar argument eliminates the possibility of integer solutions to $x^3 + y^3 + z^3 = 32$, but we can say nothing about $x^3 + y^3 + z^3 = 33$ from considerations modulo 9.

2.1 Cancellation and invertible elements

We have discussed addition, subtraction and multiplication in $\mathbb{Z}/n\mathbb{Z}$, but have avoided any mention of division or cancellation of common factors. This is because not every element of $\mathbb{Z}/n\mathbb{Z}$ has a multiplicative inverse, as is well illustrated by trying to solve simple linear congruence equations.

Example 2.18. *Consider the congruence* $2x \equiv 0 \pmod 6$. *If this were an equation in the integers, it would have a unique solution* $x = 0$. *But in* $\mathbb{Z}/6\mathbb{Z}$, *both* $x \equiv 0 \pmod 6$ *and* $x \equiv 3 \pmod 6$ *are solutions.*

On the other side of the spectrum, the congruence $2x \equiv 1 \pmod 6$ *has no solutions in* $\mathbb{Z}/6\mathbb{Z}$*! Unlike over* $\mathbb{Q}$, *linear equations are not guaranteed to have solutions, let alone unique ones!*

We can only guarantee a unique solution to linear equations for in certain cases, where the coefficient of x is invertible.

Definition 2.19. *An element* $\overline{k} \in \mathbb{Z}/n\mathbb{Z}$ *is called invertible if there exists* $\overline{\ell}$ *such that* $\overline{k} \cdot \overline{\ell} = \overline{1}$.

Note that if $\overline{k}$ is invertible, then its inverse $\overline{\ell}$ is unique, and we denote it by $\overline{k}^{-1}$.

Example 2.20. *In* $\mathbb{Z}/9\mathbb{Z}$, $\overline{2}$ *is invertible, with inverse* $\overline{5}$, *as*

$$\overline{2} \cdot \overline{5} = \overline{10} = \overline{1}$$

This means that, for example, the congruence

$$2x \equiv 3 \pmod 9$$

has a unique solution in $\mathbb{Z}/9\mathbb{Z}$ *given by*

$$x \equiv 5(2x) \equiv 5(3) \equiv 6 \pmod 9.$$

Theorem 2.21. *Let* $k \in \mathbb{Z}$, $n \in \mathbb{N}$. *Then* $\overline{k}$ *is invertible in* $\mathbb{Z}/n\mathbb{Z}$ *if and only if* $\gcd(k, n) = 1$.

Remark 2.22. *Note that since* $\gcd(k + an, n) = \gcd(k, n)$, *this condition is independent of the choice of representative of congruence class.*

Proof. The class $\overline{k}$ is invertible if and only if there exists an $\ell \in \mathbb{Z}$ such that $\overline{kl} = \overline{1}$ in $\mathbb{Z}/n\mathbb{Z}$, which occurs if and only if

$$k\ell \equiv 1 \pmod n$$

for some $\ell \in \mathbb{Z}$, which holds if and only if $n | k\ell - 1$ for some $\ell \in \mathbb{Z}$, which is true if and only if $k\ell = an + 1$ for some $a, \ell \in \mathbb{Z}$, which is equivalent to $k\ell - an = 1$ for some $a, \ell \in \mathbb{Z}$, which by Bézout's Theorem (Theorem 1.15) is equivalent to $\gcd(k, n) = 1$. □

Example 2.23. *Recall that* $\gcd(17,7)=1$. *Euclid's algorithm lets us compute* $\overline{7}^{-1}$ *in* $\mathbb{Z}/17\mathbb{Z}$*:*

$$\begin{aligned} 17 &= 2(7)+3, \\ 7 &= 2(3)+1. \end{aligned}$$

and hence $1 = 5(7)-2(17)$. *Thus,* $5(7) \equiv 1 \pmod{17}$*, and so* $\overline{5} = \overline{7}^{-1}$ *in* $\mathbb{Z}/17\mathbb{Z}$.

We can also characterise invertibility in terms of our ability to cancel "common factors", and hence find unique solutions to linear equations.

Theorem 2.24. *A congruence class* $\overline{k} \in \mathbb{Z}/n\mathbb{Z}$ *is invertible if and only if*

$$\overline{k} \cdot \overline{A} = \overline{k} \cdot \overline{B} \Rightarrow \overline{A} = \overline{B}$$

for all $\overline{A}, \overline{B} \in \mathbb{Z}/n\mathbb{Z}$.

Proof. If $\overline{k}$ is invertible, then

$$\begin{aligned} \overline{k} \cdot \overline{A} = \overline{k} \cdot \overline{B} &\Rightarrow \overline{k}^{-1} \cdot \overline{k} \cdot \overline{A} = \overline{k}^{-1} \cdot \overline{k} \cdot \overline{B} \\ &\Rightarrow \overline{1} \cdot \overline{A} = \overline{1} \cdot \overline{B} \\ &\Rightarrow \overline{A} = \overline{B}. \end{aligned}$$

Conversely, if

$$\overline{k} \cdot \overline{A} = \overline{k} \cdot \overline{B} \Rightarrow \overline{A} = \overline{B}$$

for all $\overline{A}, \overline{B} \in \mathbb{Z}/n\mathbb{Z}$, then the map

$$\begin{aligned} \mathbb{Z}/n\mathbb{Z} &\to \mathbb{Z}/n\mathbb{Z} \\ \overline{a} &\mapsto \overline{k} \cdot \overline{a} \end{aligned}$$

is injective. Since the domain and codomain are finite sets of the same cardinality, this means that it is a bijection and therefore surjective. In particular, this means there exists an $\ell \in \mathbb{Z}/n\mathbb{Z}$ such that $\overline{k} \cdot \overline{\ell} = \overline{1}$. □

Recall that, for p a prime, $\gcd(k,p)=1$ for all k not a multiple of p. As such, every non-zero congruence class in $\mathbb{Z}/p\mathbb{Z}$ is invertible! In fact, this characterises prime numbers.

Theorem 2.25. *Let* $n \in \mathbb{N}$ *be at least* 2. *Then the following are equivalent:*

1. *Every non-zero* $\overline{k} \in \mathbb{Z}/n\mathbb{Z}$ *is invertible,*
2. *For all* $\overline{k}, \overline{\ell} \in \mathbb{Z}/n\mathbb{Z}$, $\overline{k} \cdot \overline{\ell} = \overline{0}$ *if and only if* $\overline{k} = \overline{0}$ *or* $\overline{\ell} = \overline{0}$*,*
3. *n is a prime number.*

Proof. 1) ⇒ 2) If $\overline{k} \cdot \overline{\ell} = \overline{0}$, and $\overline{k} \neq \overline{0}$, then $\overline{k}$ is invertible, and hence

$$\overline{\ell} = \overline{k}^{-1} \cdot \overline{k} \cdot \overline{\ell} = \overline{k}^{-1} \cdot \overline{0} = \overline{0}.$$

2) $\Rightarrow$ 3) If 2) holds, then suppose $n = ab$ for some $1 \leq a, b \leq n$. Then

$$\overline{a} \cdot \overline{b} = \overline{ab} = \overline{n} = \overline{0}$$

and so $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$. Without loss of generality, we take $\overline{a} = \overline{0}$. This means that $n|a$, and since $1 \leq a \leq n$, we must have $a = n$, $b = 1$. Therefore n has no factors other than 1 and n, and so n is prime.

3) $\Rightarrow$ 1) If n is prime, $\gcd(k, n) = 1$ for all k not a multiple of n, i.e. $\overline{k} \neq \overline{0}$. Hence all non-zero $\overline{k}$ are invertible.

□

Definition 2.26. *We denote the set of invertible elements in $\mathbb{Z}/n\mathbb{Z}$ by $(\mathbb{Z}/n\mathbb{Z})^\times$, and define Euler's totient function*

$$\phi(n) := |(\mathbb{Z}/n\mathbb{Z})^\times| = \#\{1 \leq d \leq n \mid \gcd(k, n) = 1\}$$

We have seen $\phi(n)$ before! I gave this as an example of a multiplicative function, which is a property we will soon return to.

Example 2.27. *We have that $\phi(3) = 2$, $\phi(6) = 2$, $\phi(7) = 6$, $\phi(12) = 4$.*

Proposition 2.28. *The set $(\mathbb{Z}/n\mathbb{Z})^\times$ forms a group under multiplication*

Proof. That multiplication is associative follows from the associativity of multiplication of integers. Clearly $\overline{1} \in (\mathbb{Z}/n\mathbb{Z})^\times$ is an identity element, and every element of $(\mathbb{Z}/n\mathbb{Z})^\times$ has an inverse in $(\mathbb{Z}/n\mathbb{Z})^\times$, by definition. Thus, it really only remains to check that the product of two invertible elements is invertible. But if $\gcd(k, n) = 1$ and $\gcd(\ell, n) = 1$, we must have $\gcd(k\ell, n) = 1$, so the claim follows. □

2.2 The Chinese remainder theorem

One of the advantages of working with modular arithmetic when looking at Diophantine equations is that, for small n, you can very quickly check all possibilities. The downside is that modular arithmetic cannot tell us if a Diophantine equation has solutions, only when it fails to have solutions.

One situation where modular arithmetic can be useful in finding solutions is when we are looking for solutions of a certain size. Indeed, if I have a solution to a Diophantine equation in $\mathbb{Z}/N\mathbb{Z}$, there is at most one possible integer solution in $[0, N-1]$. But this isn't necessarily helpful, as we still might need to test essentially every possible solution. The goal of this section is to describe a method for solving a Diophantine equation modulo large N in terms of solutions modulo smaller n. As a side effect, we will show that the totient function is multiplicative.

2.2.1 Reduction maps between moduli

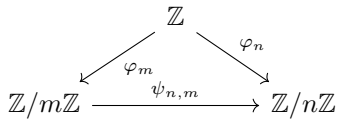
The map

$$\varphi_n : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$
$$k \mapsto \overline{k}$$

is a ring homomorphism (which is to say that it respects both the additive and multiplicative structures), often called a reduction map. I would like to construct a ring homomorphism

$$\psi_{n,m} : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$

compatible with the reduction maps in the sense that $\varphi_n = \psi_{n,m} \circ \varphi_n$, or equivalently that the following diagram commutes:



Remark 2.29. *All the notation for these maps is non-standard. To the best of my knowledge, there is no standardised notation, and so a choice was made.*

Theorem 2.30. *The map $\psi_{n,m}$ exists if and only if $n|m$.*

Proof. If such a map exists, it must map

$$\phi_m(k) = k \pmod{m} \mapsto k \pmod{n} = \phi_n(k)$$

where we use modulo notation to help keep track of our modulus. This map is well defined if and only if

$$k \equiv k' \pmod{m} \Rightarrow k \equiv k' \pmod{n}$$

which is equivalent to

$$m|(k - k') \Rightarrow n|(k - k')$$

This implication is clearly true if $n|m$. Conversely, if this implication holds, then it holds for $(k, k') = (m, 0)$ and since $m|m$, we must have $n|m$. □

Example 2.31. *We have a reduction map $\psi_{3,9} : \mathbb{Z}/9\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$ is given by*

$$\overline{0}, \overline{3}, \overline{6} \mapsto \overline{0},$$
$$\overline{1}, \overline{4}, \overline{7} \mapsto \overline{1},$$
$$\overline{2}, \overline{5}, \overline{8} \mapsto \overline{2}.$$

We cannot have a reduction map $\psi_{4,9}$, and to see this note that $\overline{13} = \overline{4}$ in $\mathbb{Z}/9\mathbb{Z}$, but

$$\overline{13} = \overline{1} \neq \overline{0} = \overline{4}$$

in $\mathbb{Z}/4\mathbb{Z}$.

2.2.2 The Chinese remainder problem

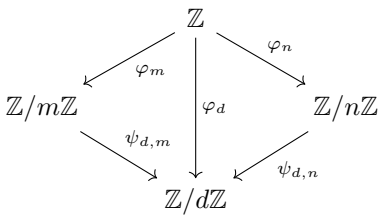
The Chinese remainder problem asks when we can solve the simultaneous congruences

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n}. \end{cases}$$

Using reduction maps, we can give a necessary condition for a solution to exist, but not immediately a sufficient condition. Let $d = \gcd(m, n)$. Then, if

$$\begin{cases} k \equiv a \pmod{m}, \\ k \equiv b \pmod{n} \end{cases}$$

for some $k \in \mathbb{Z}$, we must have that k has the same image via both sides of the diagram



which is to say that we must have

$$\psi_{d,m}(\overline{a}) = \psi_{d,m}(\varphi_m(k)) = \varphi_d(k) = \psi_{d,n}(\varphi_n(k)) = \psi_{d,n}(\overline{b}).$$

Example 2.32. *We cannot solve*

$$\begin{cases} x \equiv 4 \pmod{9}, \\ x \equiv 3 \pmod{6}, \end{cases}$$

as

$$\psi_{3,6}(\overline{3}) = \overline{0} \neq \overline{1} = \psi_{3,9}(\overline{4})$$

The easiest way to guarantee that this condition holds is to restrict our attention to the case where the moduli m and n are coprime.

Theorem 2.33. *Let $m, n \in \mathbb{N}$ be such that $\gcd(m, n) = 1$. Then the map*

$$\begin{aligned} \Phi : \mathbb{Z}/mn\mathbb{Z} &\to (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}) \\ k \pmod{mn} &\mapsto (k \pmod{m}, k \pmod{n}) = (\psi_{m,mn}(k), \psi_{n,mn}(k)) \end{aligned}$$

is a bijection, and hence there is a unique solution to

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n}. \end{cases}$$

modulo mn.

Proof. We will construct an explicit inverse to the map Φ, proving that it is a bijection and giving the solution modulo mn simultaneously.

Since m, n are coprime, there exist $u, v \in \mathbb{Z}$ such that $mu + nv = 1$ and hence

$$\begin{aligned} mu &\equiv 1 \pmod{n} \\ nv &\equiv 1 \pmod{m} \end{aligned}$$

Thus we have that

$$\Phi(mu \pmod{mn}) = (0 \pmod{m}, 1 \pmod{n})$$

and

$$\Phi(nv \pmod{mn}) = (1 \pmod{m}, 0 \pmod{n})$$

Since the reduction maps are ring homomorphisms, i.e they preserve addition and multiplication, we must have that

$$\Phi(smu + tnv \pmod{mn}) = (t \pmod{m}, s \pmod{n})$$

for all integers $s, t \in \mathbb{Z}$. Thus Φ is clearly surjective, from a set of size mn to a set of size $m \times n = mn$, and is therefore a bijection, with inverse

$$(t \pmod{m}, s \pmod{n}) \mapsto smu + tnv \pmod{mn}.$$

□

Example 2.34. *Here we will give two methods to find $k \in \mathbb{Z}$ such that*

$$\begin{cases} k \equiv 3 \ (mod\ 7), \\ k \equiv 8 \ (mod\ 17). \end{cases}$$

The first approach uses the bijection from Theorem 2.33. We have that

$$1 = 5(7) - 2(17)$$

so $mu = 35$ and $nv = -34$. Hence $\Phi^{-1}(\overline{t}, \overline{s} = \overline{35s - 34t})$. In our case this gives

$$k \equiv 8(35) - 3(34) = 178 \ (mod\ 119)$$

as a solution. We can quickly check that

$$\begin{aligned} 178 &= 25(7) + 3, \\ 178 &= 10(17) + 8. \end{aligned}$$

By reducing 178 modulo 119, we can obtain a smaller solution of $k = 59$, and the general solution will be $k = 59 + 119\ell$ for some $\ell \in \mathbb{Z}$. An alternative approach

that can be useful when you now there multiplicative inverse of one modulus

modulo the other is to note that if $k \equiv 3 \pmod 7$*, then* $k = 7\ell + 3$*. Hence, if* $k \equiv 8 \pmod{17}$*, we must have*

$$\begin{aligned} 7\ell + 3 &\equiv 8 \pmod{17} \\ 7\ell &\equiv 5 \pmod{17} \\ 35\ell &\equiv 25 \pmod{17} \\ \ell &\equiv 25 \equiv 8 \pmod{17} \end{aligned}$$

and so $\ell = 17m + 8$*. Thus* $k = 7(17m + 8) + 3 = 119m + 59$ *is the general solution. This second approach also sometimes can be used to solve simultaneous congruences where the moduli are not coprime.*

2.2.3 Chinese remainders and the totient function

We can use the (proof of the) Chinese remainder theorem to deduce a formula for Euler's totient function $\phi(n)$ in terms of the prime factors of n, via the following proposition

Proposition 2.35. *For* $m, n \in \mathbb{N}$ *coprime, the bijection*

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

induces a bijection

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

between the groups of invertible elements.

Proof. Since Φ is already known to be a bijection, it suffices to show that

$$\Phi(\overline{k}) \in (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$$

if and only if $\overline{k} \in (\mathbb{Z}/mn\mathbb{Z})^\times$.

Suppose that $\overline{k} \in (\mathbb{Z}/mn\mathbb{Z})^\times$. Then there exist $\ell \in \mathbb{Z}$ such that $k\ell \equiv 1 \pmod{mn}$ and hence $k\ell = amn + 1$ for some $a \in \mathbb{Z}$. Thus $k\ell \equiv 1 \pmod m$ and $k\ell \equiv 1 \pmod n$, and so

$$\Phi(\overline{k}) \in (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

To show the converse, it suffices to show that if $\overline{k}$ is not invertible in $\mathbb{Z}/mn\mathbb{Z}$, then $\Phi(\overline{k})$ is not invertible in $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$. If $\overline{k}$ is not invertible in $\mathbb{Z}/mn\mathbb{Z}$, then there exists a $C \in \mathbb{Z}$ such that $kC \equiv 0 \pmod{mn}$ but $C \not\equiv 0 \pmod{mn}$, as a consequence of Theorem 2.24. Thus

$$kC \equiv 0 \pmod m, \quad \text{and} \quad kC \equiv 0 \pmod n.$$

If $m|C$ and $n|C$, then $mn|C$, since $\gcd(m, n) \neq 1$. Thus

$$C \not\equiv 0 \pmod m, \quad \text{or} \quad C \not\equiv 0 \pmod n$$

and so $\overline{k}$ is not invertible in at least one of $(\mathbb{Z}/m\mathbb{Z})^\times$ or $(\mathbb{Z}/n\mathbb{Z})^\times$, and so $\Phi(\overline{k})$ is not invertible in $(\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$. Thus Φ restricts to a bijection. □

Corollary 2.36. *For $m, n \in \mathbb{N}$ with $\gcd(m,n) = 1$, $\phi(mn) = \phi(m)\phi(n)$: the totient function is multiplicative.*

Proof. Since we have a bijection

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

we have an equality of their cardinalities:

$$\begin{aligned}\phi(mn) &= |(\mathbb{Z}/mn\mathbb{Z})^\times| = |(\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times| \\ &= |(\mathbb{Z}/m\mathbb{Z})^\times| \times |(\mathbb{Z}/n\mathbb{Z})^\times| = \phi(m)\phi(n).\end{aligned}$$

□

Corollary 2.37. *Let $n = \prod_{p|n} p^{v_p(n)} \in \mathbb{N}$. Then*

$$\phi(n) = \prod_{p|n} (p-1)p^{v_p(n)-1} = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Proof. The final equality follows by factoring out $p^{v_p(n)}$ from each term in the product, so we just need to prove the first equality. Since $\gcd(p^a, q^b) = 1$ for distinct primes p and q, we can use the previous corollary to show that

$$\phi(n) = \phi\left(\prod_{p|n} p^{v_p(n)}\right) = \prod_{p|n} \phi(p^{v_p(n)})$$

and hence it suffices to show that $\phi(p^m) = (p-1)p^{m-1}$. But recall that

$$\begin{aligned}\phi(p^m) &= \#\{1 \le d \le p^m \mid \gcd(d, p^m) = 1\} \\ &= \#\{1 \le d \le p^m \mid \gcd(d, p) = 1\} \\ &= \#\{1 \le d \le p^m \mid p \nmid d\}\end{aligned}$$

There are p^m integers in the range $1, 2, \ldots, p^m$, and the multiples of p are $p, 2p, \ldots, p^{m-1}(p)$, we have that there are $p^m - p^{m-1}$ integers coprime to p^m in this range. Hence

$$\phi(p^m) = p^m - p^{m-1} = (p-1)p^{m-1}$$

as required. □

2.3 Order and primitive roots

Theorem 2.38. *Let S be a finite set and let $f : S \to S$ be any function. Pick some $s_0 \in S$ and define a sequence by $s_{m+1} = f(s_m)$. Then the sequence $\{s_m\}$ is ultimately periodic: there exist $k \in \mathbb{N}$ and $N \ge 0$ such that $s_{m+k} = s_m$ for all $m \ge N$.*

Proof. Since S is finite, eventually $f(s_m)$ must take a value we've seen before, from which the claim follows. □

Definition 2.39. *Let $\overline{k} \in \mathbb{Z}/n\mathbb{Z}$. We define a sequence by $s_0 = \overline{0}$ and $s_{m+1} = s_m + \overline{k}$, so that $s_m = \overline{mk}$ for all $m \geq 0$. The additive order of $\overline{k}$, denoted $\mathrm{AO}(\overline{k})$ is the period of the sequence $\{s_m\}$*

Example 2.40. *Let $\overline{k} = \overline{4} \in \mathbb{Z}/6\mathbb{Z}$. Then we have*

m	0	1	2	3	4	5
s_m	$\overline{0}$	$\overline{4}$	$\overline{2}$	$\overline{0}$	$\overline{4}$	$\overline{2}$

and so $\mathrm{AO}(\overline{4}) = 3$.

Theorem 2.41. *For $\overline{k} \in \mathbb{Z}/n\mathbb{Z}$, the sequence $s_m = \overline{mk}$ is periodic with period*

$$\mathrm{AO}(\overline{k}) = \frac{n}{\gcd(k,n)}.$$

Proof. As $s_n = \overline{nk} = \overline{0} = s_0$, we have that the sequence is periodic. As such, it suffices to find the minimal positive m such that $s_m = \overline{mk} = \overline{0}$. If $\overline{mk} = \overline{0}$, then mk is a multiple of n and a multiple of k, and hence a multiple of $\mathrm{lcm}(k,n)$. By minimality of m, we must have that $mk = \mathrm{lcm}(k,n)$. The result then follows on recalling that

$$\mathrm{lcm}(k,n)\gcd(k,n) = kn.$$

□

Definition 2.42. *Let $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$ and define a sequence by $t_0 = \overline{1}$, $t_{m+1} = \overline{k} \cdot t_m$ so that $t_m = \overline{k}^m$ for all $m \geq 0$. The multiplicative order, denoted by $\mathrm{MO}(\overline{k})$ is the period of t_m.*

Example 2.43. *Let $\overline{k} = \overline{4} \in \mathbb{Z}/7\mathbb{Z}$. Then*

m	0	1	2	3	4	5	6
t_m	$\overline{1}$	$\overline{4}$	$\overline{2}$	$\overline{1}$	$\overline{4}$	$\overline{2}$	$\overline{1}$

and so $\mathrm{MO}(\overline{4}) = 3$.

Theorem 2.44 (Euler's theorem)**.** *For all $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$, $\overline{k}^{\phi(n)} = \overline{1}$.*

Corollary 2.45. *We can view this as a consequence of Lagrange's theorem, which say that for any finite group G, and any $g \in G$, $g^{|G|} = 1$. $(\mathbb{Z}/n\mathbb{Z})^\times$ is a group of order $\phi(n)$, so we get the claim. We can alternatively write the elements of*

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{\overline{m}_1, \ldots, \overline{m}_{\phi(n)}\}.$$

As $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$ is invertible, the map

$$(\mathbb{Z}/n\mathbb{Z})^\times \to (\mathbb{Z}/n\mathbb{Z})^\times,$$
$$\overline{m} \mapsto \overline{km}$$

is a bijection and so

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{\overline{km}_1, \ldots, \overline{km}_{\phi(n)}\}.$$

Multiplying every element in $(\mathbb{Z}/n\mathbb{Z})^\times$ using both of these presentations, we get that

$$\overline{m}_1\overline{m}_2 \ldots \overline{m}_{\phi(n)} = (\overline{km}_1 \ldots \overline{km}_{\phi(n)} = \overline{k}^{\phi(n)} \cdot \left(\overline{m}_1 \ldots \overline{m}_{\phi(n)}\right).$$

Cancelling $\overline{m}_1 \ldots \overline{m}_{\phi(n)}$, we get that

$$\overline{k}^{\phi(n)} = \overline{1}.$$

Corollary 2.46 (Fermat's Little Theorem)**.** *For all $\overline{k} \in \mathbb{Z}/p\mathbb{Z}$, $\overline{k}^p = \overline{k}$.*

Proof. For $\overline{k} \neq \overline{0}$, Euler's theorem gives that $\overline{k}^{p-1} = \overline{1}$ and hence $\overline{k}^p = \overline{k}$. The result also holds trivially for $\overline{k} = \overline{0}$. □

Corollary 2.47. *For all $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$, the sequence $t_m = \overline{k}^m$ is periodic with period $\mathrm{MO}(\overline{k})|\phi(n)$.*

Proof. The sequence t_m is periodic as a consequence of Theorem 2.44 Let $m = \mathrm{MO}(\overline{k})$, so that $\overline{k}^m = \overline{1}$. This is the minimal such positive integer, so $m \leq \phi(n)$, and so we can write $\phi(n) = am + r$ for some $a > 0$ and $0 \leq r < m$. Then

$$\begin{aligned}\overline{1} = \overline{k}^{\phi(n)} &= \overline{k}^{am+r} \\ &= (\overline{k}^m)^a \cdot \overline{k}^r = \overline{1}^a \cdot \overline{k}^r = \overline{k}^r.\end{aligned}$$

So, by the minimality of m, we must have $r = 0$ and hence $m|\phi(n)$. □

Corollary 2.48. *For all $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$, and so all $s, t \in \mathbb{Z}$, $s \equiv t \ (mod\ \phi(n))$ implies that $\overline{k}^s = \overline{k}^t$ in $(\mathbb{Z}/n\mathbb{Z})^\times$.*

Proof. If $s \equiv t \pmod{\phi(n)}$, then $s = a\phi(n) + t$ for some $a \in \mathbb{Z}$. Hence

$$\overline{k}^s = (\overline{k}^{\phi(n)})^a \cdot \overline{k}^t = \overline{1}^a \cdot \overline{k}^t = \overline{k}^t.$$

□

Example 2.49. *What is the final digit of $(1127)^{2024}$? The last digit of any integer is determined by its value modulo 10. Note that*

$$\phi(10) = 10\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{5}\right) = 4$$

and that $2024 \equiv 0 \ (mod\ 4)$. Therefore

$$7^{2024} \equiv 7^0 \equiv 1 \ (mod\ 10)$$

and so the last digit of $(1127)^{2024}$ is 1.

2.3.1 Primitive roots

Definition 2.50. *Let $k \in \mathbb{Z}$, and $n \in \mathbb{N}$ be coprime. We call k (and $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$) a primitive root modulo n if $\mathrm{MO}(\overline{k}) = \phi(n)$.*

Remark 2.51. *A residue $\overline{k}$ is a primitive root if it is a root of $x^{\phi(n)} - 1$ in $\mathbb{Z}/n\mathbb{Z}$, and not of any $x^d - 1$ for $d < \phi(n)$.*

Example 2.52. $\mathrm{MO}(\overline{4}) = 3 < \phi(7) = 6$ *in* $(\mathbb{Z}/7\mathbb{Z})^\times$, *so* 4 *is not a primitive root modulo* 7. *But* 3 *is:*

m	1	2	3	4	5	6
$\overline{3}^m$	$\overline{3}$	$\overline{2}$	$\overline{6}$	$\overline{4}$	$\overline{5}$	$\overline{1}$

There are no primitive roots mod 8. *We would need an element of order $\phi(8) = 4$, by $\overline{k}^2 = \overline{1}$ for every $k \in \mathbb{N}$ coprime to* 8.

Remark 2.53. *Since $|(\mathbb{Z}/n\mathbb{Z})^\times| = \phi(n)$, k is a primitive root modulo n if and only if $\overline{k}$ generates $(\mathbb{Z}/n\mathbb{Z})^\times$ as a group.*

Finding a primitive root makes determining multiplicative orders slightly easier, but it is generally hard to find one, if one even exists.

Lemma 2.54. *Let $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$. Then for all $m \in \mathbb{Z}$,*

$$\mathrm{MO}(\overline{k}^m) = \frac{\mathrm{MO}(\overline{k})}{\gcd(m, \mathrm{MO}(\overline{k}))}.$$

Proof. Clearly $(\overline{k}^m)^{\mathrm{MO}(\overline{k})} = \overline{1}$, so $\mathrm{MO}(\overline{k}^m) \leq \mathrm{MO}(\overline{k})$. In fact, by division, we must have that $\mathrm{MO}(\overline{k}^m) | \mathrm{MO}(\overline{k})$.

Letting $s = \mathrm{MO}(\overline{k}^m)$, we have that $\overline{k}^{ms} = \overline{1}$, and so by division, ms is a multiple of $\mathrm{MO}(\overline{k})$, and a multiple of m. Hence it is a multiple of $\mathrm{lcm}(m, \mathrm{MO}(\overline{k}))$. By minimality of s, we must have $ms = \mathrm{lcm}(m, \mathrm{MO}(\overline{k}))$, from which the claim follows. □

Corollary 2.55. *Let $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$, and suppose $\overline{g}$ is a primitive root. Then there exists a unique $0 \leq d < \phi(n)$ such that $\overline{g}^d = \overline{k}$ and $\mathrm{MO}(\overline{k}) = \frac{\phi(n)}{\gcd(d, \phi(n))}$.*

Corollary 2.56. *If there exist primitive roots in $(\mathbb{Z}/n\mathbb{Z})^\times$, then there exist exactly $\phi(\phi(n))$ of them.*

Proof. Let $\overline{g} \in (\mathbb{Z}/n\mathbb{Z})^\times$ be a primitive root. Then $\overline{g}^d$ is a primitive root if and only if $\mathrm{MO}(\overline{g}^d) = \phi(n)$, which occurs if and only if $\gcd(d, \phi(n)) = 1$. From the definition of the totient function, there are exactly $\phi(\phi(n))$ such d such that $0 \leq d < \phi(n)$. □

2.3.2 Finding primitive roots modulo primes

Lemma 2.57. *Let p be a prime number, and let $F(x) = x^d + \overline{a}_{d-1}x^{d-1} + \cdots + \overline{a}_1 x + \overline{a}_0$ be a polynomial with coefficients in $\mathbb{Z}/p\mathbb{Z}$. Then F has at most d distinct solutions in $\mathbb{Z}/p\mathbb{Z}$.*

Proof. We first sketch that if $\overline{k}$ is solution of $F(x) = \overline{0}$, then $F(x) = (x-\overline{k})G(x)$ for some polynomial G modulo p. Since every non-zero element of $\mathbb{Z}/p\mathbb{Z}$ is invertible, we can give the usual division argument: we write $F(x) = (x - \overline{k})G(x) + R(x)$ for some polynomial R of degree less than that of $(x - \overline{k})$, i.e $R(x) = \overline{r}$. Then, if $F(\overline{k}) = 0$, we must have $\overline{r} = 0$, from which the claim follows.

Now it suffices to show that if $\overline{\ell} \neq \overline{k}$ and $F(\overline{\ell}) = \overline{0}$, then $G(\overline{\ell}) = \overline{0}$, and so we can induct on degree. But if $F(\overline{\ell}) = \overline{0}$, then

$$(\overline{\ell} - \overline{k})G(\overline{\ell}) = \overline{0}$$

and since $\overline{\ell} - \overline{k}$ is invertible for $\overline{\ell} \neq \overline{k}$, we have that $G(\overline{\ell}) = \overline{0}$. □

Lemma 2.58. *For all $n \in \mathbb{N}$, $\sum_{d|n} \phi(d) = n$.*

Proof. Consider the fractions $\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}$. When we simplify this, we get fractions $\frac{a}{d}$ with $\gcd(a, d) = 1$. There will be exactly $\phi(d)$ simplified fractions with denominator d, and so we must have $\sum_{d|n} \phi(d) = n$. □

Theorem 2.59. *For all primes p, there are $\phi(p-1) > 0$ primitive roots in $(\mathbb{Z}/p\mathbb{Z})^\times$.*

Proof. As $\phi(p) = p-1$, this will follow from Corollary 2.56 if we can show there is at least one primitive root. To show that a primitive root exists, let

$$\psi(d) = \#\{\overline{k} \in (\mathbb{Z}/p\mathbb{Z})^\times \mid \mathrm{MO}(\overline{k} = d\}$$

If this set is empty $\psi(d) = 0$. Otherwise, then for pick an element $\overline{k}$ of order d, the set

$$\{\overline{1}, \overline{k}, \overline{k}^2, \ldots, \overline{k}^{d-1}\}$$

gives d distinct roots to the equation $x^d - 1$. Corollary 2.57 tells us that these are all the roots, and hence every element of order d is a power of $\overline{k}$. And so

$$\{\overline{\ell} \in (\mathbb{Z}/p\mathbb{Z})^\times \mid \mathrm{MO}(\overline{\ell}) = d\} = \{\overline{k}^m \mid 0 \leq m < d,\ \mathrm{MO}(\overline{k}^m = d\}$$

which is in turn, via Lemma 2.54, to the set

$$\{\overline{k}^m \in (\mathbb{Z}/p\mathbb{Z})^\times \mid 0 \leq m < d,\ \gcd(d, m) = 1\}.$$

This has exactly $\phi(d)$ elements, so if $\psi(d) \neq 0$, $\psi(d) = \phi(d)$. Either way $\psi(d) \leq \phi(d)$. Hence

$$\phi(p) = |(\mathbb{Z}/p\mathbb{Z})^\times| = \sum_{d|\phi(p)} \psi(d) \leq \sum_{d|\phi(p)} \phi(d) = \phi(p)$$

where the final equality follows from Lemma 2.58. This is only possible if $\psi(d) = \phi(d)$ for every $d|\phi(p)$. In particular, there are $\psi(\phi(p)) = \phi(\phi(p)) = \phi(p-1) > 0$. □

So modulo primes, we can always find a primitive root. This can be used to help simplify our search for primitive roots modulo other natural numbers.

Lemma 2.60. *Let $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$ and let $d \in \mathbb{N}$ be such that $\overline{k}^d = \overline{1}$. Then $\mathrm{MO}(\overline{k}) = d$ if and only if for all primes $p|d$, $\overline{k}^{\frac{d}{p}} \neq \overline{1}$.*

Proof. If $\mathrm{MO}(\overline{k}) = d$, then $\overline{k}^s \neq \overline{1}$ for any $0 < s < d$, in particular for $\frac{d}{p}$.

Conversely, if $\overline{k}^{\frac{d}{p}} \neq \overline{1}$ for any $p|d$, then $\overline{k}^{\frac{d}{c}} \neq \overline{1}$ for any $c|d$. Since $\mathrm{MO}(\overline{k})|d$, we must have that $\mathrm{MO}(\overline{k}) = d$. □

Corollary 2.61. *Let $\overline{k} \in (\mathbb{Z}/n\mathbb{Z})^\times$. Then $\overline{k}$ is a primitive root if and only if for all primes $p|\phi(n)$, $\overline{k}^{\frac{\phi(n)}{p}} = \overline{1}$.*

Example 2.62. *Can we find $\mathrm{MO}(\overline{7})$ in $(\mathbb{Z}/17\mathbb{Z})^\times$. Since $\phi(17) = 16 = 2^4$, it suffices to check whether $\overline{7}^8 = \overline{1}$. We can easily check that*

$$7^8 \equiv (49)^4 \equiv (-2)^4 \equiv 16 \not\equiv 1 \ (mod\ 17)$$

and so $\mathrm{MO}(\overline{7}) = 16$. As such, 7 is a primitive root modulo 17.

What about $\mathrm{MO}(\overline{4})$ in $(\mathbb{Z}/11\mathbb{Z})^\times$. Since $\phi(11) = 10 = 2 \times 5$, we need to check $\overline{4}^2$ and $\overline{4}^5$.

$$4^2 \equiv 5 \not\equiv 1 \ (mod\ 11),$$
$$4^5 \equiv 2^{10} \equiv 1 \ (mod\ 11),$$

and so $\mathrm{MO}(\overline{4}) = 5$.

3 Quadratic reciprocity and powers mod primes

For the the rest of this section, we will consider p a fixed prime, that we will eventually assume to be odd. The goal of this section is to answer questions about how many $\overline{k} \in (\mathbb{Z}/p\mathbb{Z})^\times$ there are such that, given $t \in \mathbb{N}$, there exists $\overline{a}$ such that $\overline{k} = \overline{a}^t$?

Example 3.1. *In $(\mathbb{Z}/5\mathbb{Z})^\times$ we have that*

x	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
x^2	$\overline{1}$	$\overline{4} = \overline{-1}$	$\overline{4} = \overline{-1}$	$\overline{1}$

so we only have two possible values of non-zero squares in $\mathbb{Z}/5\mathbb{Z}$.

For talking about t^{th} powers, it is helpful to introduce the discrete logarithm.

Definition 3.2. *Given a primitive root $\overline{g} \in (\mathbb{Z}/p\mathbb{Z})^\times$, given $\log_{\overline{g}}(\overline{k})$ to be the class in $\mathbb{Z}/(p-1)\mathbb{Z}$ of $m \in \mathbb{Z}$ such that $\overline{g}^m = \overline{k}$.*

Remark 3.3. *The discrete logarithm is well defined as a class in $\mathbb{Z}/(p-1)\mathbb{Z}$ as a consequence of Corollary 2.48.*

Proposition 3.4. *The discrete logarithm behaves like a logarithm: for all $\overline{k}, \overline{\ell} \in (\mathbb{Z}/p\mathbb{Z})^\times$, and $m \in \mathbb{Z}$, we have that*

i) $\log_{\overline{g}}(\overline{k\ell}) = \log_{\overline{g}}(\overline{k}) + \log_{\overline{g}}(\overline{\ell})$,

ii) $\log_{\overline{g}}(\overline{k}^{-1}) = -\log_{\overline{g}}(\overline{k})$,

iii) $\log_{\overline{g}}(\overline{k}^m) = \overline{m}\log_{\overline{g}}(\overline{k})$,

iv)
$log_{\overline{g}}(\overline{1}) = \overline{0}$.

Proof. If $\log_{\overline{g}}(\overline{k}) = \overline{a}$ and $\log_{\overline{g}}(\overline{\ell}) = \overline{b}$), then $\overline{k} = \overline{g}^a$ and $\overline{\ell} = \overline{g}^b$. Then

$$\begin{aligned} \overline{k\ell} &= \overline{k} \cdot \overline{\ell} = \overline{g}^a \cdot \overline{g}^b = \overline{g}^{a+b}, \\ \overline{k}^m &= (\overline{g}^a)^m = \overline{g}^{ma}, \\ \overline{1} &= \overline{g}^0, \end{aligned}$$

from which all the claims follow. □

Exercise 3.5. *Let $\overline{g}$ and $\overline{h}$ be primitive roots in $(\mathbb{Z}/p\mathbb{Z})^\times$. Show that the change of base formula for logarithms holds:*

$$\log_{\overline{g}}(\overline{k}) \log_{\overline{h}}(\overline{g}) = \log_{\overline{h}}(\overline{k}).$$

Example 3.6. *Recall that $\overline{3}$ is a primitive root in $(\mathbb{Z}/7\mathbb{Z})^\times$ and that $\overline{3}^2 = \overline{2}$. Hence $\log_{\overline{3}}(\overline{2}) = \overline{2}$. We can also compute that*

$$\log_{\overline{3}}(\overline{6}) = \log_{\overline{3}}(\overline{3}) + \log_{\overline{3}}(\overline{2}) = \overline{1} + \overline{2} = \overline{3}$$

and, indeed, $\overline{3}^3 = \overline{27} = \overline{6}$.

Corollary 3.7. *Let $t \in \mathbb{Z}$ and $\overline{k} \in (\mathbb{Z}/p\mathbb{Z})^\times$. Then $\overline{k}$ is a t^{th} power if and only if $\log_{\overline{g}}(\overline{k})$ is a multiple of $\overline{t}$ in $\mathbb{Z}/(p-1)\mathbb{Z}$.*

Proof. If $\overline{k} = \overline{a}^t$, then $\log_{\overline{g}}(\overline{k}) = \overline{t}\log_{\overline{g}}(\overline{a})$ is a multiple of $\overline{t}$. Conversely, if $\log_{\overline{g}}(\overline{k}) = \overline{qt}$, then $\overline{k} = \overline{g}^{qt} = (\overline{g}^q)^t$ is a t^{th} power. □

Theorem 3.8. *Let $t \in \mathbb{Z}$. Exactly $\frac{p-1}{\gcd(t,p-1)}$ elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ are t^{th} powers.*

Proof. Given a primitive root $\overline{g}$ (which always exists for p prime), the discrete logarithm gives a bijection

$$(\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{Z}/(p-1)\mathbb{Z}$$

taking t^{th} powers to multiples of $\overline{t}$. Therefore

$$\begin{aligned}\#\{\overline{k} \in (\mathbb{Z}/p\mathbb{Z})^\times \mid \overline{k} = \overline{a}^t\} &= \#\{\overline{\ell} \in \mathbb{Z}/(p-1)\mathbb{Z} \mid \overline{\ell} = \overline{tm} \text{ for some } m \in \mathbb{Z}\} \\ &= \#\{\overline{tm} \mid \overline{m} \in \mathbb{Z}/(p-1)\mathbb{Z}\} \\ &= \mathrm{AO}(\overline{t}) = \frac{p-1}{\gcd(t, p-1)}.\end{aligned}$$

□

Corollary 3.9. *The map*

$$\begin{aligned}(\mathbb{Z}/p\mathbb{Z})^\times &\to (\mathbb{Z}/p\mathbb{Z})^\times, \\ \overline{k} &\mapsto \overline{k}^t,\end{aligned}$$

is $\gcd(t, p-1)$*-to-*1.

Proof. Denote by

$$R_{\overline{k},t} := \{\overline{a} \in (\mathbb{Z}/p\mathbb{Z})^\times) \mid \overline{a}^t = \overline{k}$$

If $R_{\overline{k},t}$ is non-empty for some $\overline{k} \neq 1$, then multiplication by $\overline{a}^{-1}$ defines a bijection between $R_{\overline{k},t}$ and $R_{\overline{1},t}$. Hence $|R_{\overline{k},t}| = |R_{\overline{1},t}|$ for all $\overline{k}$ for which this is non-empty. Thus

$$p-1 = |(\mathbb{Z}/p\mathbb{Z})^\times| = |R_{\overline{1},t}| \times |\{\overline{a}^t \mid \overline{a} \in (\mathbb{Z}/p\mathbb{Z})^\times\}| = |R_{\overline{1},t}| \frac{p-1}{\gcd(t, p-1)}$$

from which the claim follows. □

Example 3.10. *Since* $\gcd(p-1, p-1) = p-1$*, there is exactly one* $(p-1)^{th}$ *power in* $(\mathbb{Z}/p\mathbb{Z})^\times$*, which by Fermat's Little Theorem is* $\overline{1}$*.*

Corollary 3.11. *If* $\gcd(t, p-1) = 1$*, then every element of* $(\mathbb{Z}/p\mathbb{Z})^\times$ *has a unique* t^{th} *root given by* $\sqrt[t]{\overline{k}} = \overline{k}^s$ *where* $s \in \mathbb{Z}$ *is an integer such that* $\overline{st} = 1$ *in* $\mathbb{Z}/(p-1)\mathbb{Z}$*.*

Proof. Uniqueness is immediate from Corollary 3.9. It suffices to show that

$$(\overline{k}^s)^t = \overline{k}$$

but

$$\overline{k}^{st} = \overline{k}^1 = \overline{k}$$

as $\overline{st} = \overline{1}$ in $\mathbb{Z}/(p-1)\mathbb{Z}$ and so $st \equiv 1 \pmod{p-1}$. □

Example 3.12. *In* $(\mathbb{Z}/17\mathbb{Z})^\times$*, the cube root of* $\overline{7}$ *will be* $\overline{7}^{(3 \pmod{16})^{-1}}$*. To find a representative of this exponent, we note that*

$$3(11) - 16(2) = 1$$

and so $\overline{3}^{-1} = \overline{11}$ *in* $\mathbb{Z}/16\mathbb{Z}$. *Thus* $\overline{7}^{11} = \sqrt[3]{\overline{7}}$. *We can simplify this a bit*

$$\begin{aligned}\overline{7}^{11} &= (\overline{7}^2)^5 \cdot \overline{7} \\ &= \overline{-2}^5 \cdot \overline{7} \\ &= \overline{-32} \cdot \overline{7} = \overline{2} \cdot \overline{7} = \overline{14}\end{aligned}$$

and indeed $14^3 = 2744 = 161(17) + 7$.

We've noted that every element has a unique $(p-1)^{\text{th}}$ root, but what about the other extreme? What elements have square roots. From here on out, we will assume that $p \geq 3$, in particular that it is odd. If p is odd, $p-1$ is even, so $\gcd(2, p-1) = 2$. Hence, there are exactly $\frac{p-1}{2}$ squares in $(\mathbb{Z}/p\mathbb{Z})^2$, and $\frac{p-1}{2}$ non-squares.

Remark 3.13. *We usually refer to* $\overline{k}$ *as a quadratic (non-)residue modulo* p *if* $\overline{k}$ *is a (non-)square in* $(\mathbb{Z}/p\mathbb{Z})^\times$.

Definition 3.14. *For* $k \in \mathbb{Z}$ *(or* $\overline{k} \in \mathbb{Z}/p\mathbb{Z}$*), we define the Legendre symbol*

$$\left(\frac{k}{p}\right) := \begin{cases} 0 \text{ if } \overline{k} = \overline{0} \text{ in } \mathbb{Z}/p\mathbb{Z}, \\ 1 \text{ if } \overline{k} \neq \overline{0} \text{ and } \overline{k} \text{ is a square in } \mathbb{Z}/p\mathbb{Z}, \\ -1 \text{ if } \overline{k} \neq \overline{0} \text{ and } \overline{k} \text{ is not a square in } \mathbb{Z}/p\mathbb{Z}. \end{cases}$$

Fact 3.15. *The Legendre symbol has the following properties*

i) For all $a, b \in \mathbb{Z}$,

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

ii)

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 \text{ if } p \equiv 1 \pmod 4, \\ -1 \text{ if } p \equiv -1 \pmod 4, \end{cases}$$

iii)

$$\left(\frac{2}{p}\right) = \begin{cases} 1 \text{ if } p \equiv \pm 1 \pmod 8, \\ -1 \text{ if } p \equiv \pm 3 \pmod 8, \end{cases}$$

iv) If $p \neq q$ *are odd primes, then*

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

Example 3.16. *Is* -13 *a square modulo* 71*? Well, let us compute the Legendre symbol:*

$$\begin{aligned}\left(\frac{-13}{71}\right) &= \left(\frac{-1}{71}\right)\left(\frac{13}{71}\right) = (-1)^{35}(-1)^{35\times 6}\left(\frac{71}{13}\right)\\ &= -\left(\frac{6}{13}\right) = -\left(\frac{2}{13}\right)\left(\frac{3}{13}\right)\\ &= \left(\frac{3}{13}\right) = (-1)^{1\times 6}\left(\frac{13}{3}\right) = \left(\frac{1}{3}\right) = 1.\end{aligned}$$

What about 7 *modulo* 17*?*

$$\begin{aligned}\left(\frac{7}{17}\right) &= (-1)^{8\times 3}\left(\frac{17}{7}\right) = \left(\frac{3}{7}\right)\\ &= (-1)^3\left(\frac{7}{3}\right) = -\left(\frac{1}{3}\right) = -1\end{aligned}$$

so 7 *is not a square modulo* 17*.*

Let's start proving these.

Lemma 3.17. *For all* $a \in \mathbb{Z}$*,* $a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod p$*.*

Proof. If $p|a$, we are done. Otherwise, $\overline{a} \in (\mathbb{Z}/p\mathbb{Z})^\times$. Let $\overline{b} = \overline{a}^{\frac{p-1}{2}}$. Then

$$\overline{b}^2 = \overline{a}^{p-1} = \overline{1}$$

and so

$$(\overline{b} - \overline{1})(\overline{b} + \overline{1}) = \overline{0}.$$

Since p is prime, we must therefore have that $\overline{b} = \overline{\pm 1}$.

If $\overline{a} = \overline{c}^2$, then $\overline{b} = \overline{c}^{p-1} = \overline{1}$. To see the converse, note that $x^{\frac{p-1}{2}} - \overline{1}$ has at most $\frac{p-1}{2}$ roots in $(\mathbb{Z}/p\mathbb{Z})^\times$ and every square, all $\frac{p-1}{2}$ of them, is a root. So if $\overline{a}$ is not a square, then $\overline{b} = \overline{-1}$. □

Theorem 3.18. *For all* $a, b \in \mathbb{Z}$

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

and

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 \text{ if } p \equiv 1 \pmod 4, \\ -1 \text{ if } p \equiv -1 \pmod 4. \end{cases}$$

Proof. From Lemma 3.17,

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

Since these have values 0 or ± 1, and $\overline{1} \neq \overline{-1}$ for all $p \geq 3$, this lift to an equality.

Similarly

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

must lift to an equality. □

For the next part, we need a slightly convoluted proposition first. Since p is odd, we can choose our residue classes so that

$$\mathbb{Z}/p\mathbb{Z} = \left\{\overline{-\frac{p-1}{2}}, \overline{-\frac{p-3}{2}}, \ldots, \overline{-1}, \overline{0}, \overline{1}, \ldots, \overline{\frac{p-1}{2}}\right\}$$

and hence every $\overline{k} \neq \overline{0}$ can be written uniquely as $\overline{k} = \varepsilon_k \overline{s}_k$ for some $1 \leq s_k \leq \frac{p-1}{2}$ and $\varepsilon_k = \pm 1$.

Proposition 3.19. *For all $\overline{a} \in (\mathbb{Z}/p\mathbb{Z})^\times$,*

$$\left(\frac{a}{p}\right) = \prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{ta}$$

Proof. First note that for $1 \leq t_1, t_2 \leq \frac{p-1}{2}$, $\overline{s}_{t_1 a} = \overline{s}_{t_2 a}$ if and only if $t_1 = t_2$. Indeed

$$\begin{aligned} \overline{s}_{t_1 a} = \overline{s}_{t_2 a} &\Leftrightarrow \overline{t_1 a} = \pm \overline{t_2 a} \\ &\Leftrightarrow \overline{t}_1 = \pm \overline{t}_2 \end{aligned}$$

which, for $1 \leq t_1, t_2 \leq \frac{p-1}{2}$, occurs if and only if $t_1 = t_2$.

Hence, letting $S = \{1, 2, \ldots, \frac{p-1}{2}$, the map

$$\begin{aligned} S &\to S, \\ t &\mapsto \overline{s}_{ta} \end{aligned}$$

is an injective map between finite sets of the same size, and is therefore a bijection. Now, in $(\mathbb{Z}/p\mathbb{Z})^\times$

$$\begin{aligned} \overline{a}^{\frac{p-1}{2}} \prod_{t=1}^{\frac{p-1}{2}} \overline{t} &= \prod_{t=1}^{\frac{p-1}{2}} \overline{ta} \\ &= \prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{ta} \overline{s}_{ta} \\ &= \left(\prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{ta}\right) \left(\prod_{t=1}^{\frac{p-1}{2}} \overline{s}_t a\right). \end{aligned}$$

Since $\overline{t} \mapsto \overline{s}_{ta}$ is a bijection

$$\left(\prod_{t=1}^{\frac{p-1}{2}} \overline{t}\right) = \left(\prod_{t=1}^{\frac{p-1}{2}} \overline{s}_t a\right)$$

and so

$$\overline{a}^{\frac{p-1}{2}} = \prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{ta}.$$

Thus, by Lemma 3.17

$$\left(\frac{a}{p}\right) \equiv \prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{ta} \pmod{p}$$

which must lift to an equality. □

Corollary 3.20.

$$\left(\frac{2}{p}\right) = \begin{cases} 1 \text{ if } p \equiv \pm 1 \pmod 8, \\ -1 \text{ if } p \equiv \pm 3 \pmod 8, \end{cases}$$

Proof. We just need to count how many ε_{2t} are equal to -1. Since $2 \leq 2t \leq p-1$ for $1 \leq t \leq \frac{p-1}{2}$, we have that

$$\varepsilon_{2t} = \begin{cases} 1 \text{ if } 2t \leq \frac{p-1}{2}, \\ -1 \text{ if } 2t > \frac{p-1}{2}. \end{cases}$$

Write $p = 8q + r$, where $r \in \{1, 3, 5, 7\}$. Then

$$\begin{aligned} \#\{t \in \mathbb{Z} \mid \frac{p-1}{2} < 2t \leq p-1\} &= \#\{t \in \mathbb{Z} \mid \frac{p-1}{4} < t \leq \frac{p-1}{2}\} \\ &= \#\{t \in \mathbb{Z} \mid 2q + \frac{r-1}{4} < t \leq 4q + \frac{r-1}{2}\} \\ &= \begin{cases} 2q \text{ if } r = 1, \\ 2q+1 \text{ if } r = 3, \\ 2q+1 \text{ if } r = 5, \\ 2q+2 \text{ if } r = 7. \end{cases} \end{aligned}$$

Thus

$$\prod_{t=1}^{\frac{p-1}{2}} \varepsilon_{2t} = (-1)^{2q+x} = \begin{cases} 1 \text{ if } r \in \{1, 7\}, \\ -1 \text{ if } r \in \{3, 5\}. \end{cases}$$

□

3.1 Proving quadratic reciprocity

Definition 3.21. *For any $x \in \mathbb{R}$, define the floor of x by*

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leq x\}$$

Example 3.22.

$$\lfloor 3 \rfloor = \lfloor \pi \rfloor = \lfloor 3.73 \rfloor = 3$$

and

$$\lfloor -\pi \rfloor = \lfloor -3.4 \rfloor = \lfloor -3.001 \rfloor = -4.$$

Note that if we divide with remainder, $a = bt + r$ with $0 \leq r < b$, then

$$t = \left\lfloor \frac{a}{b} \right\rfloor.$$

Theorem 3.23 (Quadratic reciprocity)**.** *For any $p \neq q$ odd primes,*

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

Eisenstein. For each integer $a \in \mathbb{Z}$, perform division with remainder to get $aq = p\left\lfloor \frac{aq}{p} \right\rfloor + r_a$, with $0 \leq r_a < p$, to determine a representative

$$aq \equiv r_a \pmod{p}.$$

If $0 \leq r_a \leq \frac{p-1}{2}$, then $s_{aq} = r_a$ and $\varepsilon_{aq} = 1$, while if $\frac{p-1}{2} < r_a \leq p-1$, then $s_{aq} = p - r_a$ and $\varepsilon_{aq} = -1$. Also, note that

$$\prod_{a=1}^{\frac{p-1}{2}} \varepsilon_{aq} = \begin{cases} 1 \text{ if } \sum_{\varepsilon_{aq}=-1} 1 \equiv 0 \pmod{2}, \\ -1 \text{ if } \sum_{\varepsilon_{aq}=-1} 1 \equiv 1 \pmod{2}, \end{cases}$$

so it suffices to determine the parity of the sum of the ε_{aq} that are equal to -1. Now, modulo 2, we have that

$$\begin{aligned} \sum_{a=1}^{\frac{p-1}{2}} r_{aq} &\equiv \sum_{\varepsilon_{aq}=1} s_{aq} + \sum_{\varepsilon_{aq}=-1} p - s_{aq} \pmod{2} \\ &\equiv \sum_{\varepsilon_{aq}=1} s_{aq} + \sum_{\varepsilon_{aq}=-1} s_{aq} + \sum_{\varepsilon_{aq}=-1} 1 \pmod{2} \\ &\equiv \sum_{a=1}^{\frac{p-1}{2}} s_{aq} + \sum_{\varepsilon_{aq}=-1} 1 \pmod{2}. \end{aligned}$$

Next, we note that

$$\begin{aligned}\sum_{a=1}^{\frac{p-1}{2}} a &\equiv q \sum_{a=1}^{\frac{p-1}{2}} a \pmod 2 \\ &\equiv \sum_{a=1}^{\frac{p-1}{2}} aq \equiv \sum_{a=1}^{\frac{p-1}{2}} p\left\lfloor \frac{aq}{p} \right\rfloor + r_a \pmod 2 \\ &\equiv \sum_{a=1}^{\frac{p-1}{2}} \left\lfloor \frac{aq}{p} \right\rfloor + \sum_{a=1}^{\frac{p-1}{2}} s_{aq} + \sum_{\varepsilon_{aq}=-1} 1 \pmod 2.\end{aligned}$$

Therefore

$$\sum_{\varepsilon_{aq}=-1} 1 \equiv \sum_{a=1}^{\frac{p-1}{2}} \left\lfloor \frac{aq}{p} \right\rfloor + \sum_{a=1}^{\frac{p-1}{2}} s_{aq} + \sum_{a=1}^{\frac{p-1}{2}} a \equiv \sum_{a=1}^{\frac{p-1}{2}} \left\lfloor \frac{aq}{p} \right\rfloor \pmod 2$$

as

$$\sum_{a=1}^{\frac{p-1}{2}} s_{aq} = \sum_{a=1}^{\frac{p-1}{2}} a$$

and so their sum vanishes modulo 2.

Thus

$$\left(\frac{q}{p}\right) = (-1)^{\sum_{a=1}^{\frac{p-1}{2}} \lfloor \frac{aq}{p} \rfloor},$$

and, similarly

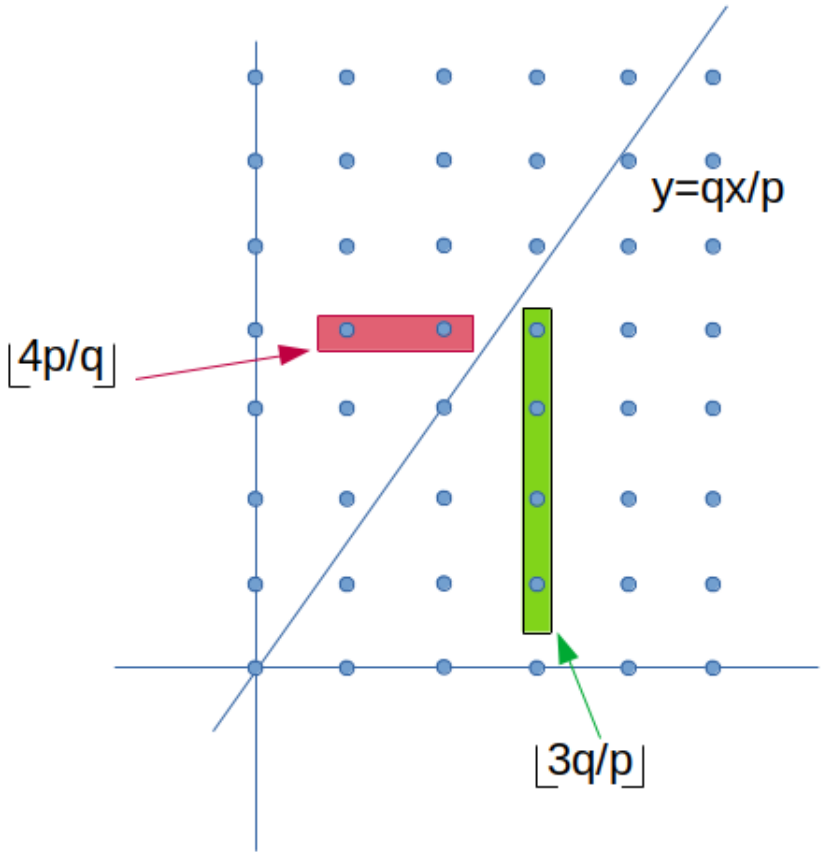
$$\left(\frac{p}{q}\right) = (-1)^{\sum_{b=1}^{\frac{q-1}{2}} \lfloor \frac{bp}{q} \rfloor}.$$

Therefore

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\sum_{a=1}^{\frac{p-1}{2}} \lfloor \frac{aq}{p} \rfloor + \sum_{b=1}^{\frac{q-1}{2}} \lfloor \frac{bp}{q} \rfloor}.$$

Now, we can think of $\left\lfloor \frac{aq}{p} \right\rfloor$ as the number of lattice points (points with integer coordinates) with positive y-coordinate and x-coordinate equal to a below the line $y = \frac{q}{p}x$, and similarly $\left\lfloor \frac{bp}{q} \right\rfloor$ is the number of lattice points with positive x-coordinate and y-coordinate equal to b to the left of the line $y = \frac{q}{p}x$, as illustrated below

Therefore, the sum in the exponent is equal to the number of lattice points with positive coordinates, such that the x-coordinate is at most $\frac{p-1}{2}$ and the y-coordinate is at most $\frac{q-1}{2}$. There are exactly $\frac{p-1}{2}\frac{q-1}{2}$ such points, and so the claim follows. □



3.2 An application of Legendre symbols to quadratic equations

Theorem 3.24. *Let $\overline{a}$, $\overline{b}$, $\overline{c}$ be elements of $\mathbb{Z}/p\mathbb{Z}$ with $\overline{a} \neq 0$. Let $\Delta = \overline{b}^2 - \overline{4ac}$. The number of solutions to $\overline{a}x^2 + \overline{b}x + \overline{c} = 0$ in $\mathbb{Z}/p\mathbb{Z}$ is*

$$\begin{cases} 2 \text{ if } \left(\frac{\Delta}{p}\right) = 1, \\ 1 \text{ if } \left(\frac{\Delta}{p}\right) = 0, \\ 0 \text{ if } \left(\frac{\Delta}{p}\right) = -1. \end{cases}$$

Proof. We can write

$$\overline{a}x^2 + \overline{b}x + \overline{c} = \overline{a}\left((x + \frac{\overline{b}}{\overline{2a}})^2 - \frac{\Delta}{(\overline{2a})^2}\right).$$

If $\Delta = \delta^2$, we can factorise this into

$$\overline{a}\left(x - \frac{-\overline{b}+\delta}{\overline{2a}}\right)\left(x - \frac{-\overline{b}-\delta}{\overline{2a}}\right).$$

Since p is prime, if this is equal to $\overline{0}$, we must have

$$x = \frac{-\overline{b}+\delta}{\overline{2a}} \quad \text{or} \quad \frac{-\overline{b}-\delta}{\overline{2a}}.$$

If these are not distinct, then $\delta = -\delta$ and so $\overline{2}\delta = \overline{0}$, which means $\delta = \overline{0}$ and hence $\Delta = \overline{0}$. Finally, if $\left(\frac{\Delta}{p}\right) = -1$, then $\frac{\Delta}{(\overline{2a})^2}$ is not a square and so there can be no solutions. □

Example 3.25. *How many solutions does $x^2 - 6x + 1$ have modulo 17? What about $x^2 - 8x + 2$?*

In the first case

$$\Delta \equiv 36 - 4 \equiv 2 - 4 \equiv -2 \ (mod\ 17)$$

and

$$\left(\frac{\Delta}{17}\right) = \left(\frac{-1}{17}\right)\left(\frac{2}{17}\right) = (-1)^8 = 1$$

so there are two distinct solutions. In the second case

$$\Delta \equiv 8^2 - 8 \equiv 13 - 8 \equiv 5 \ (mod\ 17)$$

and

$$\left(\frac{5}{17}\right) = \left(\frac{17}{5}\right) = \left(\frac{2}{5}\right) = -1$$

so there are no solutions.

Can we find the solutions to $x^2 - 6x + 1$? We know that they will be given by

$$x = \overline{2}^{-1}\left(\overline{6} - \delta\right) \quad \textit{and} \quad x = \overline{2}^{-1}\left(\overline{6} + \delta\right)$$

where $\delta^2 = \Delta = \overline{-2} = \overline{49} = \overline{7}^2$. As $\overline{2}^{-1} = \overline{9}$, we get that

$$x = \overline{15} = \overline{-2} \quad \textit{and} \quad x = \overline{8} = \overline{-9}$$

are the two solutions.

4 Fermat's Last Theorem and Pythagorean triples

The goal of this section will be to prove a special case of Fermat's Last Theorem.

Theorem 4.1. *The Diophantine equation $x^n + y^n = z^n$ has no positive integer solutions for $n > 2$.*

This is the famous result that Fermat claimed to have a "marvelous proof" of, but that this proof was too large to fit into the margins of the book in which he scribbled this claim. Given the centuries it took, it unlikely Fermat had proved this. The current proof of this is mostly attributed to Wiles, with some contributions from Taylor, who show this as a consequence of the Taniyama–Shimura conjecture, which concerns a relationship between special functions called modular forms and special curves called elliptic curves. This proof is well beyond the scope of an undergraduate course, so we will have to be satisfied with a proof of the case $n = 4$. In order to prove this, we will use the method of infinite descent, alongside a classification of Pythagorean triples.

4.1 Infinite descent

A very powerful technique for proving that an equation has no non-trivial integer solutions is the method of infinite descent. Essentially, we show that, given a non-trivial integer solution, there exists one that is smaller or closer to 0. But this process cannot repeat infinitely, as there can only be finitely many solutions between our starting solution and a trivial solution. Thus, only trivial solutions can exist. Let's see this in action, with two proofs using infinite descent.

Proposition 4.2. *The equation $x^2 + y^2 = 3z^2$ has no integer solutions other than $(x, y, z) = (0, 0, 0)$.*

Proof. We first note that if (x, y, z) is a solution, so is $(\pm x, \pm y, \pm z)$ for every choice of signs. As such, it suffices to show there are no non-negative integer solutions other than $(x, y, z) = (0, 0, 0)$.

Suppose we have a non-negative integer solution (x, y, z). By considering

$$x^2 + y^2 \equiv 3z^3 \pmod 4$$

we see that we must have

$$(x^2, y^2, z^2) \equiv (0, 0, 0) \pmod 4$$

and so x, y, z must all be even. So there exist non-negative integers a, b, $c \in \mathbb{Z}$ such that

$$x = 2a, \quad y = 2b, \quad z = 2c.$$

But this implies

$$4a^2 + 4b^2 = 12c^2$$

and hence

$$a^2 + b^2 = 3c^2$$

meaning that (a, b, c) is a smaller non-negative integer solution. We can repeat this infinitely, constructing an infinite chain of non-negative solutions. But if any of x, y, z are positive, this constructs an infinite strictly decreasing sequence of positive integers, which is impossible. Hence, we must have that $(x, y, z) = (0, 0, 0)$. □

We can also make this proof work using a mod 3 argument.

Proof. As before, we only need to consider non-negative solutions. Note also that if $z = 0$, then we must have $x = y = 0$. So suppose we have a solution with $z > 0$ and take a solution with minimal such z. Modulo 3, we must have that

$$x^2 + y^2 \equiv 0 \pmod 3$$

and hence $x \equiv y \equiv 0 \pmod 3$. Thus, there exist non-negative $a, b \in \mathbb{Z}$ such that $x = 3a$ and $y = 3b$. Hence

$$9a^2 + 9b^2 = 3z^2 \quad \Rightarrow \quad 3(a^2 + b^2) = z^2.$$

As such, $3|z^2$ and so $3|z$. Thus, there exists positive $c \in \mathbb{N}$ such that $z = 3c$, and so

$$3a^2 + 3b^2 = 9c^2 \quad \Rightarrow \quad a^2 + b^2 = 3c^2.$$

Thus, we have a solution to the equation with $0 < c < z$, contradicting the minimality of z. Therefore no solutions with non-zero z can exist, and so no non-trivial solutions can exist. □

4.2 Pythagorean triples

A triple of positive integer (a, b, c) is called a Pythagorean triple if

$$a^2 + b^2 = c^2.$$

The classical example is $3^2 + 4^2 = 5^2$. Given that any solution to

$$x^4 + y^4 = z^4$$

gives a Pythagorean triple (x^2, y^2, z^2), it will be useful to first understand the structure of Pythagorean triples.

The first thing to note is that, if we have a Pythagorean triple (a, b, c), and there exists a common divisor $d|a$, $d|b$, then $d|c$ and $(\frac{a}{d}, \frac{b}{d}, \frac{c}{d})$ gives another example of a Pythagorean triple. The same argument applies if any pair of a, b, and c have a common divisor. As such, we might as well make the simplifying assumption that a, b, and c are pairwise coprime. We call such triples *primitive* Pythagorean triples.

Theorem 4.3. *A triple (a, b, c) is a primitive Pythagorean triple if and only there exist coprime integers $u > v$, exactly one of which is odd, such that (up to swapping a and b)*

$$a = u^2 - v^2, \quad b = 2uv, \quad c = u^2 + v^2.$$

Proof. We first note that if (a, b, c) are of the given form, then

$$\begin{aligned} a^2 + b^2 &= (u^2 - v^2)^2 + 4u^2v^2 \\ &= u^4 + 2u^2v^2 + v^4 \\ &= (u^2 + v^2)^2 = c^2 \end{aligned}$$

is indeed a Pythagorean triple. It is simple to then check that they are pairwise coprime.

Conversely, suppose that (a, b, c) is a primitive Pythagorean triple. By considering both sides modulo 4, we must have that c is odd, and exactly one of a or b is odd, with the other being even. (If two were even, then we would not have a primitive triple)

Without loss of generality, assume that a is odd and b is even. Then $c - a$ and $c + a$ are even, so there exist $x, y \in \mathbb{Z}$ such that

$$c + a = 2x, \quad c - a = 2y, \quad x > y.$$

Note that $c = x + y$ and $a = x - y$. Since $\gcd(a, c) = 1$, there exist $M, N \in \mathbb{Z}$ such that

$$Ma + Nc = 1 \quad \Rightarrow \quad (M + N)x + (N - M)y = 1$$

and hence $\gcd(x, y) = 1$. Next observe that

$$4xy = (c - a)(c + a) = c^2 - a^2 = b^2.$$

Since b is even, $\frac{b}{2} \in \mathbb{N}$, and so $xy = \left(\frac{b}{2}\right)^2$ is an integer equation. For any prime p, we must have that

$$v_p(x) + v_p(y) = v_p(xy) = v_p\left(\left(\frac{b}{2}\right)^2\right) = 2v_p\left(\frac{b}{2}\right)$$

is even. Since $\gcd(x, y) = 1$, $v_p(x) \neq 0$ implies that $v_p(y) = 0$ and vice versa. Hence $v_p(x)$ and $v_p(y)$ must be even for every prime p, and so x and y are squares: there exist $u, v \in \mathbb{Z}$, such that $x = u^2$ and $y = v^2$. Clearly, we must have that $u > v$ and $\gcd(u, v) = 1$, as these hold for their squares. The claim then follows if we can show that u and v cannot have the same parity.

If u and v were both even, they would not be coprime. If u and v were both odd, then $c = u^2 + v^2$ would be even. Thus, exactly one of u and v is odd. □

Example 4.4. *Let $u = 5$ and $v = 2$. Then $a = 21$, $b = 20$ and $c = 29$, and indeed*

$$(21)^2 + (20)^2 = 441 + 400 = 841 = (29)^2.$$

4.3 Fermat's last theorem for $n = 4$

Theorem 4.5. *There are no positive integer solutions to $x^4 + y^4 = z^4$.*

Proof. We first note that, similarly to the case of Pythagorean triples, we might as well restrict our considerations to cases where x, y, and z are pairwise coprime (why?). Next we note that it would therefore suffice to prove that there are no pairwise coprime positive integer solutions to

$$x^4 + y^4 = w^2$$

as any solution to $x^4 + y^4 = z^4$ gives a solution to this via $w = z^2$.

Assume we have a solution to $x^4 + y^4 = w^2$ with $w > 0$ and minimal. Then (x^2, y^2, w) is a primitive Pythagorean triple. Without loss of generality, we can assume x^2 is odd, and so there exist coprime integers u, v such that $u > v$ and exactly one of them is odd.

As we have assumed x is odd, we have that $x^2 \equiv 1 \pmod 4$, and so

$$u^2 - v^2 \equiv 1 \pmod 4$$

which implies that u must be odd while v is even. Hence u and $2v$ are coprime. Since $y^2 = u(2v)$, we must have that both u and $2v$ must be squares.

Let $u = a^2$ and $2v = 4b^2$. Then, rearranging $x^2 = u^2 - v^2$, we see that

$$x^2 + (2b^2)^2 = (a^2)^2$$

is another primitive Pythagorean triple. Hence, there exist coprime integers c, d such that

$$x^2 = c^2 - d^2, \quad 2b^2 = 2cd, \quad a^2 = c^2 + d^2.$$

The middle equality implies that c and d are squares, so there are integers $r, s \in \mathbb{Z}$ such that $c = r^2$ and $d = s^2$, and so

$$r^4 + s^4 = a^2.$$

Thus, we have another solution to our equation. But

$$a \leq a^2 = u \leq u^2 < u^2 + v^2 = w$$

contradicting the minimality of w. □

5 Gaussian integers and sums of squares

We have just seen that the hypotenuse of any primitive right angled triangle can be written as a sum of two squares, satisfying some conditions. In this section we will drop these constraints and see what integers can be written as the sum of two squares, in general. The goal of this section will be to prove the following theorem

Theorem 5.1. *An integer $n \in \mathbb{N}$ is a sum of two squares if and only if $v_p(n)$ is even for all primes $p \equiv -1 \pmod 4$.*

Example 5.2. $2019 = 3 \times 673$ *is not a sum of two squares, but* $3 \times 2019 = 6057 = 36^2 + 69^2$.

$2020 = 2^2 \times 5 \times 101 = 24^2 + 38^2$.

$2021 = 43 \times 47$ *is not a sum of two squares.*

$2022 = 2 \times 3 \times 337$ *is not a sum of two squares.*

$2023 = 7 \times 17^2$ *is not a sum of two squares, but* $7 \times 2023 = 56^2 + 107^2$.

$2024 = 2^3 \times 11 \times 23$ *is not a sum of two squares.*

There are some results about other ways of expressing integers in terms of squares, or higher powers:

Theorem 5.3 (Legendre)**.** *An integer $n \in \mathbb{N}$ is a sum of 3 squares if and only if $n \neq 4^a(8b+7)$ for any integers $a, b \geq 0$.*

Theorem 5.4 (Lagrange)**.** *Every $n \in \mathbb{N}$ is a sum of 4 squares.*

Theorem 5.5 (Hilbert)**.** *For every $k \in \mathbb{N}$ there exists $m \in \mathbb{N}$ such that every integer $n \in \mathbb{N}$ can be written as a sum of m k^{th} powers*

This last result is closely related to the Waring problem, which asks us to find $g(k)$, the minimum such integer such that every $n \in \mathbb{N}$ can be written as $g(k)$ k^{th} powers. This is a central problem in analytic number theory, and good bounds can be given, or sometimes g(k) can be determined, using something called the circle method.

$$g(2) = 4$$
$$g(3) = 9$$
$$g(4) = 19$$
$$g(5) = 37$$

With the possible exception of Lagrange's theorem, proving these would take far too much time, so we will stick to sums of two squares for now.

5.1 Gaussian integers

The set of Gaussian integers is the subset of the complex numbers with integer real and imaginary parts:

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

Proposition 5.6. *The Gaussian integers form a ring: for any $\alpha, \beta \in \mathbb{Z}[i]$, $\alpha + \beta \in \mathbb{Z}[i]$ and $\alpha\beta \in \mathbb{Z}[i]$.*

Proof. It is easy to check that

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

and

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

are elements of $\mathbb{Z}[i]$. □

Definition 5.7. *The norm of $\alpha = a + bi \in \mathbb{Z}[i]$ is defined by*

$$\mathrm{N}(\alpha) = \alpha\overline{\alpha} = a^2 + b^2$$

where $\overline{\alpha}$ is the complex conjugate of α.

Remark 5.8. *For the Gaussian integers, the norm is just the absolute value squared. We call it a norm for a number of reasons, but primarily to have a consistent naming convenient across all "rings of integers". For example, we can define a norm on* $\mathbb{Z}[\sqrt{2}]$ *by* $\mathrm{N}(a+b\sqrt{2}) = a^2 - 2b^2$ *that will satisfy (most) properties of the norm on Gaussian integers, other than non-negativity.*

Lemma 5.9. *For any* $\alpha \in \mathbb{Z}[i]$, $\mathrm{N}(\alpha) \geq 0$ *with equality if and only if* $\alpha = 0$

Proof. $\mathrm{N}(\alpha) = a^2 + b^2$ and squares are always non-negative. We have equality if and only if $a = b = 0$. □

Lemma 5.10. *For all* $\alpha, \beta \in \mathbb{Z}[i]$, $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$.

Proof. Let $\alpha = a + bi$ and $\beta = c + di$. Then

$$\begin{aligned} N(\alpha\beta) &= (ac - bd)^2 + (ad + bc)^2 = a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 \\ &= (a^2 + b^2)(c^2 + d^2) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta) \end{aligned}$$

□

Proposition 5.11. *An integer* $n \in \mathbb{N}$ *is a sum of two squares if and only if there exists* $\alpha \in \mathbb{Z}[i]$ *such that* $\mathrm{N}(\alpha) = n$.

Proof. If $n = \mathrm{N}(\alpha)$ for $\alpha = a + bi$, then $n = a^2 + b^2$. Conversely if $n = a^2 + b^2$, then $n = \mathrm{N}(a + bi)$. □

Corollary 5.12. *Let* $m, n \in \mathbb{N}$ *both be sums of two squares. Then* mn *is a sum of two squares.*

Proof. If m and n are both the sums of two squares, then there exists $\alpha, \beta \in \mathbb{Z}[i]$ such that $\mathrm{N}(\alpha) = m$ and $\mathrm{N}(\beta) = n$. Thus

$$mn = \mathrm{N}(\alpha)\,\mathrm{N}(\beta) = \mathrm{N}(\alpha\beta)$$

is a norm and hence a sum of two squares. Explicitly, if $m = a^2 + b^2$ and $n = s^2 + t^2$, then

$$mn = (as - bt)^2 + (at + bs)^2.$$

□

Thus, to understand numbers which are sums of two squares, it suffices to understand norms of Gaussian integers. In fact, via multiplicativity, it suffices to understand norms of Gaussian "primes".

Remark 5.13. *This is not true for numbers that are sums of three squares: we have that* $2 = 1^2 + 1^2 + 0^2$ *and* $14 = 1^2 + 2^2 + 3^2$*, but* 28 *is not a sum of three squares. This means that the following approach cannot work for sums of three squares, sort of explaining why Legendre's result is less neat than for sums of two squares.*

This also gives some intuition as to why we cannot have three dimensional numbers! The sensible notion of absolute value would not be multiplicative.

Remark 5.14. *By shifting to quaternions instead of complex numbers, we can show that numbers which are sums of 4 squares are the norms of integer quaternions, and that this norm is multiplicative. Thus, a very similar approach would work for Lagrange's theorem.*

Definition 5.15. *A Gaussian integer is a called a unit if it is invertible in* $\mathbb{Z}[i]$. *That is to say that* $\alpha \in \mathbb{Z}[i]$ *is a unit if there exists* $\beta \in \mathbb{Z}[i]$ *such that* $\alpha\beta = 1$.

We denote the set of units by $\mathbb{Z}[i]^\times$.

Proposition 5.16. *A Gaussian integers* α *is a unit if and only if* $\mathrm{N}(\alpha) = 1$.

Proof. If $\alpha\beta = 1$, then

$$\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta) = 1.$$

Since $\mathrm{N}(\alpha)$ and $\mathrm{N}(\beta)$ are non-negative integers, this is only possible if $\mathrm{N}(\alpha) = \mathrm{N}(\beta) = 1$.

Conversely, if $\mathrm{N}(\alpha) = 1$, then $\alpha\overline{\alpha} = 1$, so α has a multiplicative inverse given by its complex conjugate. □

Corollary 5.17. *The set of units is* $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$.

In the integers, when defining primes, we chose them all to be positive. This wasn't strictly necessary, if just made our results cleaner. We could have chosen some primes to be negative. In the Gaussian integers, we don't have a well defined "positive", so our "primes" will only be defined up to a unit.

5.2 Division with remainder

While we cannot obtain a uniquely defined remainder in the Gaussian integers, we can still do Euclidean division!

Theorem 5.18. *Let* $\alpha, \beta \in \mathbb{Z}[i]$ *be Gaussian integers, with* $\beta \neq 0$. *Then there exist* $\gamma, \rho \in \mathbb{Z}[i]$ *such that* $\alpha = \beta\gamma + \rho$ *and* $\mathrm{N}(\rho) < \mathrm{N}(\beta)$.

Proof. We compute $\frac{\alpha}{\beta} = x + yi$ in $\mathbb{C}$, and choose $m, n \in \mathbb{Z}$ such that

$$|x - m| \leq \frac{1}{2}, \quad \text{and} \quad |y - n| \leq \frac{1}{2},$$

essentially rounding to the nearest integers. Let $\gamma = m + ni$, and $\rho = \alpha - \beta\gamma$. Then we have constructed $\gamma, \rho \in \mathbb{Z}[i]$ as in the theorem statement if we can show $\mathrm{N}(\rho) < \mathrm{N}(\beta)$.

We extend the norm to all of $\mathbb{C}$ by defining $\mathrm{N}(z) = z\overline{z} = |z|^2$ for all $z \in \mathbb{C}$. This is a multiplicative function, so

$$\begin{aligned}
\frac{\mathrm{N}(\rho)}{\mathrm{N}(\beta)} &= \mathrm{N}(\alpha - \beta\gamma)\,\mathrm{N}(\frac{1}{\beta}) \\
&= \mathrm{N}(\frac{\alpha}{\beta} - \gamma) \\
&= \mathrm{N}\left((x - m) + (y - n)i\right) \\
&= (x - m)^2 + (y - n)^2 \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.
\end{aligned}$$

Thus

$$\mathrm{N}(\rho) \leq \frac{\mathrm{N}(\beta)}{2} < \mathrm{N}(\beta).$$

□

Remark 5.19. *We are actually quite lucky that we can do Euclidean division here. If we were working in some other ring, such as $\mathbb{Z}[\sqrt{2}]$ or $\mathbb{Z}[\sqrt{-5}]$, Euclidean division can fail to be well defined!*

Example 5.20. *Let $\alpha = 11 + 27i$, $\beta = 2 + 3i$. Then*

$$\frac{\alpha}{\beta} = \frac{11+27i}{2+3i} = \frac{(11+27i)(2-3i)}{13} = \frac{103}{13} + \frac{21}{13}i \approx 8 + 2i$$

so take $\gamma = 8 + 2i$. Then

$$\begin{aligned}\rho &= \alpha - \beta\gamma \\ &= 11 + 27i - (2+3i)(8+2i) \\ &= 11 + 27i - 10 - 28i \\ &= 1 - i.\end{aligned}$$

We can check that $\mathrm{N}(\rho) = 2 < 13 = \mathrm{N}(\beta)$.

Definition 5.21. *Let $\alpha, \beta \in \mathbb{Z}[i]$. We say that $\beta|\alpha$ is there exists $\gamma \in \mathbb{Z}[i]$ such that $\alpha = \beta\gamma$.*

Remark 5.22. *For $\beta \neq 0$, this is equivalent to the remainder on dividing α by β being 0.*

The next lemma is arguably one of the most useful results in algebraic number theory, even though it is very simple.

Lemma 5.23. *For all $\alpha, \beta \in \mathbb{Z}[i]$, $\alpha|\,\mathrm{N}(\alpha)$ and if $\beta|\alpha$, then $\mathrm{N}(\beta)|\,\mathrm{N}(\alpha)$.*

Proof. As $\mathrm{N}(\alpha) = \alpha\overline{\alpha}$, $\alpha|\,\mathrm{N}(\alpha)$ by definition. If $\beta|\alpha$, there exists a Gaussian integer γ such that $\alpha = \beta\gamma$. This implies that

$$\mathrm{N}(\alpha) = \mathrm{N}(\beta\gamma) = \mathrm{N}(\beta)\,\mathrm{N}(\gamma)$$

and so $\mathrm{N}(\beta)|\,\mathrm{N}(\alpha)$. □

Definition 5.24. *We say that $\alpha, \beta \in \mathbb{Z}[i]$ are associate if $\alpha|\beta$ and $\beta|\alpha$, and write $\alpha \sim \beta$, though this is not entirely standard.*

Lemma 5.25. *Two Gaussian integers α, β are associate if and only if $\beta = \nu\alpha$ for some unit $\nu \in \mathbb{Z}[i]^\times$.*

Proof. If $\beta = \nu\alpha$ for a unit ν, then $\alpha|\beta$, and

$$\alpha = \overline{\nu}\nu\alpha = \overline{\nu}\beta$$

so $\beta|\alpha$.

Conversely, if $\alpha \sim \beta$, then $\alpha = \xi\beta$ and $\beta = \eta\alpha$ for some $\xi, \eta \in \mathbb{Z}[i]$. If $\beta = 0$, then $\alpha = 0$, so we might as well take them to be units. If $\beta \neq 0$, then

$$\beta = \eta\alpha = \eta\xi\beta$$

implies that $\eta\xi = 1$ and hence $\eta \in \mathbb{Z}[i]$. □

Definition 5.26. *Let $\alpha, \beta, \gamma \in \mathbb{Z}[i]$. We call γ a greatest common divisor of α and β if for all $\delta \in \mathbb{Z}[i]$, $\delta|\gamma$ if and only if $\delta|\alpha$ and $\delta|\beta$.*

We could also define it as a common divisor of maximal norm, but the above definition is more convenient. Following essentially the same proof as for classical integers, we can show a greatest common divisor always exists, and that most of the same properties hold, including Bézout's theorem and Gauss' lemma.

Theorem 5.27. *A greatest common divisor always exists and is unique up to multiplication by a unit.*

Sketch. As for classical integers, if $\alpha = \beta\gamma + \rho$, $\mathrm{Div}(\alpha, \beta) = \mathrm{Div}(\beta, \rho)$, so we can apply the Euclidean algorithm to compute a greatest common divisor. If $\alpha = 0$, the uniqueness is obvious, as β is a greatest common divisor, and any other greatest common divisor η must satisfy $\beta|\eta$ and $\eta|\beta$. Similarly if $\beta = 0$. If there are both non-zero, then any pair of greatest common divisors γ_1 and γ_2 must be associate, as we must have $\gamma_1|\gamma_2$ and $\gamma_2|\gamma_1$. □

Even though it is technically an abuse of notation, we will often write $\gcd(\alpha, \beta)$ for any fixed choice of greatest common divisor. In particular, we will usually write $\gcd(\alpha, \beta) = 1$ if 1 is a greatest common divisor of α and β

Corollary 5.28. *Given $\alpha, \beta \in \mathbb{Z}[i]$, the elements of $\mathbb{Z}[i]$ of the form $\alpha\xi + \beta\eta$ for $\xi, \eta \in \mathbb{Z}[i]$ are exactly the multiples of $\gcd(\alpha, \beta)$.*

Corollary 5.29. *If $\gamma|\alpha\beta$ and $\gcd(\gamma, \beta) = 1$, then $\gamma|\alpha$.*

Example 5.30. *Let us find $\gcd(11 + 27i, 2 + 3i)$. We have done the first step of Euclid's algorithm to compute*

$$11 + 27i = (8 + 2i)(2 + 3i) + (1 - i)$$

Next we note that

$$\frac{2 + 3i}{1 - i} = \frac{-1 + 5i}{2} \approx 0 + 2i$$

and that

$$2 + 3i = 2i(1 - i) + i$$

and that

$$(1 - i) = (-1 - i)(i) + 0.$$

Thus, $\gcd(11 + 27i, 2 + 3i) \sim i \sim 1$.

5.3 Gaussian primes

Definition 5.31. *An element $\pi \in \mathbb{Z}[i]$ is called irreducible if π is not a unit and, if $\pi = \alpha\beta$ is a factorisation of π in $\mathbb{Z}[i]$, then one of α or β is a unit.*

Remark 5.32. *The choice of π for irreducibles is to be consistent with Gaussian integers being represented by Greekifying their classical counterparts, in this case primes p. It is unrelated to the constant π, but it should be clear from context what is meant.*

Lemma 5.33. *If $\mathrm{N}(\pi)$ is prime in $\mathbb{Z}$, then π is irreducible in $\mathbb{Z}[i]$.*

Proof. Suppose $\mathrm{N}(\pi) = p$ is prime, and $\pi = \alpha\beta$. Then

$$\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = \mathrm{N}(\pi) = p$$

and so one of $\mathrm{N}(\alpha)$ and $\mathrm{N}(\beta)$ must equal 1, and hence one of α or β is a unit. Thus π is irreducible. □

Remark 5.34. *There are irreducible elements of non-prime norm! For example 3 has $\mathrm{N}(3) = 9$, which is not prime, but 3 is irreducible. If $3 = \alpha\beta$ has a factorisation into non-units, then*

$$\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = 9$$

and so we must have $\mathrm{N}(\alpha) = \mathrm{N}(\beta) = 3$, as neither can be 1. But 3 is not a sum of two squares, so it is not a norm. Thus, no such factorisation can exist and 3 is irreducible.

Theorem 5.35. *Every non-zero $\alpha \in \mathbb{Z}[i]$ can be factorised as*

$$\alpha = \nu\pi_1 \dots \pi_r$$

into a product of a unit ν, and irreducibles $\pi_1, \dots, \pi_r$. If

$$\alpha = \nu'\pi'_1 \dots \pi'_s$$

is another such factorisation, then $r = s$ and, up to reordering, $\pi_i \sim \pi'_i$ for each i.

Sketch. Note that if $\pi \not| \alpha$ for an irreducible π, then $\gcd(\pi, \alpha) = 1$. Thus, if $\pi|\alpha\beta$, $\pi|\alpha$ or $\pi|\beta$. With this Gaussian version of Euclid's lemma, nearly the exact proof of prime factorisation for classical integers holds. □

Example 5.36. *In $\mathbb{Z}[i]$, $2 = i(1-i)^2 = (-i)(1+i)^2$, where $(1 \pm i)$ is irreducible (as it has prime norm). These are the same factorisation up to associates and units, as $1 - i = -i(1 + i)$.*

5.4 Decomposition of primes and classification of irreducibles

Theorem 5.37. *Let $p \in \mathbb{N}$ be prime. Then:*

- *If $p \equiv 1 \pmod 4$, $p = \pi\overline{\pi}$ for some irreducible π of norm p where π and $\overline{\pi}$ are not associate,*
- *If $p \equiv -1 \pmod 4$, p is irreducible in $\mathbb{Z}[i]$,*
- *If $p = 2$, $2 = (1-i)(1+i)$*

Example 5.38. *We have already seen that 3 is irreducible. On the other hand $5 = (2-i)(2+i)$ and these are not associate.*

Remark 5.39. *For an irreducible π of norm p, its complex conjugate $\overline{\pi}$ has norm p and is therefore also irreducible.*

We will prove this in $3\frac{3}{4}$ steps: this theorem is essentially a summary of the following five lemmas.

Lemma 5.40. $2 = (1+i)(1-i)$

Lemma 5.41. *Let p be prime and suppose that p is reducible in $\mathbb{Z}[i]$. Then $p = \pi\overline{\pi}$ for some irreducible π of norm p such that $\pi = a+bi$ with $\gcd(a,b) = 1$.*

Proof. As p is reducible, we can write $p = \nu\pi_1 \dots \pi_r$ for some $r \geq 2$, ν a unit, and $\pi_1, \dots, \pi_r$ irreducibles. Thus

$$p^2 = \mathrm{N}(p) = \mathrm{N}(\nu)\,\mathrm{N}(\pi_1)\dots\mathrm{N}(\pi_r) = \mathrm{N}(\pi_1)\dots\mathrm{N}(\pi_r).$$

As $\mathrm{N}(\pi_i) \neq 1$ for any $1 \leq i \leq r$, we must have that $r = 2$, and $\mathrm{N}(\pi_1) = \mathrm{N}(\pi_2) = p$. Thus

$$\pi_1\overline{\pi}_1 = \mathrm{N}(\pi_1) = p = \mathrm{N}(\pi_2) = \pi_2\overline{\pi}_2$$

giving two factorisations nearly of the claimed form. If $\pi_1 = a + bi$ and d is a common divisor of a and b, then $d^2 | \mathrm{N}(\pi) = a^2 + b^2 = p$, so $d = 1$. Thus, the claim follows. □

Lemma 5.42. *If $p \equiv -1 \pmod 4$ is prime, p is irreducible in $\mathbb{Z}[i]$.*

Proof. Suppose p is reducible. Then by Lemma 5.41, $p = \pi\overline{\pi}$ for an irreducible $\pi = a + bi$ with $a^2 + b^2 = p$ and $\gcd(a,b) = 1$. As we cannot have $p|a$ and $p|b$, we will assume, without loss of generality, that $p \not| a$, i.e. a is invertible modulo p. Then, in $\mathbb{Z}/p\mathbb{Z}$, we have that

$$a^2 + b^2 \equiv 0 \pmod p \quad \Rightarrow \quad (ba^{-1})^2 + 1 \equiv 0 \pmod p$$

and so -1 is a square modulo p. This means that

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

and so $\frac{p-1}{2}$ is even, which means $p - 1$ is a multiple of 4, which means $p \equiv 1 \pmod 4$. □

Lemma 5.43. *If $p \equiv 1 \pmod 4$ is prime, then p is reducible in $\mathbb{Z}[i]$.*

Proof. Since $p \equiv 1 \pmod 4$, $\left(\frac{-1}{p}\right) = 1$, so there exists $c \in \mathbb{Z}$ such that $c^2 + 1 = kp$ for some $k \in \mathbb{Z}$. This implies that

$$kp = (c+i)(c-i)$$

and so $p|(c+i)(c-i)$. If p were irreducible, the Gaussian version of Euclid's lemma would imply that $p|(c+i)$ or $p|(c-i)$, and so one of $\frac{c}{p} \pm \frac{1}{p}i$ would be a Gaussian integer. In particular, $\frac{1}{p} \in \mathbb{Z}$, which is nonsense. Thus, p must be reducible. □

Lemma 5.44. *Suppose $p = \pi\overline{\pi}$ is a factorisation of a prime p into irreducibles, and $\pi \sim \overline{\pi}$. Then $p = 2$*

Proof. Write $\pi = a + bi$ where $\gcd(a, b) = 1$. As $\pi \sim \overline{\pi}$, we have that

$$\pi|\pi + \overline{\pi} = 2a \quad \text{and} \quad \pi|(-i)(\pi - \overline{\pi}) = 2b$$

and so $\pi|(2au + 2bv)$ for all $u, v \in \mathbb{Z}$. As $\gcd(a, b) = 1$, there exist u, v such that $2au + 2bv = 2$, and so $\pi|2$. By Lemma 5.23, this means that $p = \mathrm{N}(\pi)|\,\mathrm{N}(2) = 4$, and so $p = 2$. □

Having classified how primes split in $\mathbb{Z}[i]$, we have actually found every irreducible, up to multiplication by units.

Proposition 5.45. *Up to multiplication by a unit, every irreducible of $\mathbb{Z}[i]$ if one of*

i) $1 + i$,

ii) *a prime $p \in \mathbb{N}$ such that $p \equiv -1 \pmod 4$,*

iii) *π such that $\mathrm{N}(\pi)$ is a prime $p \equiv 1 \pmod 4$.*

Proof. Let π be an irreducible. Then, by Lemma 5.23, $\pi|\,\mathrm{N}(\pi)$, which is a product of prime numbers, and hence a product of irreducibles of the above type. By repeatedly applying the Gaussian version of Euclid's lemma, we must have that π divides one of these irreducibles and is therefore associate to it. □

Corollary 5.46. *Let $\pi \in \mathbb{Z}[i]$ be an irreducible. Then one of the following is true:*

i) $\mathrm{N}(\pi) = 2$ *and* $\pi \sim 1 + i$,

ii) $\mathrm{N}(\pi) = p \equiv 1 \pmod 4$, *and π is associate to exactly one of ϖ and $\overline{\varpi}$ where $p = \varpi\overline{\varpi}$,*

iii) $\mathrm{N}(\pi) = q^2$ *for a prime $q \equiv -1 \pmod 4$, and $\pi \sim q$.*

These provide some useful guidelines for factoring Gaussian integers, though it is still a bit of an art, so it is a good idea to get some practice in.

Example 5.47. *Let's factorise* $11 + 27i$ *into irreducibles. The norm is*

$$\mathrm{N}(11 + 27i) = 121 + 729 = 850 = 2 \times 5^2 \times 17.$$

Now suppose

$$11 + 27i = \nu\pi_1\pi_2 \dots \pi_r$$

is a factorisation into irreducibles and a unit ν*. Then*

$$\mathrm{N}(\pi_1) \dots \mathrm{N}(\pi_r) = 2 \times 5^2 \times 17,$$

and so we must have $r = 4$*, with one irreducible* $\pi_1 = (1 + i)$*, two irreducibles of norm* 5 *and one of norm* 17*. We can factor* $5 = (2 + i)(2 - i)$*, giving the two possible irreducibles (up to a unit) of norm* 5*. If they are both factors of* $11 + 27i$*, we would have that* $5|(11 + 27i)$*, which it clearly doesn't. Thus, the two irreducibles of norm* 5 *are either both* $(2 + i)$ *or* $(2 - i)$*. A quick calculation shows that*

$$\frac{11 + 27i}{2 + i} = \frac{49}{5} + \frac{44}{5}i \notin \mathbb{Z}[i]$$

so the irreducibles of norm 5 *must be* $2 - i$*. To determine the irreducible of norm* 17*, we compute the quotient*

$$\frac{11 + 27i}{(1 + i)(2 - i)^2} = 1 + 4i$$

which is indeed an irreducible of norm 17*. Thus*

$$11 + 27i = (1 + i)(2 - i)^2(1 + 4i)$$

is a factorisation into irreducibles.

Remark 5.48. *We can often absorb the unit into one of the irreducibles, but depending on what irreducibles of certain norm we choose, we might have to explicitly compute the unit as well. We'll see this in the next example.*

Example 5.49. *Let's factorise* $27 + 39i$ *into irreducibles. We begin with computing the norm*

$$\mathrm{N}(27 + 39i) = 2250 = 2 \times 3^2 \times 5^3.$$

From this, we see that the irreducibles must be, up to multiplication by a unit $(1 + i)$*,* 3*, and three irreducibles of norm* 5*. Choosing* $2 \pm i$ *as our possible irreducibles, we note that the irreducibles of norm* 5 *must all be* $2 + i$ *or* $2 - i$*, as otherwise* $5|27 + 39i$*. We can check that*

$$\frac{27 + 39i}{2 + i} = \frac{93}{5} + \frac{51}{5}i \notin \mathbb{Z}[i]$$

and so the irreducibles of norm 5 must be $2-i$. Because we did not compute the last irreducible by division, we need to compute the unit:

$$\nu = \frac{27+39i}{(1+i)(3)(2-i)^3} = i.$$

Therefore

$$27+39i = i(1+i)(3)(2-i)^3$$

is a factorisation into irreducibles.

We can now finally prove the theorem we began this section with:

Theorem 5.50. *An integer $n \in \mathbb{N}$ is a sum of two squares if and only if $v_p(n)$ is even for every $p \equiv -1 \pmod 4$.*

Proof. If n is a sum of two squares, then $n = \mathrm{N}(\alpha)$ for some Gaussian integer α. Factoring α into irreducibles, $\alpha = \nu\pi_1 \dots \pi_r$, we see that

$$n = \mathrm{N}(\pi_1) \dots \mathrm{N}(\pi_r)$$

where for each $1 \leq i \leq r$, $\mathrm{N}(\pi_i)$ is one of 2, p for a prime $p \equiv 1 \pmod 4$, or q^2 for a prime $q \equiv -1 \pmod 4$. Thus, for every $q \equiv -1 \pmod 4$, we have that $v_q(n)$ is even (or 0, which is even).

Conversely, if $v_q(n)$ is even for every prime $q \equiv -1 \pmod 4$, then we can write

$$n = 2^a \prod_{p \equiv 1 \pmod 4} p^{b_p} \prod_{q \equiv -1 \pmod 4} q^{2c_q}$$

for some integers $a, b_p, q_c \geq 0$. For every $p \equiv 1 \pmod 4$, we know there exists an irreducible π_p of norm p, and so we can write

$$n = \mathrm{N}\left((1+i)^a \prod_{p \equiv 1 \pmod 4} \pi_p^{b_p} \prod_{q \equiv -1 \pmod 4} q^{c_q}\right)$$

and is therefore a sum of two squares. □

Remark 5.51. *As mentioned previously, the set of sums of three squares is not closed under multiplication, so this style of argument cannot be used for classifying sums of three squares.*

In contrast, the set of sums of four squares is closed under multiplication! In fact, the set of integers that can be written as the sum of four squares is precisely the set of numbers that can be written as the norm of quaternions

$$a + bi + cj + dk$$

with integers a, b, c, d. By finding analogues of irreducible quaternions, or quaternions of prime norm for every prime, we can essentially reproduce the case of two squares.

5.5 The Gauss Circle Problem: non-examinable

Having identified which integers can be expressed as a sum of two squares, we might also ask in how many ways we can do so. There are at least 4: if $n = x^2 + y^2$, then

$$n = (-x)^2 + y^2 = x^2 + (-y)^2 = (-x)^2 + (-y)^2.$$

There could be more possibilities though

$$50 = 5^2 + 5^2 = 1^2 + 7^2,$$
$$65 = 1^2 + 8^2 = 7^2 + 4^2.$$

Can we count this? We can certainly approximate it! Imagine we have a formula $f(r)$ for the number of pairs $(a, b) \in \mathbb{Z}^2$ such that $a^2 + b^2 \leq r$ for any real r. Then the number of $(a, b) \in \mathbb{Z}^2$ such that $a^2 + b^2 = n$ would be given by the number of lattice points in the annulus

$$\{(x, y) \in \mathbb{R}^2 \mid n - \varepsilon \leq x^2 + y^2 \leq n + \varepsilon\}$$

for any $0 < \varepsilon < 1$, which would be given by $f(\sqrt{n+\varepsilon}) - f(\sqrt{n-\varepsilon})$.

The number of lattice points in the circle is approximately the area of the circle: each lattice point is the centre of a box of area 1, and these boxes cover the circle without too much excess. Thus

$$f(r) \approx \pi r^2.$$

Gauss showed that

$$|\pi^2 - f(r)| \leq 2\sqrt{2}\pi r$$

and thus

$$\begin{aligned}
|f(\sqrt{n+\varepsilon}) - f(\sqrt{n-\varepsilon})| &= |f(\sqrt{n+\varepsilon}) - \pi(n+\varepsilon) + \pi(n-\varepsilon) - f(\sqrt{n-\varepsilon}) + 2\pi\varepsilon| \\
&\leq |f(\sqrt{n+\varepsilon}) - \pi(n+\varepsilon)| + |\pi(n-\varepsilon) - f(\sqrt{n-\varepsilon})| + 2\pi\varepsilon \\
&\leq 2\pi\sqrt{2n+2\varepsilon} + 2\pi\sqrt{2n-2\varepsilon} + 2\pi\varepsilon \\
&\approx 4\pi\sqrt{2n}
\end{aligned}$$

giving a good estimate for the number of points in the annulus. Using methods from Fourier analysis and complex analysis, this bound has been improved considerably, at least for big n, to get something that behalves roughly like $\sqrt{n}$ times a constant.

6 Irrationality and continued fractions

6.1 Irrational numbers and transcendence

Definition 6.1. *A real number $x \in \mathbb{R}$ is called irrational if $x \notin \mathbb{Q}$.*

We have seen some examples of these in tutorials.

Lemma 6.2. *If $n \in \mathbb{N}$ is not the square of a natural number, $\sqrt{n}$ is irrational*

Proof. Note that $v_p(k^2) = v_p(k) + v_p(k) = 2v_p(k)$ is even, and hence the p-adic valuation of a square number is even for any prime p. Thus, if n is not a square, which is to say that if $\sqrt{n} \notin \mathbb{N}$, then $v_p(n)$ must be odd for some prime p.

Suppose $\sqrt{n} = \frac{a}{b}$. This implies that

$$a^2 = nb^2.$$

Let p be a prime such that $v_p(n)$ is odd. Then we must have

$$2v_p(a) = v_p(a^2) = v_p(nb^2) = v_p(n) + v_p(b^2) = v_p(n) + 2v_p(b).$$

But this implies $v_p(n)$ must be divisible by 2, a contraction. Therefore $\sqrt{n}$ is irrational if it is not an integer. □

This approach works for many simple irrational numbers, including more general roots $\sqrt[m]{n}$ and certain logarithms $\log_a b$. But most real numbers cannot be so easily related to rational numbers. An approach that works well for many reals, in particular those that can be written as an infinite series, is to use that rational numbers are hard to approximate by other irrational numbers.

Lemma 6.3. *If $\alpha = \frac{a}{b} \in \mathbb{Q}$ then*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{1}{|bq|}$$

for all rational numbers $\frac{p}{q} \neq \alpha$.

Proof. If $\frac{p}{q} \neq \alpha$, then

$$0 \neq \frac{a}{b} - \frac{p}{q} = \frac{aq - bp}{bq}.$$

In particular, the numerator is a non-zero integer, and hence has absolute value at least 1. Thus

$$\left|\alpha - \frac{p}{q}\right| = \left|\frac{aq - bp}{bq}\right| \geq \frac{1}{|bq|}$$

□

We can use this to show that $e = \sum_{n \geq 0} \frac{1}{n!}$ is irrational!

Proposition 6.4. *Euler's number $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ is irrational.*

Proof. Suppose $e = \frac{p}{q}$ is rational, and assume, without loss of generality, that $q > 0$. Note that the partial sum $\sum_{n=0}^{m} \frac{1}{n!}$ is a rational number, and we can take the denominator to be $q_m := m!$. Define p_m by

$$\frac{p_m}{q_m} = \sum_{n=0}^{m} \frac{1}{n!}.$$

Clearly $\frac{p_m}{q_m} \neq \frac{p}{q} = e$, as they differ by a positive number that is at least $\frac{1}{(m+1)!}$. Therefore, by Lemma 6.3, we must have that

$$\left| e - \frac{p_m}{q_m} \right| \geq \frac{1}{qq_m} = \frac{1}{q \cdot m!}$$

for all $m \geq 0$. But, we can give an upper bound for this difference where we use partial fraction expansion to telescope the final sum. Hence, we must have that

$$\frac{1}{q \cdot m!} \leq \left| e - \frac{p_m}{q_m} \right| \leq \frac{2}{(m+1)!}$$

for all $m \geq 0$, and hence $m+1 \leq 2q$ for all $m \geq 0$. But q is fixed, so this cannot be true for large n. Thus, e cannot be rational. □

Via minimal polynomials, we can give a more precise degree of irrationality to irrational numbers.

Definition 6.5. *Given $\alpha \in \mathbb{R}$, we call α transcendental if $f(\alpha) \neq 0$ for every non-zero polynomial $f \in \mathbb{Q}[x]$. Otherwise, we call α algebraic. For algebraic α, we define the minimal polynomial of α to be the unique monic (non-zero) polynomial $f \in \mathbb{Q}[x]$ of minimal degree such that $f(\alpha) = 0$. We call α algebraic of degree equal to the degree of f.*

Example 6.6. *$\sqrt{2}$ is algebraic of degree 2. $\sqrt[3]{3}$ is algebraic of degree 3. The constants π and e are transcendental.*

The following result should be familiar to you from earlier courses but we will repeat the proof here, for sake of completeness. Throughout this proof, and most of this course, the term "polynomial" will implicitly mean "polynomial with rational coefficients".

Lemma 6.7. *Minimal polynomials exist, and are unique and irreducible.*

Proof. For α algebraic, the set

$$S = \{f \in \mathbb{Q}[x] \mid f(\alpha) = 0,\ f(x) \neq 0\}$$

is non-empty, and the degree of its elements is bounded below. Hence there exists polynomials of minimal degree, which we can take to be monic, as S is closed under multiplication by non-zero rationals.

We claim that there is a unique such polynomial. Suppose f, $g \in \mathbb{Q}[x]$ are monic polynomials of minimal degree contained in S. As f and g are monic, and of the same degree, $h(x) = f(x) - g(x)$ is a polynomial of smaller degree such that

$$h(\alpha) = f(\alpha) - g(\alpha) = 0 - 0 = 0.$$

But we assumed f and g had minimal degree among non-zero polynomials vanishing at α. Thus, we must have $h(x) = 0$, i.e. $f(x) = g(x)$ is the unique monic polynomial of minimal degree vanishing at α.

To see that this polynomial must be irreducible, note that if $f(x) = a(x)b(x)$, then

$$0 = f(\alpha) = a(\alpha)b(\alpha)$$

and so one of $a(x)$ or $b(x)$ vanishes at α. But f was minimal among all such polynomials, so this can only occur if one of a or b is a constant multiple of f and the other is a constant. Thus f has no non-trivial factorisations, and is therefore irreducible. □

An immediate consequence of this definition is the following.

Remark 6.8. *A real number $\alpha \in \mathbb{R}$ is rational if and only if it is algebraic of degree 1.*

An important result is the existence of transcendental numbers. We will give two proofs of this: one due to Cantor, which is highly non-constructive, and one due to Liouville, in which we will explicitly construct a transcendental number.

Proposition 6.9. *Transcendental numbers exists.*

Cantor. The set $\mathbb{Q}$ of rational numbers is countable, and hence the set of polynomials with rational coefficients of degree at most k is countable, as it is in bijection with $\mathbb{Q}^k$. Hence, the set of algebraic numbers of degree at most k is countable, as there are at most k algebraic number associated to each such polynomial. Hence, the set of algebraic numbers, which is the countable union over $k \geq 0$ of the countable sets of algebraic numbers of degree at most k, is countable. But $\mathbb{R}$ is uncountable, so there must be elements in $\mathbb{R}$ which are not algebraic. □

The more constructive proof relies on the following lemma.

Lemma 6.10 (Liouville)**.** *Let α be an algebraic number of degree $k > 1$, i.e. an irrational algebraic number. Then there exists $C > 0$ such that*

$$\left|\alpha - \frac{p}{q}\right| > \frac{C}{|q|^k}$$

for all $\frac{p}{q} \in \mathbb{Q}$.

Proof. As we can always take the denominator of a rational number to be positive, we will omit the absolute value in the lower bound. Now, note that if

$$\left|\alpha - \frac{p}{q}\right| > 1$$

then

$$\left|\alpha - \frac{p}{q}\right| > \frac{1}{q^k}$$

so if we can find a similar bound for $\left|\alpha - \frac{p}{q}\right| \leq 1$, we can choose our constant so that both cases are covered. As such, assume $\frac{p}{q}$ is such that $\left|\alpha - \frac{p}{q}\right| \leq 1$, and

let $f(x)$ be the minimal polynomial of α, rescaled to have integer coefficients. From the mean value theorem, there exists ξ between $\frac{p}{q}$ and α such that

$$f'(\xi) = \frac{f(\frac{p}{q}) - f(\alpha)}{\frac{p}{q} - \alpha} = \frac{f(\frac{p}{q})}{\frac{p}{q} - \alpha}.$$

Since α is irrational, and its minimal polynomial is irreducible, $f(\frac{p}{q}) \neq 0$, and so $f'(\xi) \neq 0$. Thus

$$\left|\alpha - \frac{p}{q}\right| = \frac{|f(\frac{p}{q})|}{|f'(\xi)|} = \frac{1}{|f'(\xi)|}\frac{M}{q^k} \geq \frac{1}{|f'(\xi)|q^k}$$

where M is the (non-zero) integer obtained by taking a common denominator in $f(\frac{p}{q})$.

From our assumption on $\frac{p}{q}$, we must have that $\xi \in [\alpha - 1, \alpha + 1]$, so let

$$C' = \min_{\xi \in [\alpha-1, \alpha+1]} \frac{1}{|f'(\xi)}$$

to obtain a bound

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{C'}{q^k}$$

for all $\left|\alpha - \frac{p}{q}\right| \leq 1$. Then letting $C = \min(1, C')$, we obtain the desired bound. □

This essentially says that the only real numbers that can be very well approximated by rationals with "small" denominators are rational numbers and transcendental numbers.

Proposition 6.11. *The number $\beta = \sum_{k=1}^{\infty} \frac{1}{10^{k!}}$ is transcendental.*

Proof. Suppose first that β is algebraic of degree $k > 1$, and let $q_m = 10^{m!}$, define p_m by

$$\frac{p_m}{q_m} = \sum_{k=1}^{m} \frac{1}{10^{k!}}.$$

Then

$$\begin{aligned}\left|\beta - \frac{p_m}{q_m}\right| &= \sum_{k=m+1}^{\infty} \frac{1}{10^{k!}} \\ &< \frac{1}{10^{(m+1)!}}\left(1 + \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \cdots\right) \\ &= \frac{10}{9 \cdot 10^{(m+1)!}}\end{aligned}$$

since

$$10^{-(m+k)!} = 10^{-(m+1)!} \cdot 10^{(m+2)(m+3)\ldots(m+k)} < 10^{-(m+1)!}10^{1-k}.$$

Thus, by Liouville's lemma, there exists $C > 0$ such that

$$\frac{C}{q_m^k} < \left|\beta - \frac{p_m}{q_m}\right| < \frac{10}{9q_m^{m+1}}$$

for all $m \geq 1$. Hence

$$q_m^{m+1-k} < \frac{10}{9C}$$

for all $m \geq 1$. But this is impossible, as the left hand side is unbounded. Hence β is rational or transcendental.

If $\beta = \frac{p}{q}$ is rational, then $\frac{p}{q} \neq \frac{p_m}{q_m}$ and so

$$\frac{10}{9q_m^{m+1}} > \|\beta - \frac{p_m}{q_m}\| > \frac{1}{qq_m}$$

for all m, but this implies that

$$q_m^m < \frac{10v}{9}$$

which cannot hold for large m. Thus β is transcendental. □

6.2 Continued fractions and good approximations

We have seen that, for a fixed denominator, algebraic irrationals are hard to approximate. We can make this much more precise, and give a recipe for determining the best approximation via continued fractions.

Definition 6.12. *Let $a_0 \in \mathbb{Z}$ and $a_1, a_2, a_3, \ldots \in \mathbb{N}$ be a (possibly finite) sequence of integers. We define the continued fraction*

$$a_0, a_1, \ldots, a_n] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}}$$

Example 6.13. *The finite sequence $2, 3, 5, 7$ corresponds to the continued fraction*

$$[2, 3, 5, 7] = 2 + \cfrac{1}{3 + \cfrac{1}{5 + \cfrac{1}{7}}} = \frac{266}{115}$$

We will also, slightly abusively, extend this notation to allow the final entry a_n to be any positive real number. This gives us a bit of wiggle room with the length of finite continued fractions, as with this relaxation:

$$[a_0, a_1, \ldots, a_{n-1}, a_n] = [a_0, a_1, \ldots, a_{n-1} + a_n^{-1}].$$

To any real number, we can associate a canonical sequence of integers, whose finite continued fractions give surprisingly good approximations. In this context, canonical essentially means that everyone agrees this is the best way to do it, though we will later see that this is essentially the only way.

Definition 6.14. *Given a real number $x \in \mathbb{R}$, we define a (possibly finite) sequence of integers by defining $x_0 = x$, $a_n := \lfloor x_n \rfloor$, and*

$$x_{n+1} = \frac{1}{x_n - a_n}$$

for all $n \geq 0$, unless $x_n = a_n$ for some n, then we terminate the process with a finite sequence. We call this sequence the continued fraction expansion of x.

Remark 6.15. *Since $0 \leq x_n - a_n < 1$ for every n for which these are defined, $a_n \geq 1$ for every $n > 0$ for which it is defined.*

Example 6.16. *For $x = \pi$, we find that $x_0 = \pi$, so $a_0 = 3$. Then $x_1 = \frac{1}{0.1415\ldots} = 7.065\ldots$, so $a_1 = 7$. Similarly $x_2 = 15.9965\ldots$, so $a_2 = 15$. Next $x_3 = 1.0034\ldots$, and $a_3 = 1$, and $x_4 = 292.6354\ldots$, so $a_4 = 292$. Computing the associated continued fractions, we find*

$$\begin{aligned} [a_0] &= 3, \\ [a_0, a_1] &= \frac{22/7}{}, \\ [a_0, a_1, a_2] &= \frac{333}{106}, \\ [a_0, a_1, a_2, a_3] &= \frac{355}{113}, \end{aligned}$$

and so on. Note that these are all decent approximations of π, and that, since a_4 is quite large, the next fraction isn't substantially different.

Our next major goal will be to show that the sequence computed by this process determines x, specifically that

$$\lim_{n \to \infty} [a_0, a_1, \ldots, a_n] = x$$

where we interpret this as an eventually constant sequence if $a_0, a_1, \ldots$ is a finite sequence.

6.2.1 Convergence of continued fractions

We will first handle the case of finite continued fractions.

Theorem 6.17. *The sequence $a_0, a_1, a_2, \ldots$ associated to $x \in \mathbb{R}$ is finite if and only if $x \in \mathbb{Q}$.*

Proof. Suppose $x = \frac{A}{B} \in \mathbb{Q}$. Then $x_0 = \frac{A}{B}$, and $a_0 = \lfloor \frac{A}{B} \rfloor$, which we recall is the quotient in division with remainder. Thus,

$$A = a_0 B + R \text{ with } 0 \leq R < B.$$

Hence, if $x_0 \neq a_0$, $x_1 = \frac{B}{R}$. Repeating this, we find that the continued fraction expansion of x is given by the quotients in the Euclidean algorithm for computing $\gcd(A, B)$. This algorithm terminates when we get remainder 0, which corresponds to $x_n = a_n$, and thus we obtain a finite expansion.

To see the converse, note that if we get a finite continued fraction expansion, $x_n = a_n \in \mathbb{Q}$ for some n. Hence,

$$x_{n-1} = \frac{1}{x_n} + a_{n-1} \in \mathbb{Q}$$

and similarly, $x_{n-2} \in \mathbb{Q}$, $\ldots$, $x_0 = x \in \mathbb{Q}$. □

Example 6.18. *F0r $x = \frac{27}{11}$, we find that $a_0 = 2$,*

$$x_1 = \frac{1}{\frac{27}{11} - \frac{22}{11}} = \frac{11}{5},$$

so $a_1 = 2$,

$$x_2 = \frac{1}{\frac{11}{5} - \frac{10}{5}} = 5 = a_2,$$

and we indeed get a finite sequence.

Theorem 6.19. *For all $n \geq 0$, $[a_0, a_1, \ldots, a_{n-1}, x_n] = x$.*

Proof. We proceed by induction. It is clearly true for $n = 0$, so suppose it is true for some $n \geq 0$. Then

$$\begin{aligned}[a_0, a_1, \ldots, a_n, x_{n+1}] &= [a_0, a_1, \ldots, a_n + x_{n+1}^{-1}] \\ &= [a_0, a_1, \ldots, x_n] = x\end{aligned}$$

by our induction hypothesis. Thus, the claim holds. □

In particular, when $x_n = a_n$ in a finite continued fraction expansion, we obtain the following corollary.

Corollary 6.20. *Every $x \in \mathbb{Q}$ can be express as a finite continued fraction.*

Example 6.21.

$$\frac{27}{11} = 2 + \cfrac{1}{2 + \cfrac{1}{5}} = [2, 2, 5].$$

For infinite sequences, we introduce some auxiliary fractions that will let us better discuss the convergence of the associated continued fraction.

Definition 6.22. *To a sequence $a_0 \in \mathbb{Z}$, $a_1, a_2, \ldots \in \mathbb{N}$, we define two sequences*

$$p_{-2}, p_{-1}, p_0, p_1, \ldots \in \mathbb{Z},$$
$$q_{-2}, q_{-1}, q_0, q_1, \ldots \in \mathbb{Z},$$

by

$$p_{-2} = 0,\ p_{-1} = 1 \text{ and } p_n = a_n p_{n-1} + p_{n-2},$$
$$q_{-2} = 1,\ q_{-1} = 0 \text{ and } q_n = a_n q_{n-1} + q_{n-2}$$

for all $n \geq 0$. The ratios $\frac{p_n}{q_n}$ for $n \geq 0$ are called convergents of the sequence.

Note that $q_n > 0$ for all $n \geq 0$, and that both p_n and q_n are increasing. In fact, as $a_n \geq 1$ for all $n \geq 1$, both sequences grow at least as fast as the Fibonnaci numbers. Explicitly, $p_0 = a_0$, $q_0 = 1$, $p_1 = a_1 a_0 + 1$, $q_1 = a_1$, and so on.

In all that follows, we will assume we have a (possibly finite) sequence $a_0 \in \mathbb{Z}$, $a_1, a_2, \ldots \in \mathbb{N}$ given to us, rather than explicitly state this as part of every result.

Theorem 6.23. *For all $n \geq 0$, finite continued fractions are given by the corresponding convergent:*

$$[a_0, a_1, \ldots, a_n] = \frac{p_n}{q_n}$$

Proof. We proceed by induction. We first extend the notion of convergents to sequences $a_0 \in \mathbb{Q}$, and $a_1, \ldots, a_n$ positive real numbers. The claim is clearly true for $n = 0$, so suppose it is true for some $n \geq 0$. Given a sequence $a_0, a_1, \ldots, a_n, a_{n+1}$, consider the auxiliary sequences

$$a'_0 = a_0,\ a'_1 = a_1, \ldots, a'_{n-1} = a_{n-1},\ a'_n = a_n + a_{n+1}^{-1},$$
$$p'_0 = p_0,\ p'_1 = p_1, \ldots, p'_{n-1} = p_{n-1}, p'_n = a'_n p'_{n-1} + p'_{n-2} = p_n + \frac{p_{n-1}}{a_{n+1}},$$
$$q'_0 = q_0,\ q'_1 = q_1, \ldots, q'_{n-1} = q_{n-1}, q'_n = a'_n q'_{n-1} + q'_{n-2} = q_n + \frac{q_{n-1}}{a_{n+1}},$$

which agree with our original sequences the n^{th} term. By the induction hypothesis

$$[a_0, \ldots, a_n, a_{n+1}] = [a'_0, \ldots, a'_n] = \frac{p'_n}{q'_n} = \frac{a_{n+1} p_n + p_{n-1}}{a_{n+1} q_n + q_{n-1}} = \frac{p_{n+1}}{q_{n+1}}$$

□

A neat corollary of this proof is the following result about continued fractions with one real entry.

Corollary 6.24. *For all $y > 0$ and $n \geq 0$*

$$[a_0, a_1, \ldots, a_n, y] = \frac{y p_n + p_{n-1}}{y q_n + q_{n-1}}$$

By tracing back the recurrence relation, we can obtain a bit more information about the relationship between nearby convergents.

Theorem 6.25. *For all $n \geq 0$, $q_n p_{n-1} - q_{n-1} p_n = (-1)^n$ and $q_n p_{n-2} - q_{n-2} p_n = (-1)^{n-1} a_n$.*

Proof. Let $M_n = \left(\begin{smallmatrix} a_n & 1 \\ 1 & 0 \end{smallmatrix}\right)$ for every $n \geq 0$. Then

$$\begin{pmatrix} q_n & p_n \\ q_{n-1} & p_{n-1} \end{pmatrix} = \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_{n-1} & p_{n-1} \\ q_{n-2} & p_{n-2} \end{pmatrix} = M_n \begin{pmatrix} q_{n-1} & p_{n-1} \\ q_{n-2} & p_{n-2} \end{pmatrix}.$$

Iterating this, we see that

$$\begin{pmatrix} q_n & p_n \\ q_{n-1} & p_{n-1} \end{pmatrix} = M_n M_{n-1} \cdots M_n \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Taking the determinant of both sides, we find that

$$q_n p_{n-1} - q_{n-1} p_n = (-1)^{n+2} = (-1)^n.$$

To see the second claim, we apply the recursion:

$$\begin{aligned} q_n p_{n-2} - p_n q_{n-2} &= (a_n q_{n-1} + q_{n-2}) p_{n-2} - q_{n-2}(a_n p_{n-1} + p_{n-2}) \\ &= a_n (q_{n-1} p_{n-2} - q_{n-2} p_{n-1}) = (-1)^{n-1} a_n. \end{aligned}$$

□

Corollary 6.26. *The fraction $\frac{p_n}{q_n}$ is fully simplified for every $n \geq 0$, i.e.* $\gcd(p_n, q_n) = 1$.

Proof. This is an immediate consequence of Theorem 6.25 and Bezout's theorem (Theorem 1.15). □

We can now finally show that continued fraction expansions compute their associated real numbers

Theorem 6.27. *Let $x \in \mathbb{R}$, and let $a_0, a_1, \ldots$ be the associated continued fraction expansion. Then*

$$\lim_{n \to \infty} [a_0, a_1, \ldots, a_n] = \lim_{n \to \infty} \frac{p_n}{q_n} = x$$

where we interpret the limit appropriately if the continued fraction expansion is finite.

Proof. From Theorem 6.19, and Theorem 6.23, we know this to be the case if the continued fraction expansion is finite. So suppose it is infinite. Then it suffices to show

$$\lim_{n \to \infty} \frac{p_n}{q_n} = x.$$

First note that the map

$$\begin{aligned}\mathbb{R}_{>0} &\to \mathbb{R}\\ y &\mapsto [a_0, a_1, \ldots, a_{n-1}, y]\end{aligned}$$

consists of a composition of n inversions and translations. Thus it is decreasing if n is odd and increasing if n is even. Therefore, if n is even

$$\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n] < [a_0, a_1, \ldots, x_n] = x$$

and if n is odd

$$\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n] > [a_0, a_1, \ldots, x_n] = x.$$

Next note that

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^n a_n}{q_n q_{n-2}}$$

which is positive if n is even and negative if n is odd. Hence the sequence $\left\{\frac{p_{2n}}{q_{2n}}\right\}$ is increasing and bounded above by x. Thus it converges to some limit $L \leq x$. Similarly, the sequence$\left\{\frac{p_{2n+1}}{q_{2n+1}}\right\}$ is decreasing and bounded below by x. Thus it converges to some limit $R \geq x$.

Finally note that

$$\frac{p_{2n}}{q_{2n}} - \frac{p_{2n+1}}{q_{2n+1}} = \frac{(-1)^{2n+1}}{q_{2n} q_{2n+1}}$$

tends to 0 as n grows, as q_n is increasing. Thus

$$L = \lim_{n\to\infty} \frac{p_{2n}}{q_{2n}} = \lim_{n\to\infty} \frac{p_{2n+1}}{q_{2n+1}} = R.$$

Thus, $R = L \leq x \leq R$, so we must have $R = L = x$ and $\lim_{n\to\infty} \frac{p_n}{q_n} = x$. □

Thus we have show that the canonical continued fraction expansion of a real number x computes x. In fact, for irrational x, this continued fraction expansion is unique! If $x = [b_0, b_1, \ldots]$ is a continued fraction expansion, then

$$0 \leq x - b_0 = \cfrac{1}{b_1 + \cfrac{1}{\ddots}} < \frac{1}{b_1} \leq 1$$

and so we must have $b_0 = \lfloor x \rfloor$, and so on. Thus, for irrational x, we can talk about *the* continued fraction expansion.

We do not have uniqueness for rational x, as if $a_n > 1$, then

$$[a_0, a_1, \ldots, a_n] = [a_0, \ldots, a_{n-1}, a_n - 1, 1]$$

so we can only discuss talk about a canonical continued fraction expansion in this case.

6.2.2 Diophantine approximation

We have seen that convergents give arbitrarily precise approximations of real x, but so does any sequence of rational numbers tending to x. There is a sense in which, at least for irrational x, the convergents give the best approximations. Specifically, convergents provide a better approximation than any rational number whose denominator is at most as large as that of the convergent.

This is very valuable when doing numerical computations. When computing with rational numbers, it is possible to avoid the rounding errors associated to floating point arithmetic, so we can very easily control our errors when approximating irrational expressions, particular compared to the amount of memory needed. For example,

$$\pi \approx [3, 7, 15, 1] = \frac{355}{113}$$

is accurate to 6 decimal places, but only requires us to compute with three digit integers.

Let us make this precise.

Proposition 6.28. *For all irrational x and $n \geq 0$,*

$$\frac{1}{q_n(q_{n+1} + q_n)} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}$$

Proof. We know that

$$\frac{p_{2n}}{q_{2n}} < \frac{p_{2n+2}}{q_{2n+2}} < x < \frac{p_{2n+1}}{q_{2n+1}} < \frac{p_{2n-1}}{q_{2n-1}}$$

for all $n \geq 0$. Checking for both even and odd n this implies that

$$\left| x - \frac{p_n}{q_n} \right| < \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{1}{q_n} q_{n+1}.$$

Noting that if $a < b < c$, then $c - a > b - a$, and similarly with the reverse inequality, we also have that

$$\begin{aligned} \left| x - \frac{p_n}{q_n} \right| > \left| \frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} \right| &= \frac{a_{n+2}}{q_n q_{n+2}} = \frac{1}{q_n \frac{q_{n+2}}{a_{n+2}}} \\ &= \frac{1}{q_n(q_{n+1} + \frac{q_n}{a_{n+2}})} > \frac{1}{q_n(q_{n+1} + q_n)}. \end{aligned}$$

□

Corollary 6.29. *For all irrational $x \in \mathbb{R}$, $|q_n x - p_n| < \frac{1}{q_{n+1}}$, which tends to 0 as n tends to infinity.*

Corollary 6.30. *The rationals $\mathbb{Q}$ are dense in the reals $\mathbb{R}$.*

Theorem 6.31. *For all irrational $x \in \mathbb{R}$, $n \geq 0$, $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that $q \leq q_n$*

$$|qx - p| > |q_n x - p_n|$$

unless $\frac{p}{q} = \frac{p_n}{q_n}$.

Proof. Fix $n \geq 0$, and let $0 < q \leq q_n$ be a fixed integer. Suppose that $\frac{p}{q} \neq \frac{p_n}{q_n}$. Then, since

$$\det \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \pm 1$$

this matrix has an inverse with integer entries. As such, there exist $y, z \in \mathbb{Z}$ such that

$$\begin{aligned} p_n y + p_{n-1} z &= p, \\ q_n y + q_{n-1} z &= q. \end{aligned}$$

If $z = 0$, then $\frac{p}{q} = \frac{p_n}{q_n}$. Thus, we must have $z \neq 0$. If $y = 0$, then $\frac{p}{q} = \frac{p_{n-1}}{q_{n-1}}$, and Proposition 6.28 tells us that

$$|q_{n-1}x - p_{n-1}| > \frac{1}{q_n + q_{n-1}} \geq \frac{1}{a_{n+1}q_n + q_{n-1}} = \frac{1}{q_{n+1}} > |q_n x - p_n|$$

so the claim holds.

Otherwise $y, z \neq 0$, and since $q_n y + q_{n-1} z = q < q_n$, y and z must have opposite signs. Since $q_n x - p_n$ and $q_{n-1}x - p_{n-1}$ also have opposite signs, we must have that $y(q_n x - p_n)$ and $z(q_{n-1}x - p_{n-1})$ have the same sign. Thus

$$\begin{aligned} |qx - p| &= |y(q_n x - p_n) + z(q_{n-1}x - p_{n-1})| \\ &= |y(q_n x - p_n)| + |z(q_{n-1}x - p_{n-1})| \\ &> |y(q_n x - p_n)| \geq |q_n x - p_n|. \end{aligned}$$

□

This next result shows that if we have too good an approximation of x, then it must be a convergent. The proof can be a little confusing, so we might omit it in lectures, depending on how much time we have.

Theorem 6.32. *For any irrational $x \in \mathbb{R}$, $p \in \mathbb{Z}$ and $q \in \mathbb{Z}$, if $|qx - p| < \frac{1}{2q}$, then $\frac{p}{q} = \frac{p_n}{q_n}$ for some $n \geq 0$.*

Proof. If $|qx - p| < \frac{1}{2q}$, then $qx - p = \frac{\epsilon\theta}{q}$ where $\epsilon = \pm 1$ and $\theta \in [0, \frac{1}{2})$. We can rearrange this to obtain that

$$x = \frac{p + \frac{\epsilon\theta}{q}}{q}.$$

We may assume $\gcd(p, q) = 1$, as otherwise the inequality implies that $|q'x - p'| < \frac{1}{2q'}$ for

$$p' = \frac{p}{\gcd(p, q)}, \; q' = \frac{q}{\gcd(p, q)}.$$

Let $\frac{p}{q} = [b_0, b_1, \ldots, b_n]$ have convergents $\frac{s_m}{t_m}$, so that $\frac{p}{q} = \frac{s_n}{t_n}$. In fact, since we assume $\gcd(p, q) = 1$, and convergents are fully simplified, $s_n = p$ and $t_n = q$.

Recall we have an ambiguity in continued fraction expansions for rational numbers

$$[c_0, c_1, \ldots, c_m + 1] = [c_0, c_1, \ldots, c_m, 1]$$

and hence we can assume n is such that $\epsilon = (-1)^n = t_n s_{n-1} - t_{n-1} s_n$.

Let $y = \frac{1}{\theta} - \frac{t_{n-1}}{t_n}$ have continued fraction expansion

$$y = [d_0, d_1, d_2, \ldots]$$

As $\frac{1}{\theta} > 2$ and $\frac{t_{n-1}}{t_n} < 1$, $y > 1$ and hence $d_0 \geq 1$. Thus, we can make sense of the continued fraction

$$[b_0, b_1, \ldots, b_n, d_0, d_1, \ldots] = [b_0, b_1, \ldots, b_n, y]$$

which by Corollary 6.24 is equal to

$$\frac{y s_n + s_{n-1}}{y t_n + t_{n-1}} = \frac{qp - \theta(t_{n-1} s_n - t_n s_{n-1})}{q^2 - \theta t_n t_{n-1} + \theta t_n t_{n-1}} = \frac{qp - \theta\epsilon}{q^2} = x$$

Thus, by uniqueness of continued fraction expansions for irrational x, we must have that

$$b_0 = a_0,\ b_1 = a_1, \ldots, b_n = a_n$$

and hence

$$\frac{p}{q} = [b_0, \ldots, b_n] = [a_0, \ldots, a_n] = \frac{p_n}{q_n}.$$

□

Remark 6.33. *Theorem 6.31 not only tells us that the convergents of a continued fraction give the best approximations with constrained denominators, this also tells us which convergents are the most practical. Since $q_{n+1} = a_{n+1} q_n + q_{n-1}$, if a_{n+1} is very large, q_{n+1} is very large compared to q_n. As such, $\frac{p_n}{q_n}$ must be a very good approximation relative to the size of its denominator. This can be easily seen with π. $a_4 = 292$ is quite large compared to $q_3 = 113$, especially given that all prior a_n were less than 15. And indeed*

$$\frac{p_3}{q_3} = \frac{355}{113}$$

is accurate to 6 decimal places with a three digit denominator, while

$$\frac{p_4}{q_4} = \frac{103993}{33102}$$

is only accurate to 9 decimal places, despite having a five digit denominator.

Thus, an irrational number is easy to approximate if it has unusually large terms in its continued fraction expansion. At the other end of the spectrum, the irrational number given by

$$x = [1, 1, 1, 1, \ldots]$$

is going to be very poorly approximated by its convergents, as their denominators increase as slowly as is possible. Some people will argue that this means that

$$x = \frac{1+\sqrt{5}}{2}$$

is the most irrational number. Maybe the Golden Mean isn't quite so beautiful after all!

6.3 Quadratic irrationals and Pell-Fermat equations

We call an irrational $x \in \mathbb{R}$ a quadratic irrational if its minimal polynomial is quadratic, i.e. it is algebraic of degree 2. From the quadratic formula, every quadratic irrational can be written in the form $a + b\sqrt{d}$ for some $a, b \in \mathbb{Q}$ and $d \in \mathbb{N}$ not a square. We can take d to be an integer as

$$\sqrt{\frac{r}{s}} = \frac{\sqrt{rs}}{s}.$$

We denote by

$$\mathbb{Q}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$$

and similarly $\mathbb{Z}[\sqrt{d}]$. It is a simple exercise in algebra to verify the following proposition.

Proposition 6.34. *The set $\mathbb{Q}[\sqrt{d}]$ is a field: it is closed under addition, subtraction, multiplication, and division by non-zero elements. The set $\mathbb{Z}[\sqrt{d}]$ is a ring: it is closed under addition, subtraction and multiplication.*

This space has a lot in common with the Gaussian integers. We can define an analogue of complex conjugation and norm.

Definition 6.35. *Given $\beta = a + b\sqrt{d} \in \mathbb{Q}[\sqrt{d}]$, define $\overline{\beta} = a - b\sqrt{d}$, and $\mathrm{N}(\beta) = \beta\overline{\beta} = a^2 - b^2 d$.*

Furthermore these operations are compatible with the field structure:

$$\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}, \quad \overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}.$$

By using this additional structure, we can give a very nice description of quadratic irrationals in terms of their continued fraction expansions, though the proof is a bit involved. As such, we may only cover one direction in the lectures.

Theorem 6.36. *An irrational $x \in \mathbb{R}$ is a quadratic irrational if and only if its continued fraction expansion is ultimately periodic.*

Proof. We will use the notation $\overline{b_0, b_1, \ldots, b_n}$ to indicate the infinite sequence obtained by repeating the given finite sequence. Suppose that

$$x = [a_0, a_1, \ldots, a_r, \overline{b_0, b_1, \ldots, b_s}]$$

and let

$$y = \overline{[b_0, b_1, \ldots, b_s]}.$$

Then

$$y = [b_0, b_1, \ldots, b_s, y] = \frac{yp'_s + p'_{s-1}}{yq'_s + q'_{s-1}}$$

and so y is an irrational number satisfying a quadratic equation. Since $y \in \mathbb{R}$, there exist $a, b \in \mathbb{Q}$ and $d \in \mathbb{N}$ not a square such that $y = a + b\sqrt{d}$. Then

$$x = [a_0, \ldots, a_r, y] = \frac{yp_r + p_{r-1}}{yq_r + q_{r-s}} \in \mathbb{Q}[\sqrt{d}]$$

and is therefore a quadratic irrational.

Conversely, suppose that

$$x = \frac{a}{c} + \frac{b\sqrt{d}}{c}$$

is a quadratic irrational. We can assume, without loss of generality, that $b > 0$. Then, letting

$$R = a|c|,\ S = c|c|,\ D = b^2c^2d > 0$$

we have that

$$x = \frac{R + \sqrt{D}}{S}.$$

Note that $D - R^2 = c^2(b^2d - a)$ is divisible by S. Thus, in computing the continued fraction expansion, we have that

$$x_1 = \frac{1}{x - \lfloor x \rfloor} = \frac{1}{\frac{R - S\lfloor x \rfloor + \sqrt{D}}{S}} = \frac{(-R + S\lfloor x \rfloor) + \sqrt{D}}{\frac{D - R^2 + 2S\lfloor x \rfloor - S^2}{S}} = \frac{R_1 + \sqrt{D}}{S_1}$$

for integers R_1, S_1 such that $D - R_1^2 = SS_1$. Repeating this, we obtain a sequence of integers (R_n, S_n) such that $S_nS_{n+1} = D - R_{n+1}^2$ and $x_n = \frac{R_n + \sqrt{D}}{S_n}$.

As $x = [a_0, \ldots, a_{n-1}, x_n] = \frac{x_n p_{n-1} + p_{n-1}}{x_n q_{n-1} + q_{n-2}}$, we can compute that

$$x_n = -\frac{xq_{n-1} - p_{n-1}}{xq_{n-2} - p_{n-2}}$$

and hence

$$\frac{R_n - \sqrt{D}}{S_n} = \overline{x}_n = -\frac{\overline{x}q_{n-1} - p_{n-1}}{\overline{x}q_{n-2} - p_{n-2}} = \frac{-q_{n-1}}{q_{n-2}} \frac{\overline{x} - \frac{p_{n-1}}{q_{n-1}}}{\overline{x} - \frac{p_{n-2}}{q_{n-2}}}.$$

As n tends to infinite

$$\frac{\overline{x} - \frac{p_{n-1}}{q_{n-1}}}{\overline{x} - \frac{p_{n-2}}{q_{n-2}}} \to \frac{\overline{x} - x}{\overline{x} - x} = 1$$

and so, for large enough n $\overline{x}_n < 0$. Hence

$$\frac{2\sqrt{D}}{S_n} = x_n - \overline{x_n} > 1 > 0.$$

In particular, $S_n > 0$. Thus,

$$0 < S_n S_{n+1} = D - R_{n+1}^2$$

for n large enough, and so $0 \leq R_n^2 < D$ for n large enough. Also, since $S_n > 0$ for n large enough $S_n \leq S_n S_{n+1} = D - R_{n+1}^2 \leq D$.

Thus (R_n, S_n) can only take finitely many values and is therefore ultimately periodic. Hence $x_{m+k} = x_m$ for some $m, k > 0$, and so

$$x = [a_0, a_1, \ldots, a_{m-1}, \overline{a_m, \ldots, a_{m+k-1}}]$$

is ultimately periodic. □

Example 6.37. *Let $x = \frac{1+\sqrt{5}}{2}$. We can compute that $a_0 = 1$ and*

$$x_1 = \frac{1}{\frac{\sqrt{5}-1}{2}} = \frac{2}{\sqrt{5}-1} = \frac{2(1+\sqrt{5})}{4} = x$$

so $x = [\overline{1}]$.

If $x = \sqrt{6}$, then $a_0 = 2$, and $x_1 = \frac{1}{\sqrt{6}-2} = 1 + \frac{\sqrt{6}}{2}$. Thus $a_1 = 2$, and $x_2 = 2 + \sqrt{6}$. Thus $a_2 = 4$, and $x_3 = 1 + \frac{\sqrt{6}}{2} = x_1$. Hence $x = [2, \overline{2, 4}]$.

6.3.1 The Pell-Fermat equation

Continued fractions of quadratic irrationals also play an important role in solving a certain Diophantine equation, called the Pell-Fermat equation. Let $d \in \mathbb{N}$ be a non-square. A Pell-Fermat equation is an Diophantine equation

$$x^2 - dy^2 = 1$$

This has the trivial solution of $(x, y) = (\pm 1, 0)$, but finding solutions in positive integers is a non-trivial task.

Example 6.38. *If $d = 2$, then some solutions are given by $(x, y) = (3, 2)$ or $(x, y) = (17, 12)$. But as will most Diophantine equations, just trying small numbers rarely works. If $d = 61$, then the smallest positive solution is*

$$(x, y) = (1766319049, 226153980).$$

To solve this in general, we employ tactics similar to how we used Gaussian integers to investigate sums of two squares. We first note that solutions to a Pell-Fermat equation are in bijection with elements $\alpha \in \mathbb{Z}[\sqrt{d}]$ such that $\mathrm{N}(\alpha) = 1$

As the norm is multiplicative, we can recreate the proof of Proposition 5.16 to show that $\mathrm{N}(\alpha) = 1$ if and only if α has a multiplicative inverse in $\mathbb{Z}[\sqrt{d}]$. Furthermore, as the norm is multiplicative, if we have

$$x^2 - dy^2 = 1, \text{ and } z^2 - dw^2 = 1$$

then we have

$$\begin{aligned} 1 = \mathrm{N}(x + y\sqrt{d})\,\mathrm{N}(z + w\sqrt{d}) &= \mathrm{N}((x + y\sqrt{d})(z + w\sqrt{d}) \\ &= \mathrm{N}(xz + ywd + (yz + xw)\sqrt{d}) \\ &= (xz + ywd)^2 - (yz + xw)^2 d. \end{aligned}$$

Thus we can combine solutions. Combined with a powerful result from Dirichlet, we can actually general all solutions, if we can find the correct start point!

Theorem 6.39 (Dirichlet)**.** *There exists $\varepsilon = a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ with $a, b > 0$, called the fundamental unit, such that $\mathrm{N}(\alpha) = 1$ if and only if $\alpha = \pm\varepsilon^n$ for some $n \in \mathbb{Z}$.*

Thus, if we can determine the fundamental unit $\varepsilon = a_1 + b_1\sqrt{d}$, we can obtain infinitely many solutions (a_n, b_n) by computing $a_n + b_n\sqrt{d} = \varepsilon^n$.

In order to compute the fundamental unit, we first notice that, from the positivity of a_1, b_1, we have that (a_n, b_n) is increasing for $n \geq 1$. As such, the fundamental unit will correspond to a minimal positive integer solution to

$$x^2 - dy^2 = 1.$$

Finally, we need the following lemma

Lemma 6.40. *If $a, b \in \mathbb{N}$ satisfies $a^2 - db^2 = 1$, then $\frac{a}{b}$ is a convergent of $\sqrt{d}$. As such, the minimal positive integer solution to $x^2 - dy^2 = 1$ is given by $(x, y) = (p_n, q_n)$ for the minimal n for which this is a solution.*

Proof. The minimality follows from the fact that p_n and q_n are increasing. Thus it suffices to show that every positive integer solution to $x^2 - dy^2 = 1$ corresponds to a convergent. Suppose that $a^2 - db^2 = 1$. Then

$$\left|\frac{a}{b} - \sqrt{d}\right| = \frac{|a^2 - db^2|}{b(a + b\sqrt{d}} = \frac{1}{b(a + b\sqrt{d}} < \frac{1}{b(a + b)}$$

since $d > 1$, and so

$$|b\sqrt{d} - a| < \frac{1}{a + b}.$$

As $a = \sqrt{db^2 - 1} \geq \sqrt{2b^2 - 1} \geq b$ for all $b \geq 1$, we therefore have that

$$|b\sqrt{d} - a| < \frac{1}{b + b} = \frac{1}{2b}$$

which implies that $\frac{a}{b}$ is a convergent. □

Therefore, in order to solve the Pell-Fermat equation, we just need to compute convergents until we get a solution. This will give us our fundamental unit, from which we can compute all other solutions.

Example 6.41. *Let us find the fundamental solution to* $x^2 - 6y^2 = 1$. *Recall that*

$$\sqrt{6} = [2, \overline{2, 4}]$$

and so the convergents correspond to

$$(p_0, q_0) = (2, 1),$$
$$(p_1, q_1) = (5, 2),$$

and so on. The first of these does not give a solution:

$$4 - 6 = -2 \neq 1$$

while the second does:

$$25 - 24 = 1$$

and hence we obtain a fundamental unit $\varepsilon = 5 + 2\sqrt{6}$, *from which we can obtain all other solutions. For example,* $\varepsilon^2 = 49 + 20\sqrt{6}$, *and indeed*

$$49^2 - 6(400) = 1.$$

7 Summary of main results

As a possible study aid, we summarise here the major results from each section of the course. It is not a comprehensive list, but should provide a good start point. It notably does not include definitions or examples, or every useful corollary.

Chapter 1

- Theorem 1.2 - Division with remainder
- Euclid's Algorithm - Theorem 1.13 - A tool for computing gcd
- Bezout's Theorem - Theorem 1.15 - Expresses gcds as a linear combination
- Corollary 1.17 - A special case of Bezout's Theorem for coprime integers
- Gauss' Lemma - Lemma 1.19 - Divisibility when the divisor is coprime to a factor
- Euclid's Lemma - Corollary 1.23 - Divisibility of a product by a prime - You should be able to prove this as a special case of Gauss' Lemma
- The fundamental theorem of arithmetic - Theorem 1.25 - Prime factorisation
- Euclid's theorem - Theorem 1.26 - The infinitude of primes
- Theorem 1.43 - Sums of divisor formulae

Chapter 2

- Theorem 2.13 - Diophantine equations have solutions only if they have solutions modulo n
- Theorem 2.21 - Invertible elements modulo n
- The Chinese Remainder Theorem - Theorem 2.33 - Solving linear congruences with multiple moduli
- Corollary 2.37 - Formula for $\phi(n)$
- Euler's Theorem - Theorem 2.44 - Bounds multiplicative order of invertible elements
- Fermat's Little Theorem - Corollary 2.46 - Euler's theorem for prime moduli - You should be able to prove this as a special case of Euler's theorem
- Theorem 2.59 - Primitive roots exist modulo primes

Chapter 3

- Corollary 3.9 - Describes the number of possible n^{th} roots modulo p
- Fact 3.15 - Properties of the Legendre symbol
- Quadratic Reciprocity - Theorem 3.23 - Relates Legendre symbols of odd primes
- Theorem 3.24 - Number of roots of a quadratic modulo p

Chapter 4

- Theorem 4.3 - Describes all primitive Pythagorean triples

Chapter 5

- Theorem 5.1 - Describes sums of two squares
- Theorem 5.3 - Describes sums of three squares
- Theorem 5.4 - Describes sums of four squares
- Corollary 5.12 - The product of a sum of two squares is a sum of two squares
- Theorem 5.18 - Long division in the Gaussian integers
- Lemma 5.23 - Division of Gaussian integers compared to norms
- Theorem 5.27 - Computing the gcd of Gaussian integers
- Corollary 5.28 - Gaussian version of Bezout's Theorem
- Corollary 5.29 - Gaussian version of Gauss' Lemma
- Theorem 5.35 - Factorisation into irreducibles
- Theorem 5.37 - Description of all irreducibles

Chapter 6

- Lemma 6.3 - Approximation of rational numbers
- Liouville's Theorem - Lemma 6.10 - Approximation of algebraic numbers
- Theorem 6.23 - Convergents compute finite continued fractions
- Theorem 6.27 - Continued fractions converge
- Theorem 6.31 - Continued fractions give the best approximations
- Theorem 6.36 - Quadratic irrationals have periodic continued fractions
- Dirichlet's Theorem - Theorem 6.39 - There exists a fundamental unit
- Lemma 6.40 - Convergents solve the Pell-Fermat equation