M14: NUMBER THEORY

Prof. David Loeffler

Assistant: **Dr. Francesco Zerman**

Autumn 2024

Last updated 2nd December 2024

FernUni.ch
UniDistance.ch

Contents

Preamble

Acknowledgements

This course is loosely based on a lecture course taught by my former colleague Prof. John Cremona at the University of Warwick. It also incorporates a number of suggestions from Sarah Zerbes of ETH Zürich.

HTML Version

These lecture notes are also available in an HTML version and in app form.

https://apptest.fernuni.ch

The HTML version contains the lecture notes, and additional resources such as model solutions to exercises.

Updates during the semester

- *29.08.2024*: added the name "Bézout's identity" for Corollary 1.17. (This is a misnomer, since the result for integers was known long before the work of Bézout; Bézout's contribution was to prove the analogous identity for polynomials. However, the name is widely encountered in textbooks anyway, so it is useful to be aware of.)
- *13.11.2024*: sorted out a sign inconsistency with Eisenstein integers. (Sometimes ω denoted $\frac{1+\sqrt{-3}}{2}$ and sometimes $\frac{-1+\sqrt{-3}}{2}$; I standardized on the latter, so that $\omega^3 = 1$.)
- *21.11.2024*: updated Proposition 11.12 to explicitly point out that the norm on $\mathbb{Q}_p$ is nonarchimedean (which is used in the next section).
- *02.12.2024*: fixed some typos in Chapter 12, and a formatting issue that caused §8.2 to appear scrambled in certain PDF viewers.

CHAPTER 1

Divisibility and GCD

Preliminaries

Remark 1.1 (Recommended textbooks) All the material you need to know is in this script, but you might find some of the books below useful for an alternative viewpoint on the same topics:

- Chapters I–III of Davenport's classic text *The Higher Arithmetic* are a very readable account of elementary number theory (i.e. everything in chapters 1–4 and 6–7 of these notes). Since Davenport died in 1969, the copyright on this book expired long ago and you can download it for free – entirely legally – from `https://archive.org/details/h.-davenport-the-higher-arithmetic/`.
- RSA cryptography (chapter 5) is too recent to be in Davenport's book, but is covered in many more recent texts, such as Coutinho's book *The Mathematics of Ciphers*.
- For algebraic number fields (chapters 8–10), I highly recommend Stewart and Tall's *Algebraic Number Theory and Fermat's Last Theorem*, although this goes a long way beyond what we can cover here. Cox's lovely book *Primes of the form* $x^2 + ny^2$ gives a very interesting and original perspective on some of these ideas.
- Gouvêa's book *P-adic Numbers: An Introduction* is an excellent reference for p-adic arithmetic (chapters 10–11). There is also a (somewhat more advanced) text by Koblitz, *P-adic numbers, p-adic analysis and zeta-functions*, which is a good read.

Remark 1.2 (General notations) In this module we use the following symbols:

- $\mathbb{N}$ denotes the natural numbers $\{0, 1, 2, 3, \dots\}$;
- $\mathbb{Z}$ the integers (positive, negative or zero);
- $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ the fields of rational, real, and complex numbers respectively.
- $\mathbb{N}_+$ denotes the *positive*[a] integers $\{1, 2, 3, \dots\}$.
- If a is in $\mathbb{R}$ (and in particular if $a \in \mathbb{Z}$), the symbol $|a|$ means the absolute value of a, i.e. $|a| = a$ if $a \geqslant 0$, and $|a| = -a$ if $a \leqslant 0$.
- For $n \in \mathbb{N}$, we write $n!$ (read as "n factorial") for the product $1 \times 2 \times \cdots \times n$, with $0!$ defined to be 1.
- The logical symbols $\Rightarrow$, $\iff$, $\exists$, $\forall$ have their usual meanings.
- The symbol $\square$ denotes the end of a proof.
- The symbol $\circledast$ is used to mark unsolved problems and conjectures.

[a] Beware that some other texts use $\mathbb{N}$ for positive integers!

Remark 1.3 (Reminders on induction) We're going to use **induction** quite a lot in this module, so it might be a good idea to revise it if your memory has got rusty.

As a reminder: the *principle of mathematical induction*, which you saw way back in M01 Algorithmics, is a very powerful tool for proving statements about $\mathbb{N}$. It goes as follows. Suppose $P(n)$ is some statement about the natural number n, and:

- $P(0)$ is true,
- for any $n \in \mathbb{N}$ the implication $P(n) \implies P(n+1)$ is true.

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

There are a few variants of induction which are useful:
Different starting points: Let $t \in \mathbb{N}$ be given. If $P(t)$ is true, and for any $n \geqslant t$ we have $P(n) \implies P(n+1)$, then $P(n)$ is true for all $n \geqslant t$. (The usual induction is $t = 0$, but the $t = 1$ case also occurs frequently.)
This can, of course, easily be derived from "usual" induction applied to the new statement $Q(n)$ defined by "if $n \geqslant t$ then $P(n)$", which is vacuously true for $n < t$.
Strong induction: Suppose P is a statement such that:

- for any $n \in \mathbb{N}$, if $P(r)$ is true for all $r < n$, then $P(n)$ is true.

Then $P(n)$ holds for all n.
This looks far more powerful than usual induction (because we have to prove only one thing, and we're allowed to assume something that looks a lot stronger); but in fact it easily follows from usual induction.

Exercise 1.4 Deduce Strong Induction from usual Induction. (Hint: consider the statement $Q(n)$ defined as "$P(r)$ holds for all r with $r < n$".)

Minimal elements: Our final induction variant is known as the *well-ordering principle* for $\mathbb{N}$.

- Let $S \subset \mathbb{N}$ be a non-empty set. Then S has a minimal element; that is, there exists $n \in S$ such that every $m \in S$ satisfies $m \geqslant n$.

It's not immediately obvious that this has anything to do with induction at all! But it's clearly something quite special about $\mathbb{N}$: it's obviously false for $\mathbb{Z}$, or for the non-negative reals[a].
To see this, suppose S doesn't have a minimal element, and let $P(n)$ be the statement "$m \geqslant n$ for all $m \in S$". Clearly $P(0)$ holds, since every natural number is $\geqslant 0$. Now, if $P(n)$ holds, then we must have $n \notin S$, since otherwise n would be the minimal element of S. So for $m \in S$, we have $m \geqslant n$ and $m \neq n$. So $m \geqslant n+1$, and thus $P(n+1)$ holds. By induction, $P(n)$ holds $\forall n$; so S is empty, a contradiction.

Exercise 1.5 Give an example of a subset of the non-negative reals $\mathbb{R}_{\geqslant 0}$ which does not have a minimal element.

[a] Of course, it's hugely important in real analysis that any bounded-below subset of the real numbers has a *greatest lower bound*, but this is not the same thing as a *minimal element* (why?)

1.1 Divisibility

Recall the following familiar definition:

Definition 1.6 Let $a, b \in \mathbb{Z}$. We say "a divides b", or "b is a multiple of a", if there exists $n \in \mathbb{Z}$ such that $na = b$. If so, we write "$a \mid b$"; otherwise we write "$a \nmid b$"; and we say a is a *divisor* or *factor* of b.

Example 1.7 We have $3 \mid -15$, since $3 \times (-5) = -15$.

Notice that this still makes sense if a or b is zero; and since $n \cdot 0 = 0$ for all n, it follows that everything divides 0, but 0 does not divide anything except itself. On the other hand, 1 and -1 both divide everything, and nothing except ± 1 can divide them. (So 0 is the "most divisible" element of $\mathbb{Z}$, while 1 and -1 are the "least divisible".)

Remark 1.8 The "divides" symbol is a *relation*: for any given values of a and b, "$a \mid b$" is a self-contained statement which is either true or false. Don't confuse it with division a/b, which is a number (if it is defined at all, which it might not be if $b = 0$).

Exercise 1.9 Check that if $a, b \in \mathbb{N}$, then $a \mid b$ if and only if there exists $n \in \mathbb{N}$ such that $na = b$. (Take care with the case $a = 0$!)

Proposition 1.10 (Elementary properties of divisibility) *Let $a, b, c, \dots \in \mathbb{Z}$. Then:*

(i) If $a \mid b$, then $a \mid kb$ for all $k \in \mathbb{Z}$.
(ii) If $a \mid b$ and $a \mid c$, then $a \mid b \pm c$.
(iii) If $a \mid b$ and $b \mid c$ then $a \mid c$.
(iv) If $a \mid b$ and $b \mid a$, then $a = \pm b$.
(v) If $a \mid b$ and $b \neq 0$, then $|a| \leqslant |b|$; so nonzero integers have only finitely many divisors.
(vi) We have $a \mid |a|$ (the notation is awkward; read it as "a divides the absolute value of a").
(vii) If $k \neq 0$, then $a \mid b \iff ka \mid kb$.

Proof Exercise. □

The following innocent-looking proposition will turn out to be crucial in understanding divisibility and factorisation of integers:

Proposition 1.11 (Division with remainder) *Let $a, b \in \mathbb{Z}$ with $a \neq 0$. Then there exists a unique pair of integers (q, r) such that $0 \leqslant r < |a|$ and $b = qa + r$.*

Example 1.12

(i) For $a = 5$ and $b = 21$, we have $(q, r) = (4, 1)$.
(ii) For $a = 5$ and $b = -21$, we have $(q, r) = (-5, 4)$ [*not* $(-4, -1)$!]

Proof Let S be the set of of integers which are of the form $b - qa$, for some $q \in \mathbb{Z}$; and let $S' = S \cap \mathbb{N}$ be the non-negative elements of S.

The set S' is always non-empty (if $b \geqslant 0$, then $b \in S'$, and if $b < 0$, then one checks that $(|a| - 1) \cdot |b| \in S'$).

We know that a non-empty subset of $\mathbb{N}$ always has a smallest element. So let r be the smallest element of S'. If $r \geqslant |a|$, then $r - |a|$ is a strictly smaller element of S', contradiction; so $0 \leqslant r < |a|$. By definition of S' there exists q with $r = b - qa$ so we are done. □

Remark 1.13 More concretely, if $a > 0$, then q is given by $\lfloor b/a \rfloor$, where $\lfloor x \rfloor$ is the *floor* function: the function which converts a real (or rational) number to an integer by rounding towards $-\infty$ (meaning that $\lfloor 1.5 \rfloor = 1$ and $\lfloor -1.5 \rfloor = -2$). Thus we can easily compute q and r from the decimal expansion of b/a.

1.2 The greatest common divisor

Proposition 1.14 *Let $a, b \in \mathbb{Z}$. Then there exists $c \in \mathbb{Z}$ such that the following holds:*

$$\forall x \in \mathbb{Z}, \qquad x \mid c \quad \iff \quad x \mid a \text{ and } x \mid b.$$

This c is uniquely determined by a and b up to sign; and we write $\gcd(a, b)$ *for the unique non-negative c with this property, which we call the* greatest common divisor *of a and b.*

Example 1.15 If we let $a = 20$ and $b = 30$, the integers dividing a are $\{\pm 1, \pm 2, \pm 4, \pm 5, \pm 10, \pm 20\}$, and the integers dividing b are $\{\pm 1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 10, \pm 15, \pm 30\}$. The intersection of these sets, i.e. the set $\{x : x \mid a \text{ and } x \mid b\}$, is $\{\pm 1, \pm 2, \pm 5, \pm 10\}$, which are precisely the divisors of 10. So $\gcd(20, 30) = 10$.

Proof It is clear that if c and c' both satisfy the condition, then $c \mid c'$ and $c' \mid c$, so $c' = \pm c$. Conversely, if c works then $-c$ does too. So it suffices to prove existence.

If a, b are both zero, then the result is trivial; so assume not. Let T denote the set of integers of the form $ma + nb$ for $m, n \in \mathbb{Z}$, and T' its intersection with the *strictly* positive integers. We check easily that T' is non-empty (since at least one of $|a|$ and $|b|$ is in T'); so it contains a smallest element. Let c be this element. Clearly c has the form $ma + nb$, so anything which divides a and b also divides c.

We claim c itself divides both a and b. By symmetry it suffices to show $c \mid a$. By division-with-remainder, we can write $a = qc + r$, for some r with $0 \leqslant r < c$. But $r = a - qc = a - (ma + nb)$ is also in T, and it is non-negative and strictly smaller than c. If $r > 0$, then $r \in T'$, contradicting the minimality of c. So we must have $r = 0$, i.e. q divides a. □

Remark 1.16 Note that (except in the trivial case $a = b = 0$), the greatest common divisor $\gcd(a, b)$ is, as its name suggests, the largest element of the set of common divisors of a and b (the set $\{x \in \mathbb{Z} : x \mid a \text{ and } x \mid b\}$). This follows from the much stronger fact proved above that this set consists precisely of the divisors of c (and since $c > 0$, the largest divisor of c is clearly c itself).
However, if we just *defined* $\gcd(a, b)$ to be the largest element of this set, it wouldn't be clear that all other elements of this set divided it.

Corollary 1.17 (Bézout's identity) *We can always write* $\gcd(a, b)$ *in the form* $ma + nb$*, for some* $m, n \in \mathbb{Z}$*.*

Proof Clear from the proof of existence above. □

Example 1.18 Since 11 and 13 are distinct primes, their GCD must be 1; and indeed we have $6 \cdot 11 + (-5) \cdot 13 = 66 - 65 = 1$.

Exercise 1.19 (Basic Properties of GCD) For all $a, b, k, m \in \mathbb{Z}$:
(i) $\gcd(a, b) = gcd(b, a) = gcd(|a|, |b|)$;
(ii) $\gcd(ka, kb) = |k| \gcd(a, b)$;
(iii) $\gcd(a, 0) = |a|$ and $\gcd(a, 1) = 1$;
(iv) $\gcd(a, b) = \gcd(a, b + ka)$.

Pairs of numbers whose greatest common divisor is 1 are quite special. We call such pairs *coprime*.

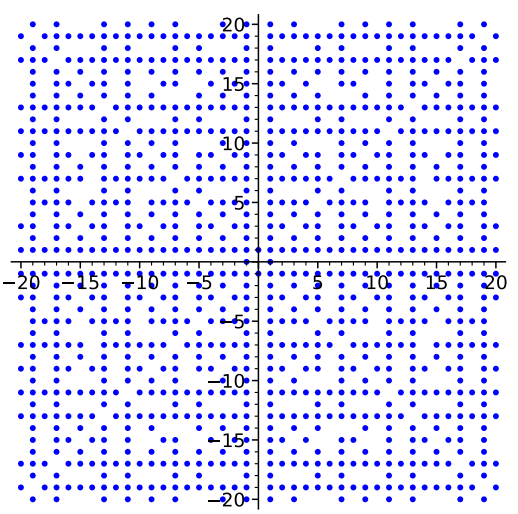

FIGURE 1.1. Pairs of coprime integers (m, n) with $\max(|m|, |n|) \leqslant 20$

Lemma 1.20 (Euler's Lemma) *If* $a \mid bc$*, and* a *and* b *are coprime, then* $a \mid c$*.*

Proof Write $1 = am + bn$. Then $c = c(am + bn) = (mc)a + n(bc)$. Since $a \mid bc$, it divides both terms in the sum, so it divides c. □

Exercise 1.21 Show that if x has the form $ma+nb$, for some $m, n \in \mathbb{Z}$, and x divides both a and b, then $x = \pm\gcd(a, b)$.

1.3 Euclid's algorithm

From our existence proof of the GCD, it's very difficult to see how one could compute it explicitly: we're asking for the smallest element of an infinite set. We can do slightly better using Remark 1.16 – in principle we can make a list of the (finitely many) divisors of both a and b, and find the greatest element appearing in both lists. But there is a way to do much better.

Proposition 1.22 *Let $a, b \in \mathbb{Z}$ with $a \neq 0$, and suppose $b = aq + r$ for some q, r. Then*

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

Proof Clear from Exercise 1.19 (iv). □

If $r \neq 0$ then we can now repeat the process, replacing the larger number with its remainder on division by the smaller. Since the quantity $\max(|a|, |b|)$ gets strictly smaller each time, we must eventually reach a remainder of zero; and since $\gcd(a, 0) = |a|$ for all a, we are done.

It's convenient to arrange this in a table. Suppose we want to calculate $\gcd(113, 251)$. Then we write

$$\begin{aligned} 251 &= \quad 2{\times}113 + 25 \\ 113 &= \quad 4{\times}25 + 13 \\ 25 &= \quad 1{\times}13 + 12 \\ 13 &= \quad 1{\times}12 + 1 \\ 12 &= \quad 12{\times}1 + 0. \end{aligned}$$

Note how the numbers move diagonally to the left each time. The grey numbers (the q's in the division-with-remainder steps) aren't important for calculating the GCD (although we'll find a different use for them in a moment); the key things are the remainders.

We claim that *the last non-zero remainder* in the table is always equal to the GCD of the original two numbers. In the above example, this is 1, so 251 and 113 are coprime. To see this, apply the last proposition repeatedly, once for each division step:

$$\begin{aligned} \gcd(251, 113) &= \gcd(113, 25) \\ &= \gcd(25, 13) \\ &= \gcd(13, 12) \\ &= \gcd(12, 1) \\ &= \gcd(1, 0) = 1. \end{aligned}$$

So we have a method for computing GCD's: *Euclid's algorithm*. It's actually a very effective algorithm in practice.

Exercise 1.23 Recall the *Fibonacci numbers*, defined by $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geqslant 2$. Show that $\gcd(F_n, F_{n-1}) = 1$ for all $n \geqslant 1$.

Now let's see how to use the grey numbers. Working up the table from the last-but-one row, we have

$$\begin{aligned} 1 &= 13 - 1 \times 12 \\ &= 13 - 1 \times (25 - 1 \times 13) && = -1 \times 25 + 2 \times 13 \\ &= -1 \times 25 + 2 \times (113 - 4 \times 25) && = 2 \times 113 - 9 \times 25 \\ &= 2 \times 113 - 9 \times (251 - 2 \times 113) && = -9 \times 251 + 20 \times 113 \end{aligned}$$

So we've written 1 as a sum of integer multiples of 251 and 113. This is a "free bonus" that Euclid's algorithm gives us: for any a, b, we can compute an expression for $\gcd(a, b)$ in the form $ma + nb$.

Remark 1.24 Finding these m, n (as well as just the GCD itself) is so useful that it has its own name: computing the triple $(\gcd(a, b), m, n)$ is called the *extended GCD problem* (XGCD). Lots of computer algebra systems have a command called `xgcd`, or something similar[a], which computes this in one step.

[a] Not to be confused with `xkcd`, an online comic strip popular with mathematicians.

Exercise 1.25 Show that 351 and 451 are coprime, and find integers m, n such that $351m + 451n = 1$.

CHAPTER 2

Prime numbers and unique factorisation

2.1 Prime numbers

I'm sure you all know this definition:

Definition 2.1 An integer $p \in \mathbb{N}$ is said to be *prime* if $p > 1$, and the only divisors of p in $\mathbb{N}$ are 1 and p itself. We write $\mathbb{P}$ for the set of primes.

The first few elements of $\mathbb{P}$ are $\{2, 3, 5, 7, 11, \dots\}$. A number which is not prime is said to be *composite*.

Exercise 2.2 Show that if $p > 1$ and p is not divisible by any integer a with $1 < a \leqslant \sqrt{p}$, then p is prime. Use this to show that 127 is prime. (Hint: $127 < 12^2 = 144$.)

Remark 2.3 It is conjectured, but not known, that there are infinitely many *twin primes* – that is, pairs (p, q) of primes with $q = p + 2$, such as $(59, 61)$. ⊛

We're going to show that any $n \in \mathbb{N}_+$ can be written *uniquely* in terms of the primes. First, we need a lemma:

Lemma 2.4 *Suppose p is prime, and $a, b \in \mathbb{Z}$. If $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof Clearly $\gcd(p, a)$ is a divisor of p, so it must be 1 or p. If $\gcd(p, a) = p$, then $p \mid a$ and we're done. If $\gcd(p, a) = 1$, then Euler's lemma (Lemma 1.20) applies and shows that $p \mid b$. □

This extends in the obvious way to products of three or more factors: if $p \mid a_1 \dots a_r$, then $p \mid a_i$ for some i.

Exercise 2.5 Prove the converse: if $p > 1$ and p is *not* prime, there exist integers a, b with $p \mid ab$ but $p \nmid a$ and $p \nmid b$.

Remark 2.6 Lemma 2.4, and its converse, show that a positive integer $n \in \mathbb{N}_+$ is a prime number iff it is a *prime element* of the ring $\mathbb{Z}$ in the sense of the *M11 Algebra* course.

FIGURE 2.1. The insect Magicicada septendecim lives most of its life underground, emerging in huge swarms every 17 years to mate and die. A related species has a 13-year cycle. There are various theories why these insects have evolved to use prime numbers of years.

2.2 Unique factorisation

Theorem 2.7 (Fundamental Theorem of Arithmetic) *Every positive integer n can be written as a product of prime numbers, and its factorisation into primes is unique up to the order of the factors.*

Note that this includes $n = 1$, which is an empty product (the product of no primes); and the primes themselves, with only one factor in the product.

Proof *Existence*: Let $n \in \mathbb{N}_+$. By Strong Induction, we may suppose the theorem is true for all m with $m < n$.

If $n = 1$, then the statement is trivial (product of no primes). So let's suppose $n > 1$. If n is prime, we're again fine (product of one prime). So n must be of the form ab with $1 < a, b < n$. By the induction hypothesis, both a and b are products of primes, hence so is n.

Uniqueness: Suppose $n = p_1 p_2 \dots p_r = q_1 q_2 \dots q_s$ are two prime factorisations of n. We want to deduce that $s = r$ and the q's can be re-ordered such that $q_i = p_i$. We shall argue by induction on r.

If $r = 0$, then $n = 1$; thus $s = 0$ as well (since any nontrivial product of primes is > 1) so we're done.

Now suppose $r \geqslant 1$ and the theorem is true for $r - 1$. Then $p_r \mid q_1 \dots q_s$. Hence $p_r \mid q_i$ for some i, and after reordering we may suppose $p_r \mid q_s$. Since q_s is prime (and $p_r > 1$), this implies $p_r = q_s$. Since p_r is not zero, we deduce that $p_1 \dots p_{r-1} = q_1 \dots q_{s-1}$. By the induction hypothesis, $r - 1 = s - 1$, so $r = s$; and $q_1, \dots, q_{s-1}$ are $p_1, \dots, p_{r-1}$ in some order. So we are done. □

Collecting together any powers of primes which occur in a prime factorization, we obtain two alternative formulations:

Corollary 2.8 *Every positive integer n may be expressed uniquely in the form*

$$n = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$$

where $k \geqslant 0$, $p_1, \ldots, p_k$ are primes with $p_1 < p_2 < \cdots < p_k$, and r_i are integers with $r_i \geqslant 1$.
Alternatively, every positive integer n may be expressed uniquely in the form

$$n = \prod_{p \in \mathbb{P}} p^{e_p}$$

where $e_p \in \mathbb{N}$ for all p, and all but finitely many e_p are zero. □

The exponent e_p which appears in this standard factorization of n is denoted $\mathrm{ord}_p(n)$; it is characterized by the following property:

$$e = \mathrm{ord}_p(n) \iff p^e | n \text{ and } p^{e+1} \nmid n.$$

For example, $700 = 2^2 \cdot 5^2 \cdot 7$, so $\mathrm{ord}_2(700) = \mathrm{ord}_5(700) = 2$, $\mathrm{ord}_7(700) = 1$, and $\mathrm{ord}_p(700) = 0$ for primes $p \neq 2, 5, 7$. Every positive integer n is uniquely determined by the sequence of exponents $\mathrm{ord}_p(n)$. From the uniqueness of the factorisation, it follows that

$$\mathrm{ord}_p(mn) = \mathrm{ord}_p(m) + \mathrm{ord}_p(n) \qquad \forall m, n \in \mathbb{N}_+, p \in \mathbb{P}. \tag{2.1}$$

Proposition 2.9 *Let $m, n \in \mathbb{N}_+$. Then $m \mid n$ iff $\mathrm{ord}_p(m) \leqslant \mathrm{ord}_p(n)$ for all $p \in \mathbb{P}$.*

Proof If $m \mid n$, then $n = km$ for some $k \in \mathbb{N}_+$. From (2.1) it follows that $\mathrm{ord}_p(n) = \mathrm{ord}_p(k) + \mathrm{ord}_p(m) \geqslant \mathrm{ord}_p(m)\ \forall p$.

Conversely, if $\mathrm{ord}_p(m) \leqslant \mathrm{ord}_p(n)$ for every p, let $k = \prod_p p^{\mathrm{ord}_p(n) - \mathrm{ord}_p(m)}$, which is in $\mathbb{N}_+$ since all the exponents are non-negative (and all but finitely many of them are zero). Then we have $n = km$. □

Corollary 2.10 *We have $\gcd(m, n) = 1$ iff there is no prime which divides both m and n.*

Proof The primes which divide $\gcd(m, n)$ are precisely the primes dividing both m and n, by the characterising property of the gcd. It follows from the existence of prime factorisations that for any $k \in \mathbb{N}_+$, we have $k > 1$ iff some prime divides k; applying this to $k = \gcd(m, n)$ we are done. □

Exercise 2.11 Show that for any $m, n \in \mathbb{N}_+$, we have

$$\gcd(m, n) = \prod_{p \in \mathbb{P}} p^{\min(\mathrm{ord}_p(m), \mathrm{ord}_p(n))}.$$

2.3 Infinitude of primes

No introductory course on number theory could possibly omit the following theorem:

Theorem 2.12 (Euclid) *There are infinitely many primes.*

Proof Suppose there are only finitely many primes $p_1, \dots, p_k$. Consider the integer $N = (p_1 p_2 \dots p_k) + 1$. Then all the p_i divide $N - 1$; so none of them can divide N (since otherwise they'd have to divide 1). But $N > 1$, so N must have some prime factors. This contradicts our assumption that $\{p_1, \dots, p_k\}$ are all the primes. □

There are lots of variants of this argument which can be used to construct primes with some special shape; we'll see a few in the next chapter.

Remark 2.13 Although there are infinitely many primes, they get "thinner and thinner" as you go further out. Gauss and Legendre conjectured around 1800 that the ratio

$$\frac{\#\{p \in \mathbb{P} : p \leqslant X\}}{X/\log X}$$

tends to 1 as $X \to \infty$. So for large X, the fraction of integers up to X which are prime is roughly $1/\log(X)$, which tends very slowly to 0.
This conjecture was open for over 100 years, until it was finally proved by Hadamard and de la Vallée Poussin in 1896. A measure of the importance of this theorem is that, among all of the thousands of theorems about prime numbers, theirs is universally known as "*the* prime number theorem".

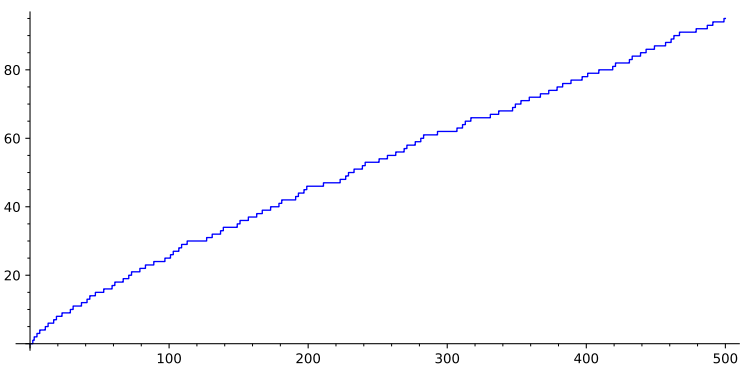

FIGURE 2.2. Graph of the number of primes $\leqslant x$, as a function of x, for $x \leqslant 500$.

Exercise 2.14 Does there exist $n \in \mathbb{N}$ such that all of the numbers $n + 1, n + 2, \dots, n + 20$ are composite?

CHAPTER 3

Congruences and modular arithmetic

3.1 Congruences

The following definition (originally due to Gauss) is a wonderful way of simplifying and organising lots of number-theoretic arguments:

Definition 3.1 Let $a, b, m \in \mathbb{Z}$, with $m \geqslant 1$. We say "a is congruent to b modulo m" if m divides $a - b$ (i.e. there exists $k \in \mathbb{Z}$ such that $a - b = km$).

Example 3.2 For example, a is congruent to 0 modulo 2 iff it's even, and to 1 modulo 2 iff it's odd.

It's easy to see that, for a fixed m, this is an *equivalence relation* in a and b. So the equivalence classes (the *congruence classes* modulo m) form a partition of $\mathbb{Z}$ into disjoint sets. There's exactly m of these congruence classes, represented by the integers $\{0, 1, \dots, m-1\}$, corresponding to the different remainders of a on division by m.

3.2 Modular arithmetic

Definition 3.3 We write $\mathbb{Z}/m\mathbb{Z}$ for the set of congruence classes modulo m.

You saw in *M11 Algebra* that this is a ring: the set $m\mathbb{Z}$ of multiples of m is an ideal of $\mathbb{Z}$, and $\mathbb{Z}/m\mathbb{Z}$ is the corresponding quotient ring. Moreover, the map $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$, sending a to its congruence class, is a ring homomorphism.

Remark 3.4 Take a moment to reflect on what this is really saying: it's saying that, for $a, b \in \mathbb{Z}$, the congruence classes of $a \pm b$ and ab are uniquely determined by the congruence classes of a and b.

That might sound like a lot of abstract nonsense; but it's actually immensely useful for solving concrete questions about $\mathbb{Z}$.

Example 3.5 *"Do there exist integer solutions to the equation $x^2 - 3y^2 = 2$?"*

Suppose (x, y) was a solution. Then, reducing modulo 3, we would have a solution to the equation $(x \bmod 3)^2 = 2$ in $\mathbb{Z}/3\mathbb{Z}$. But $x \bmod 3$ must be one of $\{0, 1, 2\}$, and we have $0^2 = 0$, $1^2 = 2^2 = 1$ in $\mathbb{Z}/3\mathbb{Z}$. So there are no solutions.

Exercise 3.6 Show that if $m, n \in \mathbb{N}_+$ with $n \mid m$, then there is a unique ring homomorphism $\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sending the congruence class $a + m\mathbb{Z}$ to $a + n\mathbb{Z}$ for every $a \in \mathbb{Z}$.

3.3 Primes in congruence classes

Notice that if p is a prime, and $p \neq 2$, then p is odd, so p must be either $1 \bmod 4$ (like 5) or $3 \bmod 4$ (like 7).

Theorem 3.7 *There are infinitely many primes p with $p = 3 \bmod 4$.*

Proof Suppose there are finitely many such primes, namely $p_1, \ldots, p_k$ (with $p_1 = 3$, $p_2 = 7$, etc). Consider the product $N = 4p_2 \ldots p_k + 3$ (note p_1 is not included!)

Clearly N can't be divisible by any of the primes $p_2, \ldots, p_k$, since these all divide $4p_2 \ldots p_k$ but don't divide 3. Moreover, it's also not divisible by $p_1 = 3$ either (since 3 doesn't divide $4p_2 \ldots p_k$, but does divide 3). Finally, it is clearly odd and thus not divisible by 2 either. Hence all of its prime factors must be $1 \bmod 4$.

However, a product of numbers that are all $1 \bmod 4$ must itself be $1 \bmod 4$, while N is obviously $3 \bmod 4$. So we have a contradiction. □

Exercise 3.8 Why doesn't this argument adapt to show that there are infinitely many primes which are $1 \bmod 4$?

Remark 3.9 This is a special case of a much more general theorem: for any $a, b \in \mathbb{N}_+$ with $\gcd(a, b) = 1$, there are infinitely many primes p with $p = a \bmod b$ (Dirichlet's theorem on primes in arithmetic progressions.) However, this is a rather deep theorem and we won't prove it in this module.

3.4 The Chinese remainder theorem

The next theorem will tell us that if m and n are coprime, then congruences mod m and congruences mod n are in some sense "independent of each other": they give totally complementary information.

Theorem 3.10 (Chinese remainder theorem, or CRT) *Let $m, n \in \mathbb{N}_+$ be coprime, and let $x, y \in \mathbb{Z}$. Then there exist integers a such that $a = x \bmod m$ and $a = y \bmod n$; and the set of integers a with this property forms a congruence class modulo mn.*

Remark 3.11 The theorem has this name because it was discovered by ancient Chinese mathematicians (long before it was known in Europe); there is a complete proof in Qin Jiushao's *Mathematical Treatise in Nine Sections* from 1247.

Exercise 3.12 Find an integer a satisfying $a = 5 \bmod 7$ and $a = 6 \bmod 9$.

Proof *Existence*: We first show that there exist integers r, s with the following property:

- $r = 1 \bmod m$ and $r = 0 \bmod n$;
- $s = 0 \bmod m$ and $s = 1 \bmod n$.

To see this, use Euclid's algorithm to write $1 = um + vn$. Then we can take $r = vn$, since $vn = 1 - um = 1 \bmod m$ and clearly $vn = 0 \bmod n$. Similarly, we can take $s = um$. This proves the claim.

Having proved the claim, for any x, y we can take $a = rx + sy$.

Uniqueness: If a is one solution, then for any integer a', it follows that a' is a solution iff $a - a'$ is divisible by both m and n. Since m and n are coprime, the set of integers that are divisible by both m and n is precisely the set of integers divisible by mn. So the set of solutions is precisely the congruence class of $a \bmod mn$, as claimed. □

Remark 3.13 (i) In more abstract language, we've shown that the natural map from $\mathbb{Z}/mn\mathbb{Z}$ to the *direct product* $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ is a bijection. Since it's also a ring homomorphism, these two rings are isomorphic.

(ii) Note that we can compute everything here explicitly, using Euclid's algorithm applied to (m, n) as the starting point.

(iii) By induction on k, one can prove the following more general theorem: if $m_1, \dots, m_k \in \mathbb{N}_+$ are *pairwise coprime*[a], and $x_1, \dots, x_k$ are arbitrary integers, then we can find an $a \in \mathbb{Z}$ with $a_i = x_i \bmod m_i$ for all i, and this a is uniquely determined modulo $m_1 m_2 \dots m_k$.

In particular, if $m = \prod_{i=1}^{k} p_i^{e_i}$ is the prime factorisation of m, then

$$\mathbb{Z}/m\mathbb{Z} \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z}).$$

[a] This means that there is no integer > 1 which divides *more than one* of the m_i. This is strictly stronger than requiring that there is no integer > 1 dividing *all* of the m_i; e.g. if $(m_1, m_2, m_3) = (6, 10, 15)$, then any two of the m_i have a prime in common, but there is no prime dividing all three.

CHAPTER 4

The group of units mod m

4.1 Units modulo m and the φ function

Recall that if R is a (commutative) ring, an element $r \in R$ is said to be *invertible*, or a *unit* in R, if there exists $s \in R$ such that $rs = 1$.

Proposition 4.1 *Let $m \in \mathbb{N}_+$, and $a \in \mathbb{Z}$. Then $a \bmod m$ is invertible in $\mathbb{Z}/m\mathbb{Z}$ if and only if $\gcd(a, m) = 1$.*

Proof We have

$$\begin{aligned} \gcd(a, m) = 1 &\iff \exists u, v \in \mathbb{Z} \text{ such that } ua + vm = 1 \\ &\iff \exists u, v \in \mathbb{Z} \text{ such that } au - 1 = vm \\ &\iff \exists u \in \mathbb{Z} \text{ such that } au = 1 \bmod m \\ &\iff a \bmod m \text{ is invertible.} \quad \square \end{aligned}$$

In particular, if p is prime, then any non-zero element in $\mathbb{Z}/p\mathbb{Z}$ is invertible, so $\mathbb{Z}/p\mathbb{Z}$ is not just a ring but a *field* (and conversely, if n is non-prime, then $\mathbb{Z}/n\mathbb{Z}$ is not a field.)

Definition 4.2 We write $U_m = (\mathbb{Z}/m\mathbb{Z})^\times$ for the units in $\mathbb{Z}/m\mathbb{Z}$ (as a group under multiplication); and we define a function $\varphi : \mathbb{N}_+ \to \mathbb{N}_+$ by $\varphi(m) = \#U_m$.

Concretely, $\varphi(m)$ is the number of integers in the range $\{0, \dots, m-1\}$ which are coprime to m. (By convention $\varphi(1) = 1$.)

Example 4.3

- We have $\varphi(12) = 4$, since the only integers in the range $\{0, \dots, 11\}$ that are coprime to 12 are $\{1, 5, 7, 11\}$.
- If p is prime, then $\varphi(p) = p - 1$, since every non-zero integer $< p$ is coprime to p.

Exercise 4.4 Notice that $\varphi(12)/12 = \frac{1}{3}$ is quite small. Can you find an integer with $\varphi(n)/n < \frac{1}{4}$?

Proposition 4.5 *If m, n are coprime, then we have an isomorphism $U_{mn} \cong U_m \times U_n$ (direct product of groups). In particular, $\varphi(mn) = \varphi(m)\varphi(n)$.*

Proof Thanks to the Chinese remainder theorem, we know that the rings $\mathbb{Z}/mn\mathbb{Z}$ and $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ are isomorphic. It follows that their unit groups are isomorphic; but the unit group of $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ is obviously just $U_m \times U_n$. □

This means $\varphi(n)$ is determined for all n by its values when n is a prime power, which are computed as follows:

Proposition 4.6 *If $n = p^k$ is a prime power, then $\varphi(p^k) = p^{k-1}(p-1)$.*

Proof An integer is coprime to p^k iff it is not a multiple of p. Out of the p^k integers $\{0, 1, \dots, p^{k-1} - 1\}$, exactly p^{k-1} of them are multiples of p. So $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$. □

Exercise 4.7

(a) Show that for any k there are only finitely many n with $\varphi(n) = k$.
(b) Does there exist an $n \in \mathbb{N}_+$ with $\varphi(n) = 14$?

Remark 4.8 *Carmichael's conjecture* is that for any k, if the equation $\varphi(n) = k$ has any solutions, then it has at least two solutions. (This has been an open problem for over 100 years.) ⊛

One of the main reasons for introducing φ is the following:

Theorem 4.9

(i) (**Euler's theorem**)*: Let $m \in \mathbb{N}_+$. Then for all $a \in \mathbb{Z}$ coprime to m, we have $a^{\varphi(m)} = 1 \bmod m$.*

(ii) (**Fermat's little theorem**)*: Let $p \in \mathbb{P}$. Then for all $a \in \mathbb{Z}$ with $p \nmid a$, we have $a^{(p-1)} = 1 \bmod p$. Moreover, for any $a \in \mathbb{Z}$ we have $a^p = a \bmod p$.*

Proof Euler's result is just Lagrange's theorem from group theory ("the order of any element of a group divides the size of the group") applied to the group U_m.

For Fermat's little theorem, specialising Euler's theorem shows that $a^{p-1} = 1 \bmod p$ for all a coprime to p, and it follows that $a^p = a \bmod p$. On the other hand, if a is not coprime to p, then $p \mid a$, so $a^p = a = 0 \bmod p$ and the result holds in this case too. □

Exercise 4.10 If $n \in \mathbb{N}$ satisfies $n > 1$ and $a^n = a \bmod n$ for all a, but n is not prime, then n is said to be a *Carmichael number*.

- Show that 561 is a Carmichael number. (Note $561 = 3 \times 11 \times 17$).
- Prove that the product of two distinct primes cannot be Carmichael.

4.2 Primitive roots

We'll now prove an important result about the structure of U_p for p prime. First we need a preparatory lemma:

Lemma 4.11 *For any $n \in \mathbb{N}_+$, we have*

$$\sum_{\substack{d \in \mathbb{N}_+ \\ d \mid n}} \varphi(d) = n.$$

Proof For each d dividing n, the map $r \mapsto \frac{n}{d} \cdot r$ gives a bijection between the sets

$$S_d = \{r \in \{0, \dots, d-1\} : \gcd(r, d) = 1\}$$

and

$$T_d = \left\{s \in \{0, \dots, n-1\} : \gcd(s, n) = \tfrac{n}{d}\right\}.$$

So we have $\#T_d = \#S_d = \varphi(d)$. However, each $s \in T = \{0, \dots, n-1\}$ lies in exactly one of the sets T_d; so the sum of their sizes must be $\#T = n$. □

Theorem 4.12 *If p is prime, then U_p is a* cyclic *group. That is, there exists an element $g \in U_p$ such that every $x \in U_p$ is equal to some power of g.*

Such a g is called a *primitive root* mod p.

Proof Note that a is a primitive root iff the order of a in U_p is exactly $p-1$ (so Euler's theorem is the "best possible" bound).

Let $n = p - 1$; and for $d \mid n$, let $\psi(d)$ denote the number of elements of U_p whose order is precisely d. We claim that for any $d \mid n$, either $\psi(d) = 0$, or $\psi(d) = \varphi(d)$.

To see this, suppose $\psi(d) > 0$. Then there exists *some* element a of order exactly d. Hence the set $\{1, a, \dots, a^{d-1}\}$ has d distinct elements, and all of them have order dividing d; that is, they are roots of $X^d - 1$. Since this polynomial has degree d (and $\mathbb{Z}/p\mathbb{Z}$ is a field), it can't have more than d roots in $\mathbb{Z}/p\mathbb{Z}$. So our set is actually *all* of the elements of U_p of order dividing d. In particular, $\psi(d)$ is the number of h in $\{0, \dots, d-1\}$ such that a^h has order exactly d. However, a^h has order exactly d iff h is coprime to d; so we conclude that $\psi(d) = \varphi(d)$.

So it certainly follows that $\psi(d) \leqslant \varphi(d)$ for every d. But every element of U_n must have some order, so

$$\sum_{d \mid n} \psi(d) = n = \sum_{d \mid n} \varphi(d).$$

It follows that in fact $\psi(d) = \varphi(d)$ for all d, and in particular $\psi(n) = \varphi(n)$. As $\varphi(n) > 0$, this shows that elements of order exactly n exist. □

Example 4.13 The integer 2 is a primitive root mod 11: we have

$$\{1, 2, 2^2, \dots, 2^{10}\} = \{1, 2, 4, 8, 5, 10, 9, 7, 3, 6\} = U_{11}.$$

However, 2 isn't a primitive root modulo 5 or 7.

(*Artin's primitive root conjecture* predicts that there are infinitely many primes p such that 2 is a primitive root mod p. This is an open problem. ⊛)

Exercise 4.14 (hard!) The converse of Theorem 4.12 is false: for instance, $(\mathbb{Z}/18\mathbb{Z})^\times$ is cyclic (but 18 is clearly not prime). Can you classify, in terms of their prime factorisations, which integers n have the property that U_n is cyclic?

CHAPTER 5

Computing in U_n and RSA cryptography

As well as being interesting just from a pure theoretical standpoint, the group of units U_n is highly important in a major real-world application of number theory: *cryptography* – devising codes for securely transmitting secret information.

5.1 Powers mod n

Suppose we want to compute $3^{123456789} \bmod 7$ (more precisely: to compute the unique representative in $\{0, \dots, 6\}$ of its congruence class). How might we do this? One obvious idea would be to compute $3^{1234\cdots}$ as an integer, and then reduce it modulo 7.

This would work, eventually, but it would be a horrendous mess, because $3^{123456789}$ is *huge*, with millions of digits. So we need a better approach.

Using the φ function: Since 7 is prime, we know that $\varphi(7) = 6$; and $123456789 = 3 \bmod 6$, so it is $6q + 3$ for some q. Since 2 is coprime to 6, we conclude that

$$3^{123456789} = 3^{6q+3} = (3^6)^q \cdot 3^3 = 1^q \cdot 27 = 6 \bmod 7.$$

This algorithm works very well if the modulus n is small (but the exponent is large), as in the previous example. But if n is a bit bigger, there are two problems.

Example 5.1 Compute $3^{123456789} \bmod 21311$.

Here we hit two snags. Firstly, to compute $\varphi(21311)$, we have to factor 21311 into primes (which is doable on a computer, but takes a while, and would rapidly become impractical for larger moduli). Secondly, even once we've computed $\varphi(21311) = 21000$ and $123456789 \bmod 21000 = 18797$, we still have to compute 3^{18797}, which has about 9000 digits! So this is clearly not a sensible method.

We'll do this by a method called **repeated squaring**.

The idea is to write the exponent as a *sum of powers of 2*, which we can always do; this is just the binary expansion of n. (This is easy to compute from the base-10 expansion, and if you're working on a computer, the computer probably converted your input to binary as soon as you entered it.) Now, we can easily make a table of values of $3^{(2^i)} \bmod 21311$

for small i by repeated squaring:

$$\begin{aligned} 3^2 &= 3^2 = 9 \\ 3^4 &= 9^2 = 81 \\ 3^8 &= 81^2 = 6561 \\ 3^{16} &= 6561^2 = 19812 \\ 3^{32} &= 19812^2 = 9346 \\ 3^{64} &= 9346^2 = 15238 \\ &\vdots \end{aligned}$$

Because we reduce modulo $n = 21311$ *after every squaring step*, we never have to deal with integers bigger than n^2, so the computations are manageable. Once we have computed a table of $3^{(2^i)}$ for all i up to 26, we can use the formula

$$123456789 = 2^{26} + 2^{25} + 2^{24} + 2^{22} + \cdots + 2^2 + 1,$$

to compute $3^{123456789} = 20878 \bmod 21311$.

Remark 5.2 There's nothing very special about U_n here: if G is a finite group, and you have a practical way of representing elements of G on a computer and calculating the group operation, then you can use repeated squaring to efficiently compute g^n for any $g \in G$ and $n \in \mathbb{N}$.

5.2 Polynomial vs. exponential time

To formalise the ideas of "hard to compute" versus "easy to compute", we use the notion of *polynomial-time* and *exponential-time* algorithms. These compare the number of steps needed for some computational method, as a function of the *length of the input* (the amount of space required to write it down) – e.g. the number of decimal (or binary) digits needed to write down an integer. We say some algorithm is *polynomial-time* if the number of steps required, for input of length N, is bounded above by a constant multiple of N^k for some constant k. Similarly, if it's bounded above by a constant multiple of C^N for some C, we say it's *exponential-time*. Since exponentials grow much faster than polynomials, any polynomial-time algorithm will eventually beat any exponential-time one.

Remark 5.3 Note that since we ignore constant factors, it doesn't matter exactly how we measure the input length, as long as we stay within a constant factor of the original measure. E.g. if the input is a number, we could count its decimal digits, or its binary digits (bits); since these differ by a factor $\log_2 10$, this does not change whether an algorithm is polynomial or exponential time.

For example, computing the product $a \cdot b$ of two integers via the standard school-book "long multiplication" method requires approximately $N_a N_b$ steps, where N_r is the number of binary digits of r. Since $N_a N_b \leqslant \frac{1}{4}(N_a + N_b)^2$, and $N_a + N_b$ is the total length of the input, this is clearly a polynomial-time algorithm. The "repeated squaring" algorithm above, for computing $a^b \bmod N$, is also polynomial-time.

On the other hand, testing whether a number r is prime by trying all potential factors up to $\sqrt{r}$ ("trial division") involves at least $\sqrt{r}$ steps, which is clearly exponential in N_r.

There's a big difference here between *primality testing* – answering the yes/no question "is N prime?" – and *factorisation* – computing the prime factors of N. These might seem like the same problem, but they aren't: there are situations where you know N cannot be prime without being able to produce a specific factor of N.

Example 5.4 For instance, suppose you compute $2^{N-1} \bmod N$, and it's not 1. Then N cannot be prime, since otherwise it would contradict Fermat's little theorem). So you know that N has a nontrivial factor; but there's no obvious way to work out what that factor *is* using the information you have about $2^{N-1} \bmod N$.

- Primality testing can be done in polynomial time. This was proved by Miller in 1976 assuming an open conjecture in analytic number theory, the *generalised Riemann hypothesis*. In 2004, Agrawal, Kayal and Saxena gave a different algorithm, for which they could prove unconditionally (without assuming any conjectures) that it gave the correct answer in polynomial time.
- It's widely believed that factorisation *cannot* be done in polynomial time on a conventional computer[1]. There are algorithms (such as the *Number Field Sieve*) which are much better than trial division, but they are still much slower than any polynomial time algorithm.

It is this "gap" – that the complexity of *factoring* integers grows much faster than the complexity of *testing* whether integers are prime – that is vitally important in many applications of number theory.

5.3 Public key cryptography

We'll now learn about applications to secure communication – the science of cryptography. This could be used by a spy sending intelligence reports back to his home base; or it could be something much more mundane, like you logging into your bank account from a smartphone. This has two steps: *encryption* – the process the sender uses to transform a message into a coded form – and *decryption*, the opposite process that recovers the readable text from the coded message.

Traditional cryptographic techniques (prior to the 1970's) relied on the existence of a *shared secret*: both *sender* and *recipient* needed to know some piece of information which, if revealed to an outsider, would allow them to read the secret message themselves. This can be difficult to achieve: it requires coordination in advance between the sender and recipient.

Remark 5.5 Sometimes the entire system *is* the shared secret; but then any security lapse means redesigning the whole system from scratch. So most practical cryptographic systems rely on choosing a "secret key" which is an arbitrary number, string of letters, etc; and then scrambling up the input message in a way that depends on this secret key. It doesn't matter if an attacker knows how the system

[1] "Conventional" as opposed to "quantum". Quantum computers could, theoretically, factorise large numbers much faster than any conventional computer could; but building quantum computers on a realistic scale has proved to be somewhat difficult. This would be an interesting project topic.

works, as long as they don't know the secret key that was used for a particular message. That way, if one of your agents is captured, you just need to choose a new key, not a whole new algorithm!

Public key cryptography refers to a class of systems where the information needed to *encrypt* a message is different from the information needed to *decrypt* it. In such systems, each participant has a *public key* and a *private key*. If Alice wants to send a message to Bob, she can encrypt it knowing only Bob's public key, but only someone knowing his private key can encrypt it again. So Bob doesn't need to tell Alice – or anybody else – what his private key is; and as long as he keeps his private key secret, he can announce his public key openly to the world, without compromising the security of the system.

Of course, such a system can only work if it is impossible to determine the private key from the public one without an impractically lengthy computation. This is where number theory comes in: primes and prime factorisations are a rich source of difficult calculations!

Remark 5.6 There are some obvious security holes, of course. If Bob asks Alice a question that has only a few possible answers (e.g. just "yes" or "no") then an attacker can try encrypting both "yes" and "no" with the public key. This will give two different gibberish messages, but if one of those exactly matches the gibberish message Alice has just radioed to Bob, then the attacker knows the message. (This is typically solved by padding messages with randomly chosen nonsense phrases.)

5.4 The RSA cryptosystem

The first practical public-key cryptosystem is the *RSA* algorithm, announced by Rivest, Shamir and Aldeman in 1977.[2]

RSA relies on the following observation: *factorising large numbers into primes is difficult*. If I give you two 20-digit numbers p, q, then you can compute $N = pq$ in a few minutes. But if I give you a 40-digit number, and tell you that it's the product of two 20-digit primes, then it would take a very long time indeed to compute those prime factors.

In RSA, each participant chooses the following data:

- two large prime numbers p, q;
- an *encryption exponent* e, with $1 < e < \varphi(pq) = (p-1)(q-1)$ and e coprime to $\varphi(pq)$.

They announce to the world the product $N = pq$ and the encryption exponent e, but keep the factors p and q secret. Using this secret information, they can compute the *decryption exponent*

$$d = e^{-1} \bmod \varphi(N).$$

Suppose one participant (Bob) wants to send information to another (Alice). Bob finds out Alice's modulus N and encryption exponent e. He converts his message into a series

[2] They were not in fact the first to discover it; 20 years later it was revealed that Brtitish security services had already discovered the algorithm in 1973, but kept the discovery secret, and Rivest et al. rediscovered it independently.

of chunks, each of which is represented by an integer m in the range $1 < m < N$, and for each chunk he computes

$$c = m^e \bmod N.$$

These c's are the encrypted message he transmits to Alice.

Alice then takes each chunk c and computes

$$c^d \bmod N = (m^e)^d = m^{de} \bmod N.$$

Since $de = 1 \bmod \varphi(N)$, this is just $m \bmod N$, recovering the original message.

The security of this system relies on the fact that it's impossible to compute $\varphi(N)$ from N without factorising N, and factorising large integers is hard – much harder than any of the other steps in the algorithm.

Example 5.7 Suppose Alice's public key is

$$N = 21311, \qquad e = 11$$

Bob wants to send the message "TINKER".
Bob converts this into 3-letter blocks 'TIN | KER' and converts each one into a number in base 26,

$$TIN = 13065, \qquad KER = 6881.$$

For the first block, he computes $13065^{11} = 2460 \bmod 21311$, and the second $6881^{11} = 14867 \bmod 21311$. So he sends the message 02460 14867.
Alice knows that $21311 = 101 \times 211$, so $\varphi(N) = 21000$, and hence the decryption exponent is 19091, since $11 \times 19091 = 21001$. So she just computes $2460^{19091} = 13065 \bmod N$, etc, and recovers the original message.

Remark 5.8 In real-world applications, p, q would be chosen so that N has roughly 600 digits (corresponding to 2048 binary bits). With keys this size, the encryption and decryption steps are still reasonably practical[a] (each encryption taking fractions of a second). However, to crack the code – computing the private key from the public one, by factorising N – would take longer than the age of the universe, even using all the computing power of Google's datacentres put together.

[a] That said, RSA is becoming less popular nowadays because other algorithms – typically based on *elliptic curves* – can offer similar levels of security while using smaller keys and quicker encryption/decryption times. The widely used elliptic-curve algorithm ECDSA, used with a 256-bit key, is estimated to be roughly as secure as RSA with a 3000-bit key.

CHAPTER 6

Quadratic residues

We're now going to investigate what the image of the *squaring* map $x \mapsto x^2$ on $\mathbb{Z}/m\mathbb{Z}$ looks like. The elements which are in the image have a special name:

Definition 6.1 We say $a \in \mathbb{Z}/m\mathbb{Z}$ is a *quadratic residue* (QR) modulo m if there exists $x \in \mathbb{Z}/m\mathbb{Z}$ with $x^2 = a$.

For example, in $\mathbb{Z}/6\mathbb{Z}$, we have

x	0	1	2	3	4	5
x^2	0	1	4	3	4	1

so $\{0, 1, 3, 4\}$ are quadratic residues mod 6, and $\{2, 5\}$ are not.

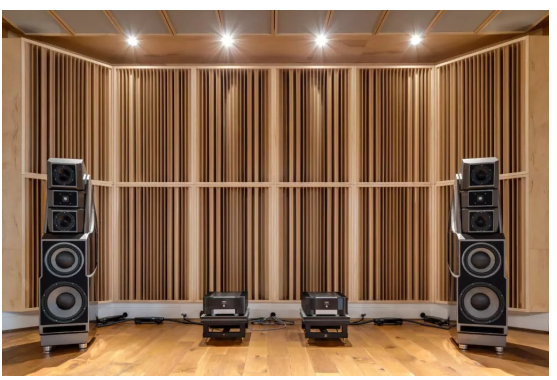

FIGURE 6.1. Quadratic residues are used in the design of echo-reducing wall panels for recording studios and concert halls. (Image: Dennis Foley, acousticfields.com)

6.1 Reducing to the prime case

From the Chinese remainder theorem, it's clear that if $n = \prod_i p_i^{e_i}$, then a is a QR mod n iff it's a QR modulo $p_i^{e_i}$ for all i. Rather less obvious is the following:

Proposition 6.2 *Let $p \in \mathbb{P}$ with $p > 2$, and let $a \in \mathbb{Z}$, with $p \nmid a$. If a is a QR mod p, then a is a QR mod p^k, for every $k \geqslant 1$.*

Proof Let's prove this by induction on k, the case $k = 1$ being true by assumption. So suppose $b \in \mathbb{Z}$ is such that $b^2 = a \bmod p^k$, for some $k \geqslant 1$, and let's try to cook up a solution modulo p^{k+1}.

By assumption, we have $b^2 = a + p^k r$, for some r. Let's consider integers of the form $b' = b + p^k s$. Then we have

$$(b')^2 = (b + p^k s)^2 = (a + p^k r) + 2bp^k s + p^{2k} s^2$$

and modulo p^{k+1} this is just $a + p^k(r + 2bs)$. So it suffices to show that we can choose s such that $r + 2bs = 0 \bmod p$.

Since $b^2 = a \neq 0 \bmod p$, and $p \neq 2$, it follows that $2b$ is a unit mod p, and we are done. □

Remark 6.3

- The argument breaks down for $p = 2$: if $a = 5$, then a is a QR modulo 2 and modulo 4, but not modulo 8. However, one can adapt the proof to show that an odd integer is a QR modulo every power of 2 iff it is 1 mod 8.
- The argument above is a preview of a much more general theorem called *Hensel's Lemma* which we'll see in the last chapter of the course.

6.2 QRs modulo primes

We can now concentrate on quadratic residues when the modulus is a **prime p with $p \neq 2$**. We first note that any nonzero quadratic residue mod p always has exactly 2 square roots mod p (if x is one, then $-x$ is the other, and $x \neq -x$). Since each unit mod p has to square to something, it follows that there are exactly $(p-1)/2$ nonzero quadratic residues; in other words, *exactly half* of the elements of U_p are squares.

Definition 6.4 Let p be an odd prime, and $a \in \mathbb{Z}$. Then the *Legendre symbol* is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a = 0 \bmod p, \\ 1 & \text{if } a \text{ is a nonzero quadratic residue mod } p, \\ -1 & \text{if } a \text{ is a non-residue mod } p. \end{cases}$$

Then we have the following:

Theorem 6.5 (Euler, 1748) *We have*

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

Proof If $a = 0 \bmod p$ the result is obvious, so assume $p \nmid a$. Then $\left(a^{(p-1)/2}\right)^2 = 1 \bmod p$ by Fermat's little theorem, so $a^{(p-1)/2}$ must be either 1 or -1 modulo p.

If $a = b^2 \bmod p$ for some x, then $a^{(p-1)/2} = b^{(p-1)} = 1$, again by Fermat's little theorem. So every nonzero QR is a root of $X^{(p-1)/2} - 1 = 0$. However, since this polynomial has degree $(p-1)/2$, and we've just exhibited $(p-1)/2$ roots of it, there can't be any more. So all quadratic non-residues a must satisfy $a^{(p-1)/2} = -1 \bmod p$. □

Here's an easy consequence:

Proposition 6.6 *-1 is a quadratic residue modulo the odd prime p if $p = 1 \bmod 4$, and a non-residue if $p = 3 \bmod 4$.* □

Another important consequence is the *multiplicativity* of the Legendre symbol:

Corollary 6.7 *For any integers a, b we have*

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

Proof Since both sides are equal to 0 or ± 1, and $p > 2$, it suffices to show that $\left(\frac{ab}{p}\right)$ and $\left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$ are congruent mod p. But this follows from Euler's criterion and the formula

$$(ab)^{(p-1)/2} = a^{(p-1)/2} b^{(p-1)/2}.$$ □

Remark 6.8 It's quite a strange and surprising thing that the product of two *non*-squares is always a square. This can be seen in an elementary way as follows. Take $a \in U_p$ which isn't a square, and consider the map $U_p \to U_p$ sending b to ab. This is a bijection; and it sends squares to non-squares, because if $b = x^2$ and $ab = y^2$ are both (nonzero) squares, then $a = (y/x)^2$ would have to be a square itself. Since there are equally many squares and non-squares, that "uses up" all the possible non-square images. Hence the non-squares have to go to squares, i.e. if b is non-square then ab is square.

Exercise 6.9 How many quadratic residues are there mod 15? How many of the *units* mod 15 are quadratic residues?

Give an example of integers a, b such that a, b and ab are all units and all quadratic non-residues mod 15.

Exercise 6.10 Suppose that there are finitely many primes p with $p = 1 \bmod 4$. By considering the integer $4(p_1 \dots p_k)^2 + 1$ where $\{p_1, \dots, p_k\}$ is the set of all such primes, deduce a contradiction.

CHAPTER 7

The reciprocity law

7.1 The statement

In the previous section the prime p was fixed, and we are asking "which a are quadratic residues mod p"? But we can also do something else: we can fix an integer a, and ask "for which (odd) primes $p \nmid a$ is a a quadratic residue mod p?" For instance, with $a = 5$, we see the following:

Residue: $\{11, 19, 29, 31, 41, 59, 61, 71, \dots\}$

Non-residue: $\{3, 7, 13, 17, 23, 37, 43, 47, \dots\}$

Notice the last digits! Amazingly, the answer seems to depend only on p mod 5 – which is strange, since the question is about $5 \bmod p$, not $p \bmod 5$, and these are totally different things.

If you try other values of a, the answer doesn't always depend on $p \bmod a$, but it's not far off – it suffices to know $p \bmod 4a$. This is the first hint at the following beautiful and important theorem:

Theorem 7.1 (Gauss' law of quadratic reciprocity) *If p, q are two distinct odd primes, then*

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)}{2} \cdot \frac{(q-1)}{2}} = \begin{cases} 1 & \text{if at least one of } p, q \text{ is } 1 \bmod 4 \\ -1 & \text{if both are } 3 \bmod 4. \end{cases}$$

Along with Gauss' law there are two related theorems (the "supplements to quadratic reciprocity") – one for $a = -1$ (which we have already proved as Proposition 6.6 above), and another for $a = 2$ (which will be Theorem 7.6 below). These say that for any odd prime p we have

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{(p-1)}{2}} = \begin{cases} 1 & \text{if } p = 1 \bmod 4 \\ -1 & \text{if } p = 3 \bmod 4 \end{cases}$$

and

$$\left(\frac{2}{p}\right) = (-1)^{\frac{(p^2-1)}{8}} = \begin{cases} 1 & \text{if } p = \pm 1 \bmod 8 \\ -1 & \text{if } p = \pm 3 \bmod 8. \end{cases}$$

The quadratic reciprocity law has many different proofs; Gauss himself published six different proofs in his lifetime, and hundreds more have been found since. However, none of them are particularly easy – whichever why you approach it, you have to do some genuine work. We'll give a proof shortly, which is quite close to one of Gauss' original arguments. First, we note that this does explain the observations above:

Corollary 7.2 *Let $a \in \mathbb{Z}$ be non-zero, and p, q odd primes, not dividing a, such that $p = q \bmod 4|a|$. Then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.*

Proof Considering the prime factorisation of $|a|$ and using the multiplicativity of the Legendre symbol, we may suppose that we are in one of three cases: $a = -1$, $a = 2$, or a is an odd prime. The first two cases are OK by the two supplementary laws, so we suppose we are in the third case.

Since $p = q \bmod 4|a|$, either $p = q = 1 \bmod 4$ or $p = q = 3 \bmod 4$. If $p = q = 1 \bmod 4$, or if $a = 1 \bmod 4$, then we have

$$\left(\frac{a}{p}\right) = \left(\frac{p}{a}\right) = \left(\frac{q}{a}\right) = \left(\frac{a}{q}\right).$$

If a, p, q are all $3 \bmod 4$, then we have similarly

$$\left(\frac{a}{p}\right) = -\left(\frac{p}{a}\right) = -\left(\frac{q}{a}\right) = \left(\frac{a}{q}\right). \qquad \square$$

7.2 Gauss' Lemma

Proposition 7.3 (Gauss' Lemma) *Let p be an odd prime, and let $L \subset U_p$ be the set given by the residue classes of the integers $\{1, \dots, \frac{p-1}{2}\}$. Then $\left(\frac{a}{p}\right) = (-1)^s$, where s is the number of $x \in L$ such that $ax \notin L$.*

Example 7.4 Take $p = 13$ and $a = 11$; then we reduce $11, 22, 33, 44, 55, 66$ modulo 13 to $-2, -4, -6, 5, 3, 1$. As expected by the proof of the Proposition, these are, up to sign, the integers between 1 and 6; and the minus sign appears exactly 3 times, so $\left(\frac{11}{13}\right) = (-1)^3 = -1$.
On the other hand, if $p = 13$ and $a = 10$, we reduce $10, 20, 30, 40, 50, 60$ to $-3, -6, 4, 1, -2, -5$, with 4 minus signs, so $\left(\frac{10}{13}\right) = (-1)^4 = 1$. Indeed, we check that $6^2 = 36 = 10 \bmod 13$.

Proof Let us compute $\prod_{x \in L} ax$.

On the one hand, we can pull out all the factors of L and get

$$\prod_{x \in L} ax = a^{(\#L)} \cdot \prod_{x \in L} x = a^{(p-1)/2} \left(\tfrac{p-1}{2}\right)!.$$

On the other hand, for $x \in U_p$, exactly one of x and $-x$ is in L; let's write $\lambda(x) = x$ if $x \in L$, and $\lambda(x) = -x$ otherwise, and $\epsilon(x) = x/\lambda(x) \in \{\pm 1\}$. So we can write

$$\begin{aligned}\prod_{x \in L} ax &= \prod_{x \in L} \epsilon(ax)\lambda(ax) \\ &= \left(\prod_{x \in L} \epsilon(ax)\right) \cdot \left(\prod_{x \in L} \lambda(ax)\right) \\ &= (-1)^s \cdot \left(\prod_{x \in L} \lambda(ax)\right),\end{aligned}$$

since the first product has s terms -1 and all the rest $+1$.

We claim $x \mapsto \lambda(ax)$ is a bijection $L \to L$, so the second product is just a re-ordering of the product $\prod_{x\in L} x = \left(\frac{p-1}{2}\right)!$. It suffices to show the map is injective, since L is finite. If $\lambda(ax) = \lambda(ay)$ for some $x, y \in L$, then $ax = \pm ay$. Since a is a unit, this implies $x = \pm y$, and since $x, y \in L$, this forces $x = y$, as required.

So we have $\prod_{x\in L} ax = (-1)^s \left(\frac{p-1}{2}\right)!$. Comparing this with the previous formula for the same product, we have

$$a^{(p-1)/2} \left(\tfrac{p-1}{2}\right)! = (-1)^s \left(\tfrac{p-1}{2}\right)!$$

and cancelling the (nonzero) common factor $\left(\frac{p-1}{2}\right)!$ gives the lemma. □

Exercise 7.5 The quantity $\left(\frac{p-1}{2}\right)! \bmod p$, which appears in the above proof, turns out to be a very interesting quantity! Can you show that

$$\left(\left(\tfrac{p-1}{2}\right)!\right)^2 \bmod p = \begin{cases} 1 & \text{if } p = 1 \bmod 4 \\ -1 & \text{if } p = 3 \bmod 4? \end{cases}$$

[Notice that this gives a direct proof that -1 is a square mod p if $p = 1 \bmod 4$. On the other hand, if $p = 3 \bmod 4$, then we must have either $\left(\frac{p-1}{2}\right)! = 1$ or $\left(\frac{p-1}{2}\right)! = -1$. Can you spot any pattern governing which case occurs? (⊛)

Theorem 7.6 *For p an odd prime, we have*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} 1 & \text{if } p = \pm 1 \bmod 8 \\ -1 & \text{if } p = \pm 3 \bmod 8. \end{cases}$$

Proof Clearly, for an integer $a \in \{1, \dots, \frac{p-1}{2}\}$, we have $2a \in L$ if $1 \leqslant a \leqslant \frac{p-1}{4}$, and $2a \notin L$ if $\frac{p-1}{4} < a \leqslant \frac{p-1}{2}\}$. So $\left(\frac{2}{p}\right)$ is $(-1)^s$, where $s = \#\{a : \frac{p-1}{4} < a \leqslant \frac{p-1}{2}\}$.

Now we consider cases:

- $p = 8q + 1$: then $s = \#\{a : 2q < a \leqslant 4q\} = 2q$ even
- $p = 8q + 3$: then $s = \#\{a : 2q + \frac{1}{2} < a \leqslant 4q + 1\} = 2q + 1$ odd
- $p = 8q + 5$: then $s = \#\{a : 2q + 1 < a \leqslant 4q + 2\} = 2q + 1$ odd
- $p = 8q + 7$: then $s = \#\{a : 2q + \frac{3}{2} < a \leqslant 4q + 3\} = 2q + 2$ even. □

7.3 Eisenstein's lemma and the final proof

To complete the proof of Quadratic Reciprocity we need one more lemma.

Proposition 7.7 (Eisenstein's Lemma) *For a odd, let*

$$t = \sum_{i=1}^{(p-1)/2} \left\lfloor \tfrac{ia}{p} \right\rfloor,$$

where $\lfloor x \rfloor$ denotes, as usual, the greatest integer $\leqslant x$. Then $\left(\frac{a}{p}\right) = (-1)^t$.

Proof It suffices to show that $t = s \bmod 2$, where s is as in Gauss' lemma. For each i, we write

$$ia = p \cdot \lfloor \tfrac{ia}{p} \rfloor + r_i$$

where r_i is the unique integer in $\{1, \dots, p-1\}$ congruent to $ia \bmod p$. Adding these equations together, and reducing mod 2 (remembering that a and p are odd), we obtain

$$(1 + \cdots + \tfrac{p-1}{2}) = t + \sum r_i.$$

As in the proof of Gauss' lemma, for each $\ell \in L$, exactly one of ℓ or $p - \ell$ occurs among the r_i. Moreover, the number of times that $p - \ell$ occurs is s. Since $p - \ell = 1 + \ell \bmod 2$, we have $\sum r_i = s + (1 + \cdots + \frac{p-1}{2})$. Plugging this into the above, we deduce that

$$(1 + \cdots + \tfrac{p-1}{2}) = t + s + (1 + \cdots + \tfrac{p-1}{2}) \bmod 2,$$

so $t = -s = s \bmod 2$. □

Proof of Quadratic Reciprocity We're going to do this by "counting lattice points". Consider the rectangle R in the (x, y) plane with vertices at $(0, 0)$, $(p/2, 0)$, $(p/2, q/2)$ and $(0, q/2)$. The diagonal $y = \frac{q}{p}$ x divides this into two triangles.

We want to count the pairs $(x, y) \in \mathbb{Z}^2$ which lie in the interior of R. Evidently there are exactly $\frac{p-1}{2} \cdot \frac{q-1}{2}$ of these in total; and none of them lie exactly on the diagonal (since p and q are distinct primes).

On the other hand, how many lie below the diagonal? For each $i = 1, \dots, \frac{p-1}{2}$, the number of points below the diagonal in the "vertical column" with $x = i$ is exactly $\lfloor \frac{iq}{p} \rfloor$. Thus the total number is the quantity t from the last lemma; so $(-1)^t = \left(\frac{q}{p}\right)$.

Reversing the roles of p and q, the number of lattice points above the diagonal, t', satisfies $(-1)^{t'} = \left(\frac{p}{q}\right)$. Since every point must lie above or below the diagonal, $t + t' = \frac{(p-1)}{2} \cdot \frac{(q-1)}{2}$, and hence

$$\left(\frac{q}{p}\right) \cdot \left(\frac{p}{q}\right) = (-1)^{t+t'} = (-1)^{\frac{(p-1)}{2} \cdot \frac{(q-1)}{2}}.$$

which proves the theorem. □

CHAPTER 8

Gaussian integers

Having investigated the arithmetic of $\mathbb{Z}$ quite thoroughly, we're now going to look at how factorisation, primes, etc work out in some other algebraic structures – in particular, some subrings of the complex numbers which behave a bit like $\mathbb{Z}$.

8.1 Definitions

Definition 8.1 For any $\alpha \in \mathbb{C}$, we let $\mathbb{Z}[\alpha]$ denote the subgroup of $(\mathbb{C}, +)$ generated by the powers $\{1, \alpha, \alpha^2, \dots\}$.

This is clearly a sub*ring* of $\mathbb{C}$, not just an additive subgroup, and in fact it's the smallest subring containing α. It is always an integral domain (since it's a subring of $\mathbb{C}$, which is a field and hence an integral domain, and any subring of an integral domain is an integral domain.)

Definition 8.2 The ring of *Gaussian integers* is the ring $\mathbb{Z}[i]$, where $i = \sqrt{-1}$ as usual.

Since $i^2 = -1$, any element of $\mathbb{Z}[i]$ can be written uniquely as $a + bi$ for some $a, b \in \mathbb{Z}$; so $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Z}^2$ as an additive group. We can visualise it as a "square lattice" inside the complex plane:

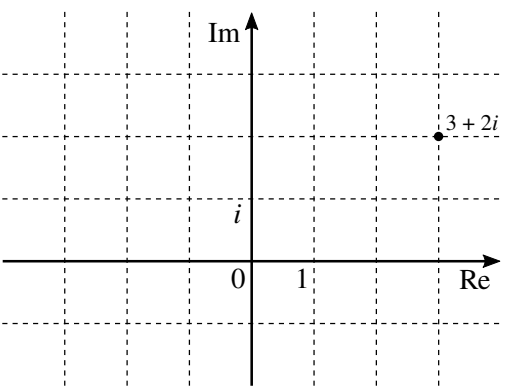

FIGURE 8.1. Gaussian integer grid (image: Wikipedia)

Definition 8.3 If $x = a + bi \in \mathbb{Z}[i]$, we define $N(x) = |x|^2 = a^2 + b^2$.

Note that $N(x) \in \mathbb{N}$, and $N(xy) = N(x)N(y)$. Moreover, we have $N(x) = x\bar{x}$, where $\bar{x}$ is the complex conjugate of x (which is in $\mathbb{Z}[i]$ if x is).

Let's use this to compute the *units* in $\mathbb{Z}[i]$. If x is invertible in $\mathbb{Z}[i]$, then $N(x)$ is invertible in $\mathbb{N}$; so it must be 1. Conversely, if $N(x) = 1$, then x is invertible, since its inverse is $\bar{x}$. So the units are exactly the x with $N(x) = 1$.

However, for integers a, b we have $a^2 + b^2 > 1$ unless $(a, b) = (\pm 1, 0)$ or $(0, \pm 1)$. So we've shown that:

Proposition 8.4 *The set $\mathbb{Z}[i]^\times$ of invertible Gaussian integers consists precisely of $\{1, -1, i, -i\}$.*

So we have more invertible elements than we do in $\mathbb{Z}$ (where the only units are ± 1). This means we need to take care of them when making divisibility statements. So we'll introduce the following notation:

Definition 8.5 We say $\alpha, \beta \in \mathbb{Z}[i]$ are *associates* if $\alpha = u\beta$ for a unit u.

Clearly this is an equivalence relation; moreover, α and β are associates iff $\alpha \mid \beta$ and $\beta \mid \alpha$.

8.2 Euclidean division

Proposition 8.6 *Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\alpha \neq 0$. Then there exist $\kappa, \rho \in \mathbb{Z}[i]$ such that*

- $\beta = \kappa\alpha + \rho$,
- $0 \leqslant N(\rho) < N(\alpha)$.

Proof Let $q = \beta/\alpha \in \mathbb{C}$. Clearly $q = u + vi$ with $u, v \in \mathbb{Q}$; but u and v won't necessarily be in $\mathbb{Z}$.

We shall define $\kappa = x + yi \in \mathbb{Z}[i]$ by rounding u, v to the nearest integer, so that $|x - u| \leqslant \frac{1}{2}$ and $|y - v| \leqslant \frac{1}{2}$. Then we compute that

$$\rho = \beta - \kappa\alpha = \alpha\left((u + vi) - (x + yi)\right).$$

Since the norm on $\mathbb{C}$ is multiplicative, we have

$$N(\rho) = N(\alpha) \cdot \left((u - x)^2 + (v - y)^2\right).$$

But both $|u - x|$ and $|v - y|$ are $\leqslant \frac{1}{2}$, so $(u - x)^2 + (v - y)^2 \leqslant (\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. So we've shown that

$$N(\rho) \leqslant \frac{1}{2} N(\alpha) < N(\alpha).$$

□

Example 8.7 For $\beta = 11 + 8i$ and $\alpha = 2 + 3i$, we compute

$$\frac{\beta}{\alpha} = \frac{(11 + 8i)(2 - 3i)}{(2 + 3i)(2 - 3i)} = \frac{46}{13} + \frac{-17}{13} i.$$

Rounding $\frac{46}{13}$ and $\frac{-17}{13}$ to the *nearest* integers we obtain $\kappa = u + vi = 4 - i$, and hence

$$\rho = \beta - \kappa\alpha = (11 + 8i) - (2 + 3i) \cdot (4 - i) = -2i.$$

Remark 8.8 Note that, unlike in the case of $\mathbb{Z}$, we haven't claimed any uniqueness for κ and ρ. Can you find a different pair (κ, ρ) which also works, for the same (α, β) as above?

Proposition 8.6 is precisely the statement that $\mathbb{Z}[i]$ is a *Euclidean ring* (Algebra, Chapter 9). This is exactly what we need to make the Euclidean algorithm work in $\mathbb{Z}[i]$: for any two elements α, β there exists an (explicitly computable) element $\gcd(\alpha, \beta)$, well-defined up to multiplication by units, such that we have

$$\forall x \in \mathbb{Z}[i], \qquad x \mid \gcd(\alpha, \beta) \iff x \mid \alpha \text{ and } x \mid \beta.$$

Moreover, $\gcd(\alpha, \beta)$ can always be written as $r\alpha + s\beta$ for $r, s \in \mathbb{Z}[i]$.

Remark 8.9 Note that in general there are four equally valid possibilities for the GCD – it is only well-defined up multiplication by $\{\pm 1, \pm i\}$ and there's no obvious "best" choice among these four options.

Example 8.10 From the calculation above, $\gcd(11+8i, 2+3i) = \gcd(2+3i, -2i)$. We also have

$$(2 + 3i) = (-1 + i) \cdot (-2i) + i,$$

so

$$\gcd(2 + 3i, -2i) = \gcd(-2i, i).$$

Since i is a unit, this shows that $11 + 8i$ and $2 + 3i$ are coprime in $\mathbb{Z}[i]$.

Corollary 8.11 *Let $\alpha \in \mathbb{Z}[i]$. Then the following are equivalent:*

- *α is an* indecomposable element*: that is, if $\beta \mid \alpha$, then either β is a unit or it is an associate of α.*
- *α is a* prime element*: that is, if $\rho, \sigma \in \mathbb{Z}[i]$ and $\alpha \mid \rho\sigma$, then $\alpha \mid \rho$ or $\alpha \mid \sigma$.*

Proof Cf. *Algebra*, Prop 9.21. Since $\mathbb{Z}[i]$ is Euclidean, it is a PID; and in a PID, prime elements and indecomposable elements coincide. Alternatively, we can repeat exactly the same argument as for $\mathbb{Z}$, using Euler's Lemma 1.20. □

Remark 8.12 Recall that you saw in *Algebra* that in the similar-looking ring $\mathbb{Z}[\sqrt{-5}]$, the element 3 is indecomposable but not prime, since it divides $(1 - \sqrt{-5})(1+\sqrt{-5})$ but doesn't divide either factor. So this is something rather special about $\mathbb{Z}[\sqrt{-1}]$.

Corollary 8.13 (Fundamental Theorem of Arithmetic for $\mathbb{Z}[i]$) *Any non-zero $\alpha \in \mathbb{Z}[i]$ can be written as a product of prime elements. Moreover, if*

$$\alpha = \pi_1\pi_2 \dots \pi_r = \mu_1\mu_2 \dots \mu_s$$

are two factorisations of α as products of prime elements, then $r = s$, and we can re-order the factors so that μ_i is an associate of π_i for $i = 1, \dots, r$.

Exactly as before, we can also gather together the factors and write

$$\alpha = u \cdot \prod_i \pi_i^{e_i},$$

with u a unit, $e_i \in \mathbb{N}_+$, and π_i primes which are pairwise non-associate.

8.3 Gaussian primes

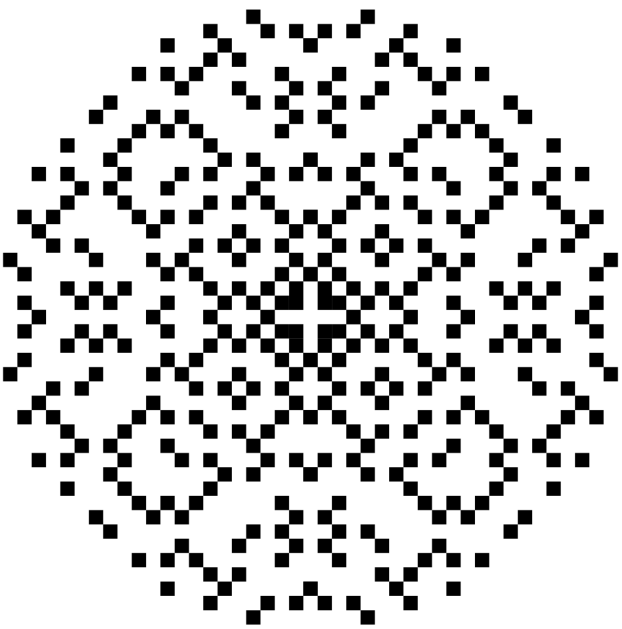

FIGURE 8.2. Gaussian primes α with $N(\alpha) \leqslant 500$ (image: Wikipedia)

We'll now classify all the primes in $\mathbb{Z}[i]$. We start with the following easy remark:

Proposition 8.14 *Suppose $\alpha \in \mathbb{Z}[i]$ is a prime element. Then there is a unique prime integer $p \in \mathbb{P}$ such that α divides p. (We say α* lies above *the prime integer p.)*

Proof Consider the norm $N(\alpha)$, which is a non-zero integer. Since $\alpha\bar{\alpha} = N(\alpha)$, we have $\alpha \mid N(\alpha)$. From the factorisation theory of $\mathbb{Z}$, we can write $N(\alpha)$ as a product of prime integers; but since α is prime, it must divide one of these factors. This shows that α must divide some $p \in \mathbb{P}$. But if α divides two distinct elements $p, q \in \mathbb{P}$, then it must divide $mp + nq$ for all $m, n \in \mathbb{Z}$; so it must divide 1, which is a contradiction since α is not a unit. □

So we can study all Gaussian primes by asking, for each $p \in \mathbb{P}$, which Gaussian primes lie above it.

Proposition 8.15 *Given $p \in \mathbb{P}$, exactly one of the following two possibilities occurs:*

- *p factors as $\alpha\bar{\alpha}$, for some prime $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) = p$. Then the primes above p are the associates of α and $\bar{\alpha}$.*
- *p is itself a prime element of $\mathbb{Z}[i]$ (so the only primes above p are $\pm p$ and $\pm ip$).*

Moreover, the first case occurs if and only if there exist integers (x, y) with $p = x^2 + y^2$.

Proof First let us assume that (x, y) exists with $p = x^2 + y^2$. Then $\alpha = x + yi \in \mathbb{Z}[i]$ satisfies $N(\alpha) = p$. Since $N(\alpha) = \alpha\bar{\alpha}$, we have $\alpha \mid p$; and α must be indecomposable, and hence prime, since if α factors as a product $\beta\gamma$ then we must have $N(\beta)N(\gamma) = N(\alpha) = p$, so one of β and γ has norm 1 and is thus a unit. Since $N(\bar{\alpha}) = N(\alpha)$ we see that $\bar{\alpha}$ is also prime. Moreover, any prime above p must divide $\alpha\bar{\alpha}$; so it divides one of α and $\bar{\alpha}$, and must therefore be an associate of it, since they are both prime.

Conversely, if no such (x, y) exists, then p is indecomposable, since any nontrivial factor β of p would have to satisfy $N(\beta) = p$. □

We'd like to know which $p \in \mathbb{P}$ remain prime, and which do not. Clearly $p = 2$ factors as $(1 + i)(1 - i)$, so we can restrict to odd p. It turns out that the answer depends only on p mod 4. One direction is easy:

Proposition 8.16 *Let $p \in \mathbb{P}$. If $p = 3 \bmod 4$, then p is a Gaussian prime.*

Proof The only squares mod 4 are 0 and 1, so if $p = 3 \bmod 4$, the equation $x^2 + y^2 = p$ has no solutions mod 4 and hence no solutions in $\mathbb{Z}$. □

It turns out that the converse is also true, but this is a much deeper theorem:

Theorem 8.17 (Fermat) *Let $p \in \mathbb{P}$ with $p = 1 \bmod 4$. Then p is not a Gaussian prime. Equivalently, p is the sum of two integer squares.*

Proof Consider the equation $X^2 + 1 = 0 \bmod p$. This has a solution, since $p \neq 3 \bmod 4$. Choose $t \in \mathbb{Z}$ such that $t^2 + 1 = 0 \bmod p$; and let $\alpha = \gcd(t - i, p)$.

Clearly $\alpha \mid p$, but α is not an associate of p, since $p \nmid t - i$. Thus $N(\alpha) = 1$ or p; and it suffices to prove that $N(\alpha) \neq 1$, i.e. that $t - i$ and p aren't coprime in $\mathbb{Z}[i]$.

Consider the map

$$\lambda : \mathbb{Z}[i] \to \mathbb{F}_p, \qquad a + bi \mapsto a + tb \bmod p.$$

This is obviously compatible with addition; we claim it's also compatible with multiplication. This can be checked explicitly: if $u = a + bi$, $v = c + di$, then

$$\lambda(uv) = \lambda((ac - bd) + (ad + bc)i) = (ac - bd) + t(ad + bc) \bmod p,$$

while

$$\lambda(u)\lambda(v) = (a + tb)(c + td) = (ac + t^2 bd) + t(ad + bc) \bmod p,$$

and since $t^2 = -1 \bmod p$ these are the same.

Clearly λ kills both p and $t - i$. So if these elements were coprime, there would be u, v with $up + v(t - i) = 1$, and we'd have $1 = \lambda(up + v(t - i)) = 0 + 0 = 0 \bmod p$, which is a contradiction. So p and $t - i$ can't be coprime. □

Remark 8.18 Fermat announced that he had proved the theorem in a letter dated Christmas Day 1640, but he never revealed his method of proof (sound familiar?). Just like Quadratic Reciprocity, this theorem now has many different proofs; these include a famous "one-sentence proof" due to Don Zagier (link).

Exercise 8.19 Show that if $p = 1 \bmod 4$ and $\alpha \in \mathbb{Z}[i]$ satisfies $N(\alpha) = p$, then $\bar{\alpha}$ is not an associate of α. (Hint: if α divides $t - i$, for some $t \in \mathbb{Z}$ as above, then $\bar{\alpha}$ divides $t + i$.)

8.4 Euclidean rings

Recall the following construction from Algebra:

Definition 8.20 Let R be an integral domain. A *Euclidean function* on R is a map

$$\delta : R \setminus \{0\} \to \mathbb{N}$$

such that for every $a, b \in R$ with $b \neq 0$, we can find $q, r \in R$ with $a = bq + r$ such that either $r = 0$, or $\delta(r) < \delta(b)$.

We know that $\mathbb{Z}$, $k[X]$ for k a field, and $\mathbb{Z}[i]$ are examples of Euclidean domains. One can check similarly that $\mathbb{Z}[\sqrt{-2}]$ (i.e. the subring of $\mathbb{C}$ consisting of numbers of the form $a + b\sqrt{-2}$, with $a, b \in \mathbb{Z}$), is a Euclidean domain, with the Euclidean function again given by $N(x) = x\bar{x}$.

Exercise 8.21 Prove this. (You will need the fact that if $|p|, |q| \leqslant \frac{1}{2}$ then $p^2 + 2q^2 \leqslant \frac{3}{4} < 1$.)

It follows that factorisation in $\mathbb{Z}[\sqrt{-2}]$ works in just the same elegant way as before; the ring is a PID and a UFD, and and we can characterise exactly which primes remain prime in $\mathbb{Z}[\sqrt{-2}]$ in terms of congruences mod 8.

Exercise 8.22 (hard!) Show that an odd prime p has the form $x^2 + 2y^2$ iff $p = 1, 3 \bmod 8$, and not if $p = 5, 7 \bmod 8$.
[Hint: First show that -2 is a square mod p iff $p = 1, 3 \bmod 8$, using the last two parts of Quadratic Reciprocity.]

8.5 The Eisenstein integers

On the other hand, the ring $\mathbb{Z}[\sqrt{-3}]$ is *not* Euclidean. It can't be, because

$$(1 - \sqrt{-3})(1 + \sqrt{-3}) = 4 = 2 \cdot 2$$

and 2 does not divide either factor on the right. So 2 is not a prime element; but it is obviously indecomposable since $a^2 + 3b^2 = 2$ has no solutions. So $\mathbb{Z}[\sqrt{-3}]$ is not a PID, and hence not Euclidean either. However, we can fix this by embedding $\mathbb{Z}[\sqrt{-3}]$ inside a slightly larger ring:

Definition 8.23 The *Eisenstein integers* is the subring $\mathbb{Z}[\omega] \subset \mathbb{C}$, where $\omega = \frac{-1+\sqrt{-3}}{2}$.

This clearly contains $\mathbb{Z}[\sqrt{-3}]$ (since $\sqrt{-3} = 2\omega + 1$), but it is slightly larger, since $\omega \notin \mathbb{Z}[\sqrt{-3}]$.

One can check that $\mathbb{Z}[\omega]$ consists precisely of the linear combinations $a + b\omega$ with $a, b \in \mathbb{Z}$. This is because ω satisfies the equation $\omega^2 = -1 - \omega$, and we can use this (and induction on n) to show that ω^n is a $\mathbb{Z}$-linear combination of 1 and ω for all $n \in \mathbb{N}$.

Exercise 8.24 Show that the abelian-group quotient $\mathbb{Z}[\omega]/\mathbb{Z}[\sqrt{-3}]$ has order 2.

One can picture $\mathbb{Z}[\omega]$ as a triangular lattice inside the complex plane, as in the following figure[1]:

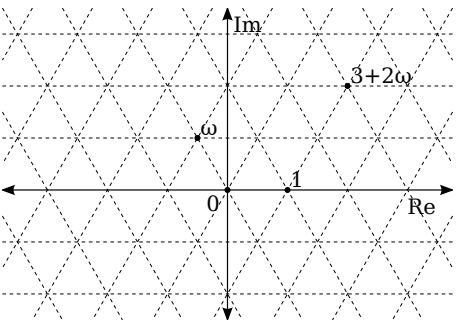

FIGURE 8.3. Eisenstein integer grid (image: Wikipedia)

Exercise 8.25 Find all the units of $\mathbb{Z}[\omega]$. (Hint: There are 6 of them.)

From Figure 8.3, it's easy to convince yourself that for every $x \in \mathbb{C}$, there exists $y \in \mathbb{Z}[\omega]$ with $|x - y| < 1$. (In fact we can do a little better: we're never more than $\frac{1}{\sqrt{3}} \cong 0.58$ away from an element of $\mathbb{Z}[\omega]$.) This suffices to show that $\mathbb{Z}[\omega]$ is Euclidean, with $N(x) = x\bar{x}$ as the Euclidean function, just as before. So $\mathbb{Z}[\omega]$ is a PID and a UFD; and we can deduce the following:

Proposition 8.26 *Let $p \in \mathbb{P}$. Then $p = N(\alpha)$ for some $\alpha \in \mathbb{Z}[\omega]$ if and only if $p = 1 \bmod 3$.*

Exercise 8.27 Can you show that, despite $\mathbb{Z}[\sqrt{-3}]$ not being a PID, nonetheless every prime that is 1 mod 3 has the form $x^2 + 3y^2$? (Hint: Show that if $\alpha \in \mathbb{Z}[\omega]$ then at least one of its associates lies in the subring $\mathbb{Z}[\sqrt{-3}]$.)

[1] From Wikipedia, with thanks to Wikipedia contributor `gunther`.

CHAPTER 9

Arithmetic in number fields

9.1 Algebraic integers

Remember the following definition from *Algebra*: let $\alpha \in \mathbb{C}$; then α is *algebraic* if there is a non-constant polynomial $f(X) \in \mathbb{Q}[X]$ with $f(\alpha) = 0$.

Of course, the set $\overline{\mathbb{Q}}$ of all algebraic numbers is too big to have any interesting factorisation theory (it's a field, so every non-zero element is a unit); we want to pick out the algebraic numbers which "don't have any denominators" in some sense. It turns out the good definition is the following:

Definition 9.1 We say $\alpha \in \mathbb{C}$ is an *algebraic integer* if there exists a *monic* polynomial $f(X) \in \mathbb{Z}[X]$ with $f(\alpha) = 0$. We write $\bar{\mathbb{Z}}$ for the set of algebraic integers.

Example 9.2 Clearly we have $\mathbb{Z} \subseteq \bar{\mathbb{Z}}$, since for any $n \in \mathbb{Z}$, $f(X) = X - n$ is a monic polynomial that it satisfies. Moreover, if $n \in \mathbb{Z}$ then $\sqrt{n} \in \bar{\mathbb{Z}}$.
Less obviously, $\omega = \frac{-1+\sqrt{-3}}{2} \in \bar{\mathbb{Z}}$, since it satisfies $X^2 + X + 1 = 0$.

Proposition 9.3 *For any algebraic number α, there exists some $N \in \mathbb{N}_+$ such that $N\alpha \in \bar{\mathbb{Z}}$.*

Proof Exercise. (Hint: if $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0 \in \mathbb{Q}[X]$ is the minimal polynomial of α, and $\beta = N\alpha$ for some N, then what is the minimal polynomial of β?) □

What's less obvious is how one would show that anything is *not* an algebraic integer! Fortunately, we have the following criterion:

Proposition 9.4 *An algebraic number $\alpha \in \mathbb{C}$ is an algebraic integer if and only if its* minimal *polynomial has integer coefficients.*

Proof Let $f \in \mathbb{Q}[X]$ be the minimal polynomial of α. If $f \in \mathbb{Z}[X]$, then clearly f is an algebraic integer.

Conversely, suppose f does not have integer coefficients, but there is some (larger-degree) monic integral polynomial h with $h(\alpha) = 0$. Then we must have $h(X) = f(X)g(X)$ for some $g \in \mathbb{Q}[X]$.

Let C be the least common multiple of the denominators of the coefficients of f, so that $Cf \in \mathbb{Z}[X]$, and similarly D for g. Then we clearly have $(Cf)(Dg) = (CD)h$. Now let p be a prime dividing CD. Clearly at least one coefficient of Cf is not divisible by p (since otherwise C/p would be the LCM of the denominators). Similarly at least one of the coefficients of Dg is not divisible by p. So $Cf \bmod p$ and $Dg \bmod p$ are non-zero in $\mathbb{F}_p[X]$. But their product CDh is zero, since $p \mid CD$ and h has integral coefficients. This contradicts the fact that $\mathbb{F}_p[X]$ is an integral domain. So CD must in fact be 1, i.e. both f and g are integral. □

Example 9.5

- If $x \in \mathbb{Q} - \mathbb{Z}$, then x is not an algebraic integer. (That is, we have $\bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}$).
- The number $\frac{1+\sqrt{3}}{2}$ is not an algebraic integer: it is a root of the polynomial $x^2 - x - \frac{1}{2}$, and since it clearly isn't in $\mathbb{Q}$, this must be the minimal polynomial.

It follows that a rational number is an algebraic integer iff it's an integer in the usual sense.

Exercise 9.6 (Warning!) Give a counterexample to show that is *not* true that if α is an algebraic integer, then every monic polynomial that f satisfies has to have integral coefficients.

For doing arithmetic with algebraic integers, the following characterisation is useful:

Proposition 9.7 *$\alpha \in \mathbb{C}$ is an algebraic integer if and only if $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group.*

Proof If α satisfies a polynomial $f(X) = X^n + a_{n-1}X^{n-1} + \dots$, then α^n is in the $\mathbb{Z}$-span of $1, \dots, \alpha^{n-1}$, and by induction one can show that α^{n+1}, α^{n+2} etc are also in this span.

Conversely, if this group is finitely generated, then each generator can only mention finitely many powers of α, so there is some N such that $\{1, \dots, \alpha^N\}$ is a generating set. Hence α^{N+1} is in the $\mathbb{Z}$-span of $\{1, \dots, \alpha^N\}$, giving a monic integral polynomial that α satisfies. □

Corollary 9.8 *If α, β are algebraic integers then so are $\alpha \pm \beta$ and $\alpha\beta$.*

Proof Suppose α, β satisfy polynomials of degree M, N respectively. Consider the subgroup of $\mathbb{C}$ generated by $\{\alpha^i\beta^j : 0 \leqslant i < N, 0 \leqslant j < M\}$. This is finitely generated and contains $\alpha^r\beta^s$ for all $r, s \in \mathbb{N}$, so in particular it contains $(\alpha\beta)^j$ and $(\alpha \pm \beta)^k$ for all j, k. Since a subgroup of a finitely generated abelian group is finitely generated, the result follows. □

Thus the set $\bar{\mathbb{Z}}$ of all algebraic integers is a subring of of $\mathbb{C}$.

Exercise 9.9 Find a monic polynomial $f(X) \in \mathbb{Z}[X]$ with $f(\sqrt{2} + \sqrt{3}) = 0$.

9.2 Number fields

Definition 9.10 A *number field* is a subfield $K \subset \mathbb{C}$ such that $[K : \mathbb{Q}] < \infty$ (i.e. K is finite-dimensional as a $\mathbb{Q}$-vector space).

Note that every number field consists of algebraic numbers. Conversely, if α is an algebraic number, then $\mathbb{Q}(\alpha)$, the field extension generated by α (cf. *Algebra* chapter 10) is a number field. (However, the field $\bar{\mathbb{Q}}$ of *all* algebraic numbers isn't a number field – it's too big.)

Definition 9.11 If K is a number field, then we define $\mathcal{O}_K$, the *ring of integers of* K, as $K \cap \bar{\mathbb{Z}}$.

Note that if α is an algebraic integer, $\mathbb{Z}[\alpha]$ might not be equal to the ring of integers of $\mathbb{Q}(\alpha)$. For instance, $\mathbb{Z}[\sqrt{-3}]$ is not the ring of integers of $\mathbb{Q}(\sqrt{-3})$, because it doesn't contain ω.

Proposition 9.12 (Rings of integers of quadratic fields) *Let $d \in \mathbb{Z}$ with $d \neq 1$, and suppose d is not divisible by n^2 for any $n > 1$ (d is "square-free"). Then the ring of integers of $\mathbb{Q}(\sqrt{d})$ is given by*

$$\mathcal{O}_{\mathbb{Q}(\sqrt{d})} = \begin{cases} \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] & \text{if } d = 1 \bmod 4, \\ \mathbb{Z}[\sqrt{d}] & \text{otherwise.} \end{cases}$$

Proof First, note that $\frac{1+\sqrt{d}}{2}$ is a root of $X^2 - X + \frac{1-d}{4}$, so it is an algebraic integer iff $d = 1 \bmod 4$.

Conversely, let $\alpha = u + v\sqrt{d}$ with $u, v \in \mathbb{Q}$, and suppose $\alpha \in \bar{\mathbb{Z}}$. Then $\alpha' = u - v\sqrt{d}$ is also in $\bar{\mathbb{Z}}$, since it satisfies the same polynomial that α does; and hence $\alpha + \alpha' = 2u \in \bar{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}$. Similarly, $\alpha - \alpha' = 2v\sqrt{d} \in \bar{\mathbb{Z}}$; thus $(2v)^2 d \in \mathbb{Z}$, but since d is squarefree, this implies that $2v \in \mathbb{Z}$.

So, if α is an algebraic integer but doesn't lie in $\mathbb{Z}[\sqrt{d}]$, then we can subtract a $\mathbb{Z}$-linear combination of 1 and $\sqrt{d}$ to deduce that one of $\{\frac{1}{2}, \frac{\sqrt{d}}{2}, \frac{1+\sqrt{d}}{2}\}$ is an algebraic integer. Clearly $\frac{1}{2}$ and $\frac{\sqrt{d}}{2}$ are never algebraic integers (since $4 \nmid d$); and $\frac{1+\sqrt{d}}{2}$ is an algebraic integer iff $d = 1 \bmod 4$. □

Remark 9.13 Note that $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ is isomorphic to $\mathbb{Z}^2$ as an abelian group: every element can be written uniquely in the form $a + b\lambda$ for $a, b \in \mathbb{Z}$, where $\lambda = \frac{1+\sqrt{d}}{2}$ or $\sqrt{d}$ respectively.
More generally, one can show that for any number field K, $\mathcal{O}_K$ is isomorphic to $\mathbb{Z}^d$ as an abelian group, where $d = [K : \mathbb{Q}]$. See Theorem 2.16 of Stewart & Tall.

We finish this chapter with a useful little result which will be helpful in the next chapter:

Proposition 9.14 *For any number field K and any non-zero $\alpha \in \mathcal{O}_K$, there exists a non-zero $\beta \in \mathcal{O}_K$ such that $\alpha\beta \in \mathbb{Z}$. That is, α divides some non-zero integer.*

Proof This is a disguised version of Proposition 9.3. Let $\gamma = 1/\alpha$. Then $\gamma \in \overline{\mathbb{Q}}$, so there is some $N \in \mathbb{N}_+$ such that $N\gamma$ is an algebraic integer. Let $\beta = N\gamma$ for any such N. Then $\beta = N/\alpha$ is in K, and it's an algebraic integer, so it's in $\mathcal{O}_K$; and we have $\alpha\beta = N$. □

CHAPTER 10

Ideals in number fields

10.1 Ideals

Let K be a number field. We're going to study *ideals* in the ring of integers of K. (The zero ideal is an ideal, but it's not very interesting, so henceforth, when we say "ideal" we always mean *nonzero* ideal.)

Definition 10.1 (Notation for ideals) For any commutative ring R and elements $x_1, \dots, x_k$ of R, write $\langle x_1, \dots, x_k \rangle_R$ for the set $\{r_1x_1 + \cdots + r_kx_k : r_1, \dots, r_k \in R\}$, which is an ideal of R (the ideal generated by the x_i). We omit the subscript R if it's obvious from context.

Notice that any $\alpha \in \mathcal{O}_K$ gives us an ideal – the principal ideal $\langle \alpha \rangle = \{\alpha x : x \in \mathcal{O}_K\}$. However, since integer rings aren't always PIDs, there can be more ideals which aren't of this form.

Example 10.2 Let $R = \mathbb{Z}[\sqrt{-5}]$, which is the ring of integers of $\mathbb{Q}(\sqrt{-5})$; and let I be the ideal $\langle 2, 1 - \sqrt{-5} \rangle$ of R. We claim this ideal is not principal.
Assume for contradiction that α is a generator. Then α must divide 2, so $N(\alpha) \mid N(2) = 4$; and also $N(\alpha) \mid N(1 - \sqrt{-5}) = 6$. So $N(\alpha)$ must be 1 or 2.
If $N(\alpha)$ were equal to 1, then I would be the unit ideal. But this is not possible, since every element of I has the form $x + y\sqrt{-5}$ with $x = y \bmod 2$ (exercise!), so $1 \notin I$.
Hence $N(\alpha)$ must be 2. But the equation $x^2 + 5y^2 = 2$ obviously has no solutions, so we have a contradiction. □

Exercise 10.3 Generalise the above! Show that if $d \in \mathbb{N}_+$ is square-free with $d \neq 3 \bmod 4$, $p \nmid d$ is a prime such that $\left(\frac{-d}{p}\right) = 1$, and t is a square root of $-d \bmod p$, then the ideal $\langle p, \sqrt{-d} - t \rangle$ is principal in $\mathbb{Z}[\sqrt{-d}]$ if and only if $x^2 + dy^2 = p$ has an integer solution. Can you formulate an analogue for $d = 1 \bmod 4$? What about $d < 0$?

Definition 10.4 (Product of ideals) Let I and J be ideals in $\mathcal{O}_K$. Then we define

$$IJ = \{i \cdot j : i \in I, j \in J\}.$$

You should check that ideal multiplication is compatible with element multiplication, i.e. $\langle\alpha\rangle\langle\beta\rangle = \langle\alpha\beta\rangle$. Moreover, $\langle\alpha\rangle = \langle\beta\rangle$ iff α and β are associates. So we get maps

$$(\dagger) \qquad (\mathcal{O}_K - \{0\}) \twoheadrightarrow \frac{(\mathcal{O}_K - \{0\})}{\{\text{units}\}} \hookrightarrow \{\text{nonzero ideals}\},$$

which are compatible with multiplication (and send the identity to the identity). If $\mathcal{O}_K$ is a PID (in particular if it's Euclidean), then the second map is a bijection.

The moral of the next few sections will be that there is **always** a notion of "unique prime factorisation" for *ideals*. When $\mathcal{O}_K$ is a PID, we get unique factorisation for *elements* from this using the bijectivity of the second map in (†). Conversely, when $\mathcal{O}_K$ is not a PID, we never have unique prime factorisation in $\mathcal{O}_K$; but the non-principal ideals are precisely the "extra stuff" we need to add to get unique factorisation back again.

10.2 Factoring ideals

Remember that an ideal I in any (commutative) ring A is said to be a *prime ideal* if $I \neq A$, and for all $x, y \in A$ we have $xy \in I \Rightarrow x \in I$ or $y \in I$. This obviously generalises the definition of prime *elements*: an element is prime iff the principal ideal it generates is a prime ideal.

Proposition 10.5 *Let I be a non-zero ideal in $\mathcal{O}_K$, for K a number field. Then I is prime if and only if it is* maximal*, i.e. $I \neq \mathcal{O}_K$ and there is no ideal J such that $I \supsetneq J \supsetneq \mathcal{O}_K$.*

Proof We know that I is prime iff $R = \mathcal{O}_K/I$ is an integral domain (this is just rewriting the definition).

We claim that

(a) this quotient R is *finite*,
(b) a finite integral domain is automatically a field.

To prove (a), we note that I is non-zero, so it contains a non-zero $\alpha \in \mathcal{O}_K$. Moreover, α must divide a non-zero integer C, by Proposition 9.14. Thus $C \in \mathcal{O}_K$; and $\mathcal{O}_K/C$ is finite, since $\mathcal{O}_K$ is finitely-generated and $C \neq 0$. Thus R is a quotient of a finite thing, so it's also finite.

To prove (b), suppose R is an integral domain and $0 \neq x \in R$. Then multiplication by x is a map $R \to R$ which is injective, by the integral-domain assumption. But an injection from a finite set to itself must be a bijection; so 1 is in the image and hence x is invertible.

To finish the proof, we note that for any commutative ring A and ideal I of A, the ideal I is maximal iff A/I is a field (exercise). So

$$(I \text{ prime}) \iff (R \text{ int. domain}) \iff (R \text{ field}) \iff (I \text{ maximal}). \qquad \square$$

Corollary 10.6 *Let $0 \neq \alpha \in \mathcal{O}_K$. Then:*

- *α is a prime element iff there is no ideal strictly containing $\langle\alpha\rangle$ except the unit ideal.*
- *α is indecomposable iff there is no* principal *ideal strictly containing $\langle\alpha\rangle$ except the unit ideal.*

Proof The first assertion is just the previous proposition applied to $\langle\alpha\rangle$. The second is obvious, since $\langle\beta\rangle \supset \langle\alpha\rangle$ iff $\beta \mid \alpha$. □

In particular, if $\mathcal{O}_K$ is a PID, then prime elements and indecomposable elements coincide (something you saw without proof in Algebra 1).

Theorem 10.7 (Dedekind) *Let I, J be ideals in $\mathcal{O}_K$ with $I \subseteq J$. Then there exists an ideal H such that $I = HJ$.*

This is surprisingly hard, and we're not going to prove it in this course. For a proof see Stewart & Tall.

Remark 10.8 This theorem would be false if we replaced $\mathcal{O}_K$ with a ring like $\mathbb{Z}[\sqrt{-3}]$, which isn't equal to the full ring of integers of its parent number field.

Corollary 10.9 *Multiplying by a non-zero ideal is injective: that is, if H, I, J are (nonzero!) ideals and $HI = HJ$, then $I = J$.*

Proof Firstly, we suppose H is principal, say $H = \langle x\rangle$. Then HI is exactly the set of elements $xi : i \in I$, and similarly HJ. Since multiplication by x is injective, it follows that $I = \{y : xy \in HI\} = \{y : xy \in HJ\} = J$.

For a general ideal H, we choose a non-zero element $x \in H$. Then $H \supseteq \langle x\rangle$, so $\langle x\rangle = H'H$ for some H'. So if $HI = HJ$ then $H'HI = H'HJ$, i.e. $\langle x\rangle I = \langle x\rangle J$, and the previous paragraph shows that $I = J$. □

Theorem 10.10 (Unique factorisation of ideals) *Any nonzero ideal is equal to a product of finitely many prime ideals, and its expression in this form is unique up to ordering.*

Proof For any I, there are finitely many ideals containing I, since they biject with the ideals of the finite quotient ring R/I. Hence we can find one which is maximal (not contained in any other ideal). Let P be such an ideal. Then P divides I, so $I = PJ$ for some J.

Clearly J can't be equal to I, since if $I = PI$ then $R = P$, a contradiction. So J is strictly larger than I. By induction on the size of $\mathcal{O}_K/I$, we may assume that J is a product of maximal ideals, hence so is I.

The proof of uniqueness proceeds exactly as before. □

Example 10.11 (important) We can now understand what was really going on in our $\mathbb{Z}[\sqrt{-5}]$ example. Remember that we had two different factorisations of 6 into indecomposable elements:

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

One checks that the ideals

$$\mathfrak{p} = \langle 2, 1+\sqrt{-5}\rangle,$$
$$\mathfrak{q}_1 = \langle 3, 1+\sqrt{-5}\rangle,$$
$$\mathfrak{q}_2 = \langle 3, 1-\sqrt{-5}\rangle$$

are all prime; but none of them can be principal, since that would contradict the indecomposability of 2 and 3 in $\mathbb{Z}[\sqrt{-5}]$.
Now, one can show (exercise!)

$$\mathfrak{p}^2 = \langle 2\rangle, \qquad \mathfrak{q}_1\mathfrak{q}_2 = \langle 3\rangle,$$
$$\mathfrak{p}\mathfrak{q}_1 = \langle 1+\sqrt{-5}\rangle, \qquad \mathfrak{p}\mathfrak{q}_2 = \langle 1-\sqrt{-5}\rangle.$$

So the (unique) factorisation of the *ideal* $\langle 6\rangle$ is

$$\langle 6\rangle = \mathfrak{p}^2\mathfrak{q}_1\mathfrak{q}_2,$$

and the rival factorisations of the *element* 6 into indecomposables correspond to the ways of grouping the factors into subsets whose product is principal:

$$\langle 6\rangle = (\mathfrak{p}^2)(\mathfrak{q}_1\mathfrak{q}_2) = (\mathfrak{p}\mathfrak{q}_1)(\mathfrak{p}\mathfrak{q}_2).$$

Exercise 10.12 Compute the factorisation of $\langle 21\rangle$ into prime ideals in $\mathbb{Z}[\sqrt{-5}]$. Hence show that there are exactly 3 distinct factorisations of 21 into indecomposable elements, up to units and re-ordering.

10.3 The class group

We're now going to cook up an algebraic object which *measures* how badly ideals can fail to be principal (and thus how badly unique factorisation fails for elements).

Definition 10.13 A *fractional ideal* of $\mathcal{O}_K$ is a subset of K of the form

$$\mathcal{O}_K x_1 + \cdots + \mathcal{O}_K x_n$$

for some $x_1, \ldots, x_n \in K$ (not all of which are zero).

Thus, a fractional ideal contained in $\mathcal{O}_K$ is just an ideal, but things like $\frac{1}{2}\mathcal{O}_K$ are also fractional ideals.

Note that one can multiply fractional ideals to get new fractional ideals; and it follows from Dedekind's theorem that every fractional ideal has an inverse. Along with some easy checks for associativity etc, this shows that fractional ideals form an abelian group.

Definition 10.14 The *class group* of K is the quotient

$$\mathrm{Cl}_K = \frac{\{\text{fractional ideals}\}}{\{\text{principal fractional ideals}\}}.$$

We'll now state one of the most important theorems in algebraic number theory:

Theorem 10.15 *For any number field K, the class group Cl_K is* finite.

This is one of the key theorems of algebraic number theory. We're not going to prove it in this course (it's quite hard)[1]. It says that although unique factorisation can fail – because there are non-principal ideals – it only "fails finitely badly".

Example 10.16 Going back to Example 10.11, the ideal $\mathfrak{p}$ is not principal (since $x^2+5y^2 = 2$ has no solutions) but $\mathfrak{p}^2$ is principal, so $[\mathfrak{p}]$ is a nontrivial element of Cl_K of order 2. Since $\mathfrak{p}\mathfrak{q}_1$ and $\mathfrak{p}\mathfrak{q}_2$ are principal, all three of the ideals $\{\mathfrak{p}, \mathfrak{q}_1, \mathfrak{q}_2\}$ all lie in this nontrivial ideal class.
It turns out that this is the only non-trivial element of the class group, so $\mathrm{Cl}_K \cong C_2$.

10.4 Cyclotomic fields, and Fermat's Last Theorem

Definition 10.17 The n-th *cyclotomic field* is the number field $\mathbb{Q}(\zeta_n)$, where $\zeta_n = \exp(2\pi i/n)$.

This is indeed a number field, because $(\zeta_n)^n = 1$, so ζ_n is algebraic. One can check that the ring of integers is equal to $\mathbb{Z}[\zeta_n]$.

Theorem 10.18 (Kummer) *Let p be an odd prime, and suppose that p does not divide the class number of the field $\mathbb{Q}(\zeta_p)$, where ζ_p is a nontrivial p-th root of 1. Then there are no solutions to Fermat's equation $x^n + y^n = z^n$ with n divisible by p.*

The idea of Kummer's proof was to write $y^n = x^n - z^n$ and factor this in $\mathbb{Z}[\zeta_p]$ as $(x - z)(x - \zeta_p z)\dots(x - \zeta_p^{p-1} z)$. For simplicity, suppose $xyz \neq 0 \bmod p$; then one can show that the factors on the right are pairwise coprime.

If $\mathbb{Z}[\zeta_p]$ were a PID, then – by considering prime factorisations – each of the terms must itself be a p-th power (up to units); and this eventually gives enough information to deduce that no such x, y, z exist. Kummer realised that one can push through the same argument as long as the class group has prime-to-p order (it doesn't have to be trivial).

Remark 10.19 Several earlier mathematicians had tried to make such an argument *assuming* that unique factorisation worked in $\mathbb{Z}[\zeta_p]$. Kummer *invented* the whole machine of ideal theory and class groups in order to sort out the mess!

[1] There is a simpler proof in the special case of fields $\mathbb{Q}(\sqrt{-D})$ with $D > 0$, which would be a nice project for a bachelor thesis or similar.

CHAPTER 11

P-adic numbers

An idea we've seen a few times already is that it's sometimes easiest to attack number-theoretic problems "one prime at a time". Pursuing this idea a little further leads to the idea of *p-adic numbers*, which is the last topic we'll look at in this module.

11.1 Review of metric spaces

For this section, we'll need a few ideas which you saw in the *Calculus II* module (also known as *Analysis II*):

- **Metric spaces**: a metric space is a set X with a notion of the "distance" $d(x, y)$ between points in X (which has to satisfy some axioms, e.g. the triangle inequality).
- **Cauchy sequences**: a Cauchy sequence in a metric space is a sequence whose terms are "eventually close together", i.e. for any $\epsilon > 0$ there is an N such that $d(x_m, x_n) < \epsilon$ for all $m, n \geqslant N$.
- **Completeness**: a metric space is *complete* if every Cauchy sequence in the space has a limit.

We'll be interested in the case where X is also a *ring*; and we'll want to consider metrics which "interact nicely" with the ring structure. This leads to the following construction:

Definition 11.1 If A is a commutative ring, then an *absolute value* on A is a map $|\cdot| : A \to \mathbb{R}$ satisfying the conditions

(a) $|x| \geqslant 0$ for all x, with equality if and only if $x = 0$;
(b) $|x \cdot y| = |x| \cdot |y|$;
(c) $|x + y| \leqslant |x| + |y|$.

If we have the stronger inequality

(c') $|x + y| \leqslant \max(|x|, |y|)$

then we say $|\cdot|$ is *non-archimedean*.

Clearly if $|\cdot|$ is an absolute value, then the formula $d(x, y) = |x - y|$ gives a metric (but not all metrics arise in this way). For instance, the standard metric on $\mathbb{R}$ is induced by the standard absolute value; this satisfies (c) but not (c') (it is "Archimedean").

Exercise 11.2 Show that if an absolute value on A exists, then A must be an integral domain.

11.2 The p-adic metric

Let p be a prime number.

Definition 11.3 We define the *p-adic absolute value* on $\mathbb{Q}$ by $|0|_p = 0$, and for $x = \frac{r}{s}$ with for integers $r, s \neq 0$, then $|x|_p = p^{\text{ord}_p(s) - \text{ord}_p(r)}$.

Here $\text{ord}_p\, n$ is the highest power of p dividing n, as before. Thus $|p^k|_p = p^{-k}$ for all k; note the sign! So we've entered a strange mirror-world where raising p to a *large* power gives something *small*.

Exercise 11.4 Show that $|\cdot|_p$ is a nonarchimedean absolute value.

The p-adic metric has some strange properties. For example, if x, y, z are any three rationals, then at least two of the lengths $|x-y|_p, |y-z|_p, |x-z|_p$ are equal to each other. That is, in the p-adic world, every triangle is an isosceles triangle! (Exercise: Prove this.)

Remark 11.5 There is a notion of "equivalence" of absolute values: the absolute values $|\cdot|$ and $|\cdot|'$ are equivalent if there is a real number $e > 0$ such that $|x|' = |x|^e$ for all e. Equivalent absolute values make the same sequences Cauchy, and induce the same topology.
Ostrowski's theorem shows that any absolute value on $\mathbb{Q}$ is equivalent to precisely one of the following:

- the standard absolute value inherited from $\mathbb{R}$,
- the p-adic absolute value for some prime p,
- the trivial absolute value with $|x| = 1$ for all $x \neq 0$.

11.3 Building the completion

It turns out that $\mathbb{Q}$ is *not* complete in the p-adic metric (more generally, a countable metric space with no isolated points cannot be complete). See Gouvea's book for an explicit construction of a Cauchy sequence which doesn't converge (using Proposition 6.2).

We'll now show that there is a canonical way of "completing" it: embedding $\mathbb{Q}$ into a larger field in which any Cauchy sequence for the p-adic metric has a unique limit, just like any Cauchy sequence in $\mathbb{Q}$ for the *usual* metric has a limit in $\mathbb{R}$. (We'll skip several proofs here; you can look them up in Gouvea's book if you want to see the proofs, but you don't need to know them for the exam.)

Definition 11.6 Let $\mathcal{C}$ denote the set of sequences $(x_n)_{n\in\mathbb{N}}$ of rational numbers which are Cauchy sequences for the p-adic metric.

Notice that we can embed $\mathbb{Q}$ into $\mathcal{C}$ via $x \mapsto (x, x, x, \dots)$. We can also make $\mathcal{R}$ into a *ring* with the obvious termwise ring operations, so $(x_0, x_1, \dots) + (y_0, y_1, \dots) = (x_0 + y_0, x_1 + y_1, \dots)$ and similarly for multiplication.

Exercise 11.7 Check that the sum and product of Cauchy sequences is Cauchy, so this is well-defined.

Definition 11.8 We write $\mathcal{N} \subset \mathcal{C}$ for the set of Cauchy sequences tending to 0.

Proposition 11.9 *The set $\mathcal{N}$ is an ideal of $\mathcal{C}$.*

Proof It is easy to check that the sum of two sequences tending to 0 tends to 0, so it is a subgroup under addition. To show it is closed under multiplication, let $(x_n) \in \mathcal{N}$ and let (y_n) be any Cauchy sequence; then we can find a B such that $|y_n|_p \leqslant B$ for all sufficiently large n, so $|x_n y_n|_p \leqslant B|x_n|_p$, and for large enough n this can be made arbitrarily small. □

Definition 11.10 We define $\mathbb{Q}_p$ as the quotient ring $\mathcal{C}/\mathcal{N}$.

Proposition 11.11 *$\mathbb{Q}_p$ is a field.*

Proof What we need to show is the following: if $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence which does *not* tend to 0, then we can find another Cauchy sequence $(y_n)_{n\in\mathbb{N}}$ such that $x_n y_n - 1 \in \mathcal{N}$.

One can show that the sequence (x_n) is *eventually bounded away from 0*; that is, we can find a $c > 0$ and $N \in \mathbb{N}$ such that $|x_n|_p \geqslant c$ for all $n \geqslant N$. (This is a nice exercise, using the fact that (x_n) is a Cauchy sequence and doesn't tend to 0.) In particular, $x_n \neq 0$ for all $n \geqslant N$.

Let's define the sequence y_n by $y_n = 0$ if $x_n = 0$, and $y_n = 1/x_n$ otherwise. Then y_n is Cauchy, since for any $m, n \geqslant N$ we have

$$|y_m - y_n|_p = \frac{|x_m - x_n|_p}{|x_m x_n|_p} \leqslant \tfrac{1}{c^2}|x_m - x_n|_p,$$

so the Cauchy property for (y_n) follows from that for (x_n). So $(y_n) \in \mathcal{C}$. Moreover, $x_n y_n - 1$ is a sequence which eventually consists entirely of zeros, so it's certainly tending to 0. □

The composite $\mathbb{Q} \to \mathcal{C} \to \mathcal{C}/\mathcal{N}$ is injective, since $(x, x, x, \dots) \notin \mathcal{N}$ if $x \neq 0$. So we can identify $\mathbb{Q}$ with a subfield of $\mathbb{Q}_p$. It turns out that the p-adic absolute value extends to $\mathbb{Q}_p$:

Proposition 11.12 *For any $(x_n) \in \mathcal{C}$, the limit $|x|_p = \lim_{n\to\infty} |x_n|_p$ exists, and it depends only on the image of (x_n) in the quotient $\mathbb{Q}_p$. This defines a nonarchimedean absolute value $|\cdot|_p$ on $\mathbb{Q}_p$ whose restriction to $\mathbb{Q}$ is the p-adic absolute value defined above. Moreover, for any $x \in \mathbb{Q}_p$ we have*

$$|x|_p \in \{0\} \cup \{p^{-k} : k \in \mathbb{Z}\}.$$

(Note that since the new absolute value on $\mathbb{Q}_p$ agrees with the one we already have on $\mathbb{Q}$, there is no harm in denoting both by the same symbol.) The last theorem of this section shows that we have achieved our goal of "p-adically completing" $\mathbb{Q}$:

Theorem 11.13 *The field $\mathbb{Q}_p$ is complete for the metric induced by this absolute value; and $\mathbb{Q}$ is a dense subset of $\mathbb{Q}_p$.*

Proof Omitted. □

11.4 The p-adic integers $\mathbb{Z}_p$

Definition 11.14 We define $\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leqslant 1\}$.

The non-archimedean property implies that this is a *subring* of $\mathbb{Q}_p$, not just a subset. Moreover, it is both open and closed in the p-adic topology. We're going to show that $\mathbb{Z}_p$ is also the *closure* of $\mathbb{Z}$ in $\mathbb{Q}_p$; this follows from the following more precise statement:

Proposition 11.15 *Given any $x \in \mathbb{Z}_p$ and $n \geqslant 1$, there exists $\alpha \in \mathbb{Z}$ such that $|x - \alpha| \leqslant p^{-n}$; and the set of α with this property is a congruence class modulo p^n.*

Proof If α_0 satisfies these conditions, and α is any integer, then the nonarchimedean property implies that

$$|x - \alpha| \leqslant p^{-n} \iff |\alpha - \alpha_0| \leqslant p^{-n} \iff \alpha = \alpha_0 \bmod p^{-n}.$$

So the set of $\alpha \in \mathbb{Z}$ such that $|x - \alpha| \leqslant p^{-n}$ is either empty, or a congruence class mod p^n.

It remains to show that some α_0 with this property exists. The density of $\mathbb{Q}$ in $\mathbb{Q}_p$ shows there is a rational $\frac{a}{b}$ with $|x - \frac{a}{b}|_p \leqslant p^{-n}$. Since $|x|_p \leqslant 1$, we deduce that $|\frac{a}{b}|_p \leqslant 1$ also; so $\operatorname{ord}_p(a) \geqslant \operatorname{ord}_p(b)$, and after removing any common factors, we can suppose $p \nmid b$. Hence we can find $b' \in \mathbb{Z}$ with $bb' = 1 \bmod p^n$.

We claim $\alpha_0 = ab' \in \mathbb{Z}$ works. This follows since $|\frac{a}{b} - ab'|_p = |\frac{a - abb'}{b}|_p = |a(1 - bb')|_p$ (as $|b| = 1$), which is $\leqslant p^{-n}$, since $bb' = 1 \bmod p^n$. □

This construction defines a map $\mathbb{Z}_p \to \mathbb{Z}/p^n\mathbb{Z}$, $x \mapsto \alpha$, which is clearly a ring homomorphism, and extends the natural quotient map $\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$.

The kernel of this map is $\{x \in \mathbb{Z}_p : |x| \leqslant p^{-n}\}$, and since $|x| \leqslant p^{-n} \iff |x/p^n| \leqslant 1$, the kernel is precisely the principal ideal $p^n\mathbb{Z}_p$. That is, we've shown:

Proposition 11.16 *The inclusion $\mathbb{Z} \hookrightarrow \mathbb{Z}_p$ induces isomorphisms*

$$\mathbb{Z}/p^n\mathbb{Z} \cong \mathbb{Z}_p/p^n\mathbb{Z}_p$$

for all $n \geqslant 1$. □

(In particular, we may write "$x \bmod p^n$" for any $x \in \mathbb{Z}_p$ to mean its image in $\mathbb{Z}/p^n\mathbb{Z}$.)

Proposition 11.17 *The only ideals in $\mathbb{Z}_p$ are the zero ideal and the ideals $p^n\mathbb{Z}_p$, for $n \geqslant 1$. In particular, the only non-zero **prime** ideal is $p\mathbb{Z}_p$.*

Proof For the first claim, let J be an ideal in $\mathbb{Z}_p$, and consider the set $\{k \in \mathbb{N} : \exists x \in J \text{ such that } |x| = p^{-k}\}$. If S is empty, then $J = \{0\}$ and we're done. If S is non-empty, it has a least element, say n, and there is some x with $|x| = p^{-n}$. But then $u = p^n/x$ satisfies $|u| = 1$, so $u \in \mathbb{Z}_p$; thus $ux = p^n \in J$. Now, for any $y \in J$, we have $|y/p^n| \leqslant 1$, so $y \in p^n\mathbb{Z}_p$ and we have shown that $J = p^n\mathbb{Z}_p$.

For the second claim, we know that $p^k\mathbb{Z}_p$ isn't prime for $k > 1$, since it doesn't contain p or p^{k-1}, but does contain their product. □

Combining the last two propositions shows that going from $\mathbb{Z}$ to $\mathbb{Z}_p$ "removes" all primes except p, without changing anything modulo powers of p.

Proposition 11.18 *Suppose we have a sequence of elements*

$$(x_1, x_2, x_3, \dots) \in (\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/p^2\mathbb{Z}) \times (\mathbb{Z}/p^3\mathbb{Z}) \times \dots$$

satisfying $x_{i+1} \bmod p^i = x_i$ for all i. Then there is a unique $x \in \mathbb{Z}_p$ with $x \bmod p^i = x_i$ for all i.

Proof It's clear that the map sending $x \in \mathbb{Z}_p$ to $(x \bmod p, x \bmod p^2, \dots)$ is well-defined, and it must be injective, since if $x = x' \bmod p^i$ then $|x - x'| \leqslant p^{-i}$, and if this holds for all i then $|x - x'| = 0$, implying $x = x'$.

To show surjectivity, let $\tilde{x}_i$ be any choice of element in $\mathbb{Z}_p$ reducing to $x_i \bmod p^i$. Then $(\tilde{x}_i)$ is a Cauchy sequence, since $|\tilde{x}_i - \tilde{x}_j| \leqslant p^{-N}$ for all $i, j \geqslant N$. So it has a limit x; and letting $j \to \infty$ in the last formula we have $|\tilde{x}_i - x| \leqslant p^{-N}$ for all $i \geqslant N$, hence $|x - \tilde{x}_i| \leqslant p^{-i}$, i.e. $x \bmod p^i = x_i$. □

This gives a bijection from $\mathbb{Z}_p$ to the set of compatible sequences in $\prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$; and this is clearly a ring homomorphism, if we define addition and multiplication of sequences term-by-term.

Remark 11.19 This can be used to give an alternative, purely algebraic definition of $\mathbb{Z}_p$, although constructing $\mathbb{Q}_p$ this way is more difficult.

11.5 P-adic numbers as "power series"

Working with compatible sequences is a nice way of proving theorems about $\mathbb{Z}_p$, but it's a little bit cumbersome if you want to actually compute. The basic problem is that if you know x_n, then you can recover $x_1, x_2, \dots, x_{n-1}$ from it; so each new term in the sequence repeats a lot of information you already knew. The following proposition gives a more "concrete" way of thinking about, and calculating in, the ring $\mathbb{Z}_p$.

Proposition 11.20 *Let $x \in \mathbb{Z}_p$. Then there are uniquely determined integers $(a_0, a_1, \dots)$, with $a_i \in \{0, 1, 2, \dots, p-1\}$ for each i, such that the sum $\sum_{i=0}^{\infty} a_i p^i$*

converges to x in the topology of $\mathbb{Z}_p$. This construction gives a bijection between $\mathbb{Z}_p$ and $\{0, \dots, p-1\}^{\mathbb{N}}$.

Proof If $a = (a_0, a_1, \dots)$ is any sequence in $\{0, \dots, p-1\}^{\mathbb{N}}$, then we write $S_i(a)$, for $i \geqslant 1$, for the partial sum $\sum_{r=0}^{i-1} a_r p^r$.

If $j \geqslant i \geqslant 1$, then the extra terms in $S_j(a)$ that aren't in $S_i(a)$ are all divisible by p^i, so $S_j(a) = S_i(a) \bmod p^i$. It follows that the sequence $(S_i(a))_{i \geqslant 1}$ is Cauchy, and hence have a limit in $\mathbb{Z}_p$.

On the other hand, if two such sequences a, b have the same limit x, then $S_n(a) = S_n(b) = x \bmod p^n$ for all n; but $S_n(a)$ and $S_n(b)$ are integers in $\{0, \dots, p^n - 1\}$, so being equal mod p^n means they are equal as integers. Since an integer has a unique base p expansion, we have $a_i = a_j$ for $0 \leqslant j < n$, and as this holds for all n, the two sequences are identical.

It remains to show that every $x \in \mathbb{Z}_p$ is a limit of such a sequence. Clearly we can choose $a_0 \in \{0, \dots, p-1\}$ such that $x \bmod p = a_0$. Then $x \bmod p^2$ differs from a_0 by a multiple of p, so we can find $a_1 \in \{0, \dots, p-1\}$ such that $x \bmod p^2 = a_0 + pa_1$. Proceeding inductively we can construct a sequence $a = (a_0, a_1, \dots) \in \{0, \dots, p-1\}^{\mathbb{N}}$ such that $S_n(a) = x \bmod p^n$ for all n, so the partial sums of a tend to x. □

Corollary 11.21 *$\mathbb{Z}_p$ and $\mathbb{Q}_p$ are uncountable.*

Proof Once we know that $\mathbb{Z}_p$ bijects with (nontrivial finite set)$^{\mathbb{N}}$, the proof proceeds in the same way as Cantor's diagonal argument for the uncountability of the real numbers. □

The a_i are sometimes called the *p-adic digits* of x. To add two p-adic integers in this form, we add the terms in the sum, starting with the degree 0 term, and "carrying" powers of p upwards, just like adding usual integers written in base 10. For example if $p = 5$, and we want to compute

$$(3 + 2 \times 5 + 3 \times 5^2 + \dots) + (2 + 1 \times 5 + 2 \times 5^2 + \dots),$$

then the degree 0 terms sum to $5 = 0 + 1 \times 5$; so we write down 0, and carry the 1 to the next term to get $(2 + 1 + 1) \times 5 = 4 \times 5$, hence the sum is $0 + 4 \times 5 + \dots$.

Remark 11.22 This process will never stop, unless x and y are actually in $\mathbb{N}$; we can't compute "all" the a_i's for most p-adic numbers, any more than we can compute "all" of the decimal digits of π. But we can compute the first n digits of $x + y$ for any given n if we know the corresponding digits of x and y.

We can extend this to $\mathbb{Q}_p$ if we allow finitely many terms with negative powers of p, i.e. $x = \sum_{i=-N}^{\infty} a_i p^i$ for some N. (We can't allow an infinite negative "tail", though, since the sum wouldn't converge.)

CHAPTER 12

Equations in $\mathbb{Z}_p$ and Hensel's lemma

As in the previous chapter, p is a prime. For $x \in \mathbb{Z}_p$, we'll write $\bar{x}$ for its reduction mod p, so $\bar{x} \in \mathbb{Z}/p\mathbb{Z}$.

12.1 Roots of polynomials

Let's consider a *monic* polynomial $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$, where the a_i are in $\mathbb{Z}_p$ (or just in $\mathbb{Z}$, if you prefer). What can we say about its roots in $\mathbb{Z}_p$?

Obviously, if $\alpha \in \mathbb{Z}_p$ is a root, then its reduction $\bar{\alpha} \in \mathbb{Z}/p\mathbb{Z}$ is a root of the mod p polynomial $\bar{f} = \sum \bar{a}_i X^i$. Amazingly, this is (in a sense) all the information we need to understand solutions in $\mathbb{Z}_p$ as well!

We need the following abstract algebraic warm-up:

Proposition 12.1 *Let $\mathbb{K}$ be a field, and let $f = \sum a_i X^i$ a monic polynomial with coefficients in $\mathbb{K}$. Suppose $r \in \mathbb{K}$ satisfies $f(r) = 0$. Then the following are equivalent:*

- *r is a **simple** root of f, i.e. we can write $f(X) = (X - r)g(x)$ for some g with $g(r) \neq 0$;*
- *$f'(r) \neq 0$, where f' is defined purely formally as $\sum i a_i X^{i-1}$.*

Proof We can always write $f(X) = (X - r)g(X)$ for some g (the question is whether $g(r) = 0$ or not). But the product rule for derivatives holds for polynomials over any field, so $f'(X) = g(X) + (X - r)g'(X)$, and setting $X = r$ gives $f'(r) = g(r)$. □

Theorem 12.2 (Hensel's Lemma) *Let $f \in \mathbb{Z}_p[X]$ be a monic polynomial, and let $r \in \mathbb{Z}/p\mathbb{Z}$ be a simple root of $\bar{f}$. Then there exists a unique $\alpha \in \mathbb{Z}_p$ such that $f(\alpha) = 0$ and $\bar{\alpha} = r$.*

Proof We claim that for each $n \geqslant 1$, there exists a unique $\alpha_n \in \mathbb{Z}/p^n\mathbb{Z}$ such that $f(\alpha_n) = 0$ and $\alpha_1 = r$. This clearly suffices to prove the theorem, since the uniqueness implies that $(\alpha_n)_{n \geqslant 1}$ is a compatible sequence defining an element of $\mathbb{Z}_p$.

To prove the claim, we induct on n. The claim is obvious for $n = 1$; so let us suppose α_n exists, for some $n \geqslant 1$, and use it to construct α_{n+1}. Clearly, if it exists, it must be one of the p elements of $\mathbb{Z}/p^{n+1}$ which reduce to α_n mod p^n (otherwise this would contradict the uniqueness of α_n).

Let β be an arbitrary lifting of α_n to $\mathbb{Z}/p^{n+1}$; then all the other liftings look like $\beta + p^n\epsilon$, where ϵ varies over $\{0, \dots, p-1\}$. Moreover, $f(\beta)$ is in $\mathbb{Z}/p^{n+1}$ and is zero mod p^n, so it can be written as $p^n\mu$ for some $\mu \in \{0, \dots, p-1\}$.

For each i, the binomial theorem gives

$$a_i(\beta + p^n\epsilon)^i = a_i(\beta^i + i\beta^{i-1}p^n\epsilon + \dots)$$

where the $(\dots)$ denote terms which are divisible by p^{2n}, hence are zero mod p^{n+1}. Thus

$$f(\beta + p^n\epsilon) = f(\beta) + p^n\epsilon f'(\beta) = p^n(\mu + \epsilon f'(\beta)).$$

However, since we are working mod p^{n+1}, the expression in the brackets only matters modulo p. As $\beta = r \bmod p$, we have $f'(\beta) \bmod p = \bar{f}'(r) \neq 0$. Thus there is a unique choice of ϵ which makes the bracket zero mod p. □

Example 12.3 Suppose $p \neq 2$, and $a \in \mathbb{Z}_p$ is such that a mod p is a non-zero quadratic residue. Then $f(X) = X^2 - a$ has a root modulo p, and this root r must satisfy $\bar{f}'(r) = 2r \neq 0$; so Hensel's lemma says it has a root in $\mathbb{Z}_p$. Thus *a unit in $\mathbb{Z}_p$ is a square if and only if its image in $\mathbb{Z}/p\mathbb{Z}$ is a square*, and similarly for n-th powers as long as $p \nmid n$. (This generalises Proposition 6.2).

12.2 Explicitly constructing solutions

The proof of Hensel's lemma is constructive – it gives us a recipe for constructing the solution modulo higher and higher powers of p. This can be made even more explicit, as follows.

Proposition 12.4 (Newton's iteration) *In the situation of Hensel's lemma, choose some $x_1 \in \mathbb{Q}$ whose denominator is coprime to p and such that $x_1 = r \bmod p$. Consider the sequence defined for $n \geqslant 1$ by*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Then we have $x_m = x_n \bmod p^n$ for all $m \geqslant n \geqslant 1$, and x_n is a root of $f \bmod p^n$.

Proof This is a rephrasing of the proof of Hensel's lemma above. □

Example 12.5 For example, take $f(X) = X^2 - 11$, and start with $x_1 = 1$, which is a root of f modulo 5. Then we get a sequence of rational numbers (which are quite complicated, e.g. $x_7 = 51909324631296565268391993035 53/156512502657058511473462499 3088$); and these are tending in the 5-adic metric to a root of f in $\mathbb{Z}_5$.
(Amazingly, they're also tending in $\mathbb{R}$ to a root in $\mathbb{R}$! So the same rational-number sequence is calculating the square root of 11 in two different completions of $\mathbb{Q}$ at once. However, it diverges horribly in the p-adic topology for $p \neq 5$.)

Remark 12.6 The convergence is actually much better than the theorem claims: rather than just getting one more correct 5-adic digit of $\sqrt{11}$ with each step, we

actually *double* the number of correct digits (on average). But this takes a little more work to show.

There are various generalisations of Hensel's lemma; for instance, we can deal with non-simple roots, as long as we start with a root modulo p^k, for some large enough k. We can also consider systems of polynomials in several variables, assuming that a solution exists mod p and a suitable matrix of partial derivatives mod p is non-singular.

12.3 P-adic logarithms and the structure of $\mathbb{Z}_p^\times$

In this section we'll suppose $p \neq 2$, for simplicity.

Recall that in real analysis the logarithm function has a Taylor series expansion around 1,

$$\log x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(x-1)^n}{n},$$

convergent for $|x-1| < 1$. Amazingly the same thing works in the p-adics:

Proposition 12.7 *For all $x \in \mathbb{Q}_p$ with $|x-1|_p < 1$, the sum $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}(x-1)^n}{n}$ converges in $\mathbb{Q}_p$.*

Proof For any $n \geqslant 1$ we have $|\frac{1}{n}|_p \leqslant n$ (exercise). So if $|x-1| = r < 1$, then

$$|\frac{(-1)^{n-1}(x-1)^n}{n}|_p \leqslant nr^n,$$

which tends (rather rapidly) to 0. The nonarchimedean property (and completeness) of $\mathbb{Q}_p$ implies that any sum whose terms tend to 0 is convergent[1]. □

This defines a function $\log_p : U \to \mathbb{Q}_p$, where U denotes the subgroup $\{x : |x-1| < 1\}$ of $\mathbb{Q}_p^\times$. One can check that

$$\log_p(xy) = \log_p x + \log_p y \quad \forall x, y \in U,$$

and moreover, $\log_p$ is a bijection from U to $p\mathbb{Z}_p$ (which is clearly isomorphic to $\mathbb{Z}_p$ as an additive group). So we have shown:

Proposition 12.8 *There is an isomorphism of abelian groups $L : (U, \times) \cong (\mathbb{Z}_p, +)$.* □

Corollary 12.9 *The group U does not contain any nontrivial root of unity (i.e. any u with $u \neq 1$, but $u^k = 1$ for some $k > 1$).*

Proof Suppose $u \in U$ has $u \neq 1$ but $u^k = 1$; and let $t = L(u) \in \mathbb{Z}_p$. Since L converts multiplication to addition, we have $kt = L(u^k) = L(1) = 0$; but since $u \neq 1$, we have $t \neq 0$, and also $k \neq 0$, since $k > 1$ and $\mathbb{Z}$ injects into $\mathbb{Z}_p$. So this contradicts the fact that $\mathbb{Z}_p$ is an integral domain. □

[1] One of the many ways that p-adic analysis is easier than real analysis!

Theorem 12.10 *The group of roots of unity in $\mathbb{Z}_p$ is finite and cyclic of order $p-1$; and for each $a \in (\mathbb{Z}/p\mathbb{Z})^\times$, there is a unique $(p-1)$-st root of unity in $\mathbb{Z}_p$ mapping to a mod p (the "Teichmüller lift" of a).*

Proof Firstly, the polynomial $X^{(p-1)} - 1$ has $p-1$ distinct roots mod p – every element of $(\mathbb{Z}/p\mathbb{Z})^\times$ is a root, and clearly these must be simple (since the degree is $p-1$). So Hensel's lemma says that there is a unique root in $\mathbb{Z}_p$ lifting each of these.

On the other hand, suppose N is any integer $\geqslant 1$, and $\zeta \in \mathbb{Z}_p^\times$ is a root of unity of order N. Then there is some $(p-1)$-st root of unity ω such that $\zeta = \omega \bmod p$. Thus $\zeta/\omega = 1 \bmod p$, and ζ/ω is a root of unity lying in U. From the last corollary, $\zeta/\omega = 1$, so ζ is in fact a $(p-1)$-st root of unity. □

Definition 12.11 The map $\tau : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{Z}_p^\times$, sending a to the unique $(p-1)$-st root of unity in $\mathbb{Z}_p$ that reduces mod p to a, is actually a group homomorphism; it is called the *Teichmüller character*.

Theorem 12.12 *There is an isomorphism*

$$\mathbb{Z}_p^\times \cong \mathbb{Z}_p \times C_{p-1},$$

sending x to $\left(L(\frac{x}{\tau(x)}), \tau(x)\right)$.

Proof We already know that $\mathbb{Z}_p^\times$ has a subgroup (namely U) isomorphic to $\mathbb{Z}_p$, such that the quotient (namely $(\mathbb{Z}/p\mathbb{Z})^\times$) is cyclic of order $p-1$. So to show the above isomorphism, it suffices to find a subgroup of $\mathbb{Z}_p^\times$ mapping isomorphically to $(\mathbb{Z}/p\mathbb{Z})^\times$, and the Teichmüller lifting construction does exactly this. □

12.4 Local-to-global principles

Now let's suppose we're trying to solve a polynomial equation in $\mathbb{Q}$ (or a system of many equations in many variables). Clearly, if it has a solution in $\mathbb{Q}$, then it has a solution in $\mathbb{Q}_p$ for *every* p, and also in $\mathbb{R}$. Often this gives us a cheap way of *ruling out* solutions in $\mathbb{Q}$: for instance, $X^2 - 37Y^2 = 0$ has no solutions in $\mathbb{Q}_5$ except $X = Y = 0$ (exercise!) so it has no non-trivial rational solutions either.

It would be nice if this were the *only* obstruction to existince of rational solutions. Unfortunately, this doesn't always work. For instance, Gouvea gives the following exercise:

Exercise 12.13 Show that the equation $(X^2-2)(X^2-17)(X^2-34) = 0$ has roots in $\mathbb{Q}_p$ for every p, and in $\mathbb{R}$, but no roots in $\mathbb{Q}$.

However, in some cases – for nice classes of equations – we *can* deduce solvability in $\mathbb{Q}$ from solvability in the completions.

Theorem 12.14 (Hasse–Minkowski) *Let*

$$F(X_1, \dots, X_n) = \sum_{i,j} c_{i,j} X_i X_j$$

be a homogenous quadratic polynomial in n variables with rational coefficients. Then there exist non-trivial solutions of $F(X_1, \dots, X_n) = 0$ in $\mathbb{Q}$ if and only if there exist non-trivial solutions in $\mathbb{R}$, and in $\mathbb{Q}_p$ for every p.

(For a proof see Serre's book *A Course in Arithmetic*.)

This is an example of a *local-to-global principle*, showing that (under suitable hypotheses) we can recover information in $\mathbb{Q}$ (a "global" field) from information about its completions ("local" fields). Investigating when local-global principles hold – and if not, whether one can quantify "how badly" they fail – is a hugely important theme in number theory research today.