ALGEBRAIC NUMBER THEORY

BEN GREEN

Contents

Preface

A brief introduction. Every positive integer greater than one may be factored into primes, and this factorisation is unique up to the ordering of the primes. You have known this fact since school (though the first time you saw a *proof* may well have been last year, in Part A). It is impossible to imagine doing number theory without it.

Does unique factorisation into primes generalise? To understand why one might care about this question, let us look at some theorems about *diophantine equations* (equations to be solved in integers) that have been proven by mathematicians in the past.

- considered by Fermat and Euler: the only solutions to $y^2 + 2 = x^3$ are $x = 3$, $y = \pm 5$.
- Fermat: if p is a prime, $p = x^2 + y^2$ has a solution if and only if $p \equiv 1 (\mathrm{mod}\, 4)$.
- Euler: if $x^3 + y^3 = z^3$, one of x, y, z is zero (the case $n = 3$ of Fermat's last theorem).
- Nagell (conjecture of Ramanujan): if $x^2 + 7 = 2^n$, then $n = 3, 4, 5, 7, 15$.

A common feature of these equations is that they admit natural factorisations, but not over the integers. Respectively, they may be factored as

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3,$$

$$p = (x + iy)(x - iy),$$

$$(x + y)(x + \zeta y)(x + \zeta^2 y) = z^3$$

(where $\zeta = e^{2\pi i/3}$) and

$$(x + \sqrt{-7})(x - \sqrt{-7}) = 2^n.$$

To proceed further, one needs to understand the more general "number systems" in which we have written these factorisations. This – especially the question of unique factorisation into primes – is the main theme of the course.

Synopsis. The official synopsis of the course is as follows.

Field extensions, minimum polynomial, algebraic numbers, conjugates, discriminants, Gaussian integers, algebraic integers, integral basis

Examples: quadratic fields

Norm of an algebraic number

Existence of factorisation

Factorisation in $\mathbf{Q}(\sqrt{d})$

Ideals, $\mathbf{Z}$–basis, maximal ideals, prime ideals

Unique factorisation theorem of ideals

Relationship between factorisation of number and of ideals

Norm of an ideal

Ideal classes
Statement of Minkowski convex body theorem
Finiteness of class number
Computations of class number

These notes. These notes are expanded from previous ones by Victor Flynn, building on earlier notes of Neil Dummigan, Alan Lauder and Roger Heath-Brown. In particular, most of the illustrative examples are lifted directly from those notes. I am grateful to students Yutong Dai and Keyang Li for drawing my attention to a number of typos in earlier versions of the notes. Please send any further comments and corrections to

`ben.green@maths.ox.ac.uk`.

Prerequisites. These notes are relatively self-contained. We repeat a certain amount of material from *Rings and Modules*, sometimes with proof, but sometimes not. I would regard *Rings and Modules* as an essential prerequisite. *Galois Theory*, whilst listed as an essential prerequisite, is not quite so vital and a student not having taken that course ought to be able to follow the course, even though a couple of nonexaminable proofs do use the language of Galois theory.

I would expect all students attending this course would have been to *Part A Number Theory*. If you haven't, I advise reading the notes (for example, my notes from 2019, available on my webpage), especially

- The language of modular arithmetic;
- The statement (but not the proof) of quadratic reciprocity.

On examinable material. I have starred some subsections, as well as the final section of the notes. This means they are non-examinable according to the synopsis (in my interpretation) and, if time is short, I may not even cover them in lectures.

The appendices are definitely not examinable.

Material that is principally in other courses (*Rings and Modules, Galois Theory*) will not be examined.

A couple of calculations which we need, but which essentially have nothing to do with algebraic number theory, are outscoured to "Sheet X". This is entirely non-examinable.

In practice, past exams have focussed for the most part on *techniques* (computing integral bases, computing class numbers, solving equations, factoring into ideals) and it seems very unlikely that will change.

1. Algebraic numbers

Algebraic numbers. Minimal polynomials. We start with some very basic definitions.

Definition 1.1. A complex number α is *algebraic* if it is the solution to some polynomial equation with coefficients in $\mathbf{Q}$. The set of all algebraic numbers is denoted by $\overline{\mathbf{Q}}$.

Examples. Every rational is algebraic, as are i, $\sqrt{2}$, $3^{1/5}\dots$ but *not* e, π (though we shall not prove this here!). $\overline{\mathbf{Q}}$ is countable, since one may enumerate the polynomials over $\mathbf{Q}$, and each has only finitely many roots.

Lemma 1.2 (Minimal polynomial)**.** *Suppose that $\alpha \in \overline{\mathbf{Q}}$. Then there is a unique nonzero monic irreducible polynomial $m_\alpha(X)$ satisfied by α, which we call the* minimal polynomial *of α. If $f \in \mathbf{Q}[X]$ is any other polynomial satisfied by α then $m_\alpha \mid f$.*

Proof. Take m_α to be a monic nonzero polynomial of least degree satisfied by α. If m_α were reducible, say $m_\alpha(X) = f(X)g(X)$ with $\deg f, \deg g < \deg m_\alpha$, then since $m_\alpha(\alpha) = 0$ we have $f(\alpha)g(\alpha) = 0$, whence either $f(\alpha) = 0$ or $g(\alpha) = 0$, contrary to the minimality of $\deg m_\alpha$.

Now let $f \in \mathbf{Q}[X]$ be some polynomial satisfied by α. By the Euclidean algorithm we may write $f(X) = m_\alpha(X)q(X) + r(X)$ with $\deg r < \deg m_\alpha$. If $f(\alpha) = 0$ then, since $m_\alpha(\alpha) = 0$, we have $r(\alpha) = 0$. By the minimality of $\deg(m_\alpha)$, we must have $r = 0$, and so $m_\alpha | f$.

The uniqueness of m_α follows immediately, since the only monic irreducible f for which $m_\alpha \mid f$ is m_α itself. $\square$

Examples. The minimal polynomial $m_i(X)$ is $X^2 + 1$. The minimal polynomial $m_{\sqrt{2}}(X)$ is $X^2 - 2$. If $\omega = e^{2\pi i/3}$ is a primitive third root of unity then $m_\omega(X)$ is *not* $X^3 - 1$, since this is a reducible polynomial; rather, $m_\omega(X) = X^2 + X + 1$.

Given any complex number α, write $\mathbf{Q}(\alpha)$ for the smallest field containing $\mathbf{Q}$ and α; this will consist of all fractions $p(\alpha)/q(\alpha)$, where $p, q \in \mathbf{Q}[X]$ are polynomials. Recall that if K, L are two fields with $K \supseteq L$ then the *degree* $[K : L]$ is the degree of K, considered as a vector space over L (it may be infinite).

Lemma 1.3. *Let $\alpha \in \mathbf{C}$. Then α is algebraic if, and only if, $[\mathbf{Q}(\alpha) : \mathbf{Q}] < \infty$. Suppose that α* is *algebraic. Then $\mathbf{Q}(\alpha) = \mathbf{Q}[\alpha]$. Suppose that m_α, the minimal polynomial for α, has degree n. Then a basis for $\mathbf{Q}(\alpha)$ as a vector space over $\mathbf{Q}$ is $1, \alpha, \dots, \alpha^{n-1}$, that is to say $\mathbf{Q}(\alpha)$ may be identified with the polynomials in α of degree $< n$, and hence $[\mathbf{Q}(\alpha) : \mathbf{Q}] = \deg m_\alpha = n$.*

Proof. Suppose first that $[\mathbf{Q}(\alpha) : \mathbf{Q}]$ is finite, say equal to n. In particular, the numbers $1, \alpha, \dots, \alpha^n$ must be linearly dependent over $\mathbf{Q}$, which means precisely that α satisfies some polynomial equation with coefficients in $\mathbf{Q}$ (of degree $\leqslant n$) and hence is algebraic.

In the other direction, suppose that $\alpha \in \overline{\mathbf{Q}}$, and that m_α is the minimal polynomial of α, with $\deg m_\alpha = n$. Consider the evaluation map $\mathbf{Q}[X] \rightarrow \mathbf{Q}[\alpha]$, which sends $f(X)$ to $f(\alpha)$. This is a surjective ring homomorphism whose kernel is the set of polynomials in $\mathbf{Q}[X]$ satisfied by α. As we saw above, this is precisely (m_α), the ideal generated by m_α. Therefore

$$\mathbf{Q}[\alpha] \cong \mathbf{Q}[X]/(m_\alpha).$$

Now (m_α) is a maximal ideal in $\mathbf{Q}[X]$ (since all ideals in $\mathbf{Q}[X]$ are principal, and if $(m_\alpha) \subseteq (f)$ then $f | m_\alpha$ and so $(f) = (1)$ or (m_α)). Therefore the quotient $\mathbf{Q}[X]/(m_\alpha)$ is actually a *field*. We have shown that the polynomial ring $\mathbf{Q}[\alpha]$ is in fact a field, and so of course it must be $\mathbf{Q}(\alpha)$.

Suppose that $f(\alpha) \in \mathbf{Q}[\alpha]$. By the Euclidean algorithm,

$$f(X) = m_\alpha(X) q(X) + r(X)$$

where $\deg r < n$, and so $f(\alpha) = r(\alpha)$. That is, $\mathbf{Q}[\alpha]$ is spanned by $1, \alpha, \dots, \alpha^{n-1}$. In the other direction, these elements are independent over $\mathbf{Q}$ since otherwise there would be a nonzero polynomial of degree $< n$ satisfied by α. ∎

Remark. To help in understanding all this, let us explain a little more explicitly and algorithmically why inverses exist in $\mathbf{Q}[\alpha]$, a fact which is surprising at first sight. Let $f(\alpha) \in \mathbf{Q}[\alpha]$, $f(\alpha) \neq 0$. Then f is not divisible by m_α and so is coprime it. By the Euclidean algorithm there are polynomials q, p such that $f(X)q(X) + m_\alpha(X)p(X) = 1$. Thus $f(\alpha)q(\alpha) = 1$, so $q(\alpha)$ is the inverse of $f(\alpha)$.

Examples. The field $\mathbf{Q}(i) = \{a + bi : a, b \in \mathbf{Q}\}$, with the inverse being given by $\frac{1}{a+bi} = \frac{a-bi}{a^2+b^2}$. Similarly $\mathbf{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbf{Q}\}$, with $\frac{1}{a+b\sqrt{2}} = \frac{a-b\sqrt{2}}{a^2-2b^2}$.

Corollary 1.4. *Suppose that α satisfies an equation of degree n over $\mathbf{Q}$. Then $[\mathbf{Q}(\alpha) : \mathbf{Q}] \leqslant n$.*

Proof. The minimal polynomial of α has degree $\leqslant n$, so the result follows straight away from Lemma 1.3. □

Arbitrary fields. Everything in this section in fact holds with $\mathbf{Q}$ replaced by an arbitrary field k, and the proofs are essentially the same. The definitions of *algebraic* and *minimal polynomial* must all be taken with respect to k. We did not state results in this generality, because our main concern in this course is the case $k = \mathbf{Q}$. In this case we can cheat somewhat, at least from the point of view of exposition, because we already have the complex numbers $\mathbf{C}$ at our disposal, in which we may locate $\overline{\mathbf{Q}}$ as a specific subset. For general fields k, extensions $k(\alpha)$ and an algebraic closure $\overline{k}$ must be constructed abstractly. (This is probably the "correct" way to proceed when $k = \mathbf{Q}$ as well.) For the details, see the *Galois theory* course.

Let us particularly note the following.

Lemma 1.5. *Let k be a field. If α satisfies a polynomial of degree n over k, then $k[\alpha] = k(\alpha)$ is a field and $[k(\alpha) : k] \leqslant n$. If α satisfies an irreducible monic polynomial of degree n over k, then $[k(\alpha) : k] = n$.*

We will need this twice. In Lemma 1.6 we will need it when $k = \mathbf{Q}(\alpha)$ in which case, since this field is contained in $\mathbf{C}$, the proof goes *exactly* as for $k = \mathbf{Q}$. Later, in Lemma 9.1, we will need the case $k = \mathbf{F}_p$.

The algebraic numbers are a field. We now turn to some basic field-theoretic properties of algebraic numbers.

Lemma 1.6. *Suppose that α, β are algebraic. Then*

$$[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}(\alpha)] \leqslant [\mathbf{Q}(\beta) : \mathbf{Q}].$$

Proof. Let m_β be the minimal polynomial of β. Suppose it has degree n, thus $[\mathbf{Q}(\beta) : \mathbf{Q}] = n$. Now m_β may *also* be regarded as a polynomial of degree n over $k = \mathbf{Q}(\alpha)$, and of course it is satisfied by β (it might not be the minimal polynomial for β over k, though). Therefore by Lemma 1.5 we have $[k(\beta) : k] \leqslant n$. □

Corollary 1.7. *Suppose that α, β are algebraic. Then*

$$[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}] \leqslant [\mathbf{Q}(\alpha) : \mathbf{Q}][\mathbf{Q}(\beta) : \mathbf{Q}].$$

Proof. If $K_1 \subset K_2 \subset K_3$ are fields then

$$[K_3 : K_1] \leqslant [K_3 : K_2][K_2 : K_1]. \tag{1.1}$$

Indeed if $e_1, \dots, e_n$ is a basis for K_2 over K_1, and $f_1, \dots, f_m$ a basis for K_3 over K_2, then an easy exercise shows that

$$\{e_i f_j : 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\} \tag{1.2}$$

spans K_3 over K_1. (In fact (1.1) is an equality, the so-called *tower law* for field extensions. This is because (1.2) is actually a *basis* for K_3 over K_1, which is another easy exercise, and also in the *Galois theory* course). Applying (1.1) with $K_1 = \mathbf{Q}$, $K_2 = \mathbf{Q}(\alpha)$, and $K_3 = \mathbf{Q}(\alpha, \beta)$ we get

$$[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}] = [\mathbf{Q}(\alpha, \beta) : \mathbf{Q}(\alpha)][\mathbf{Q}(\alpha) : \mathbf{Q}].$$

The result now follows immediately from Lemma 1.6. □

Proposition 1.8. *The algebraic numbers $\overline{\mathbf{Q}}$ are a field.*

Proof. Suppose that $\alpha, \beta \in \overline{\mathbf{Q}}$. By Corollary 1.7, $[\mathbf{Q}(\alpha, \beta) : \mathbf{Q}]$ is finite. Since $\mathbf{Q}(\alpha + \beta) \subseteq \mathbf{Q}(\alpha, \beta)$, $[\mathbf{Q}(\alpha + \beta) : \mathbf{Q}]$ is finite, and so by Lemma 1.3 $\alpha + \beta$ is algebraic. Similarly, $\alpha\beta$ is algebraic. □

Number fields. The primitive element theorem. We have seen that if α is algebraic then $\mathbf{Q}(\alpha)$ is a finite degree extension of $\mathbf{Q}$.

Definition 1.9. A *number field* K is a subfield of $\mathbf{C}$ which is a finite degree extension of $\mathbf{Q}$.

Lemma 1.10. *Let $\alpha \in \mathbf{C}$. Then α is algebraic if and only if it lies in some number field K.*

Proof. If α is algebraic, take $K = \mathbf{Q}(\alpha)$. Conversely, if $\alpha \in K$, where $[K : \mathbf{Q}] = n$, observe that $1, \alpha, \alpha^2, \dots, \alpha^n$ are linearly dependent over $\mathbf{Q}$ and so α satisfies some polynomial equation over $\mathbf{Q}$. □

Proposition 1.11 (Primitive element theorem)**.** *Every number field K is of the form $\mathbf{Q}(\theta)$ for some algebraic number θ.*

Proof. *The key fact we will need is that there are only finitely many fields intermediate between $\mathbf{Q}$ and K. This follows from the fundamental theorem of Galois theory: consider some $\tilde{K} \supseteq K$ (for example, the normal closure) such that $\tilde{K}/\mathbf{Q}$ has finite degree and is Galois. Then the subfields of $\tilde{K}$ are in one-to-one correspondence with the subgroups of $\mathrm{Gal}(\tilde{K}/\mathbf{Q})$. This being a finite group, it only has finitely many subgroups.

Turning to the proposition at hand, certainly every number field is finitely generated, that is to say $K = \mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ for some n (if not, keep adding new elements; the degree increases each time).

By induction, it suffices to check that any number field $K = \mathbf{Q}(\alpha, \beta)$ generated by two elements is in fact generated by one element. By the key fact (and the pigeonhole principle), there must be two different rationals c_1, c_2 such that $\mathbf{Q}(\alpha + c_1\beta) = \mathbf{Q}(\alpha + c_2\beta)$. Take $\theta = \alpha + c_1\beta$. Then $\alpha + c_2\beta \in \mathbf{Q}(\theta)$ and hence both α and β lie in this field since they may be expressed as a rational combination of $\alpha + c_1\beta$ and $\alpha + c_2\beta$. $\square$

Remark. θ is not unique – in fact a "generic" $\theta \in K$ is likely to work. For instance, $\mathbf{Q}(\sqrt{2})$ is generated by any $a + b\sqrt{2}$ with $b \neq 0$.

More examples. We now give some examples. The first is the most important for this course: all of the examples and calculations in this course will be quadratic fields.

Example. (Quadratic fields). Suppose the minimal polynomial m_α is an irreducible quadratic over $\mathbf{Q}$, say $m_\alpha(X) = X^2 + bX + c$. The roots of this are of course $\frac{-b\pm\sqrt{d}}{2}$, where $d = b^2 - 4c$. The field generated by either root is $\mathbf{Q}(\sqrt{d})$; the irreducibility of m_α manifests in the fact that d is not a square. By clearing denominators and removing square factors, one may assume that d is in fact a squarefree integer, other than 1. For instance, $\mathbf{Q}(\sqrt{\frac{12}{19}}) = \mathbf{Q}(\sqrt{12 \cdot 19}) = \mathbf{Q}(\sqrt{3 \cdot 19}) = \mathbf{Q}(\sqrt{57})$.

Moreover, all these fields are distinct. To see this, suppose that $\mathbf{Q}(\sqrt{d_1}) = \mathbf{Q}(\sqrt{d_2})$, with d_1, d_2 squarefree integers. Then $u + v\sqrt{d_1} = \sqrt{d_2}$ for some $u, v \in \mathbf{Q}$, which implies that $2uv\sqrt{d_1} = d_2 - u^2 - d_1 v^2$. This can only happen if $uv = 0$. If $v = 0$ then $d_2 = u^2$, contrary to the fact that d_2 is squarefree. If $u = 0$, $d_2 = d_1 v^2$, which again cannot happen for squarefree integers d_1, d_2 (consider prime factorisations).

Example. (Cubic fields). We have already discussed the example $\mathbf{Q}(2^{1/3})$. This is an example of a *pure* cubic field. More generally, one may consider

α with a minimal polynomial $m_\alpha(X) = X^3 + pX + q$; there is more on this, including the criterion for irreducibility, on the first example sheet. This is the most general type of cubic field since one may always remove the X^2 term from a cubic $X^3 + aX^2 + bX + c$ by substituting $Y = X - \frac{a}{3}$, and the resulting field will be the same. We will occasionally touch on cubic fields as a source of examples on the sheets, but already they can be difficult to work with by hand.

Example. (Cyclotomic fields). These are fields $\mathbf{Q}(\zeta_n)$ where ζ_n is a primitive nth root of unity, satisfying the polynomial $X^n - 1 = 0$. (This will not be the minimal polynomial, as $X^n - 1$ is reducible.) The case $n = p$ prime is an important and interesting example and takes up a portion of Sheet 2.

Example. (Quartic fields). General quartic (i.e. degree 4) fields are too complicated as a source of examples in this course. However we will occasionally look at *biquadratic* fields such as $K = \mathbf{Q}(\sqrt{2}, \sqrt{3})$. In this case, the primitive element theorem is not obvious just by looking at the field; on Sheet 1, we will show that indeed $K = \mathbf{Q}(\theta)$ where $\theta = \sqrt{2} + \sqrt{3}$ (for example).

Conjugates and embeddings. Suppose that α is an algebraic number with minimal polynomial m_α of degree n. Then the roots of m_α are called the *conjugates* of α.

Example. The conjugates of $\sqrt{2}$ are $\pm\sqrt{2}$. The conjugates of i are $\pm i$. The minimal polynomial of $2^{1/3}$ is $X^3 - 2$ (which is irreducible by Gauss's lemma (see Lemma C.1) since it has no integer root, or alternatively by Eisenstein's criterion). Hence the conjugates of $2^{1/3}$ are $\omega 2^{1/3}$ and $\omega^2 2^{1/3}$; note in particular that these do not lie in $K = \mathbf{Q}(2^{1/3})$.

In Lemma 1.13 below we will show that the field $\mathbf{Q}$ is *perfect*, which means that the roots (in $\overline{\mathbf{Q}}$) of any irreducible polynomial are distinct. Thus the conjugates of any algebraic number are distinct. These facts will be familiar to anyone having taken a course on Galois theory, but we include the (short) proof here.

We isolate a general lemma from the proof. We state it for general fields since we will need the case $k = \mathbf{F}_p$ later, for a different purpose.

Lemma 1.12. *Let k be a field, and suppose that $f(X), g(X) \in k[X]$. Suppose that f, g gave a common root in some field extension of k. Then $f(X)$ and $g(X)$ have a common factor in $k[X]$.*

Proof. Suppose not. Then $f(X), g(X)$ are coprime in $k[X]$, and so by Euclid's algorithm there are polynomials $a(X), b(X) \in k[X]$ such that $f(X)a(X)+g(X)b(X) = 1$. If α is a common root of f, g (in some extension field of k) then substituting $X = \alpha$ immediately gives a contradiction. □

Lemma 1.13. *Let $f(X) \in \mathbf{Q}[X]$ be irreducible. Then the roots of f in $\overline{\mathbf{Q}}$ are distinct. Thus the conjugates of any algebraic number are distinct.*

Proof. If f had a repeated root β in $\mathbf{C}$ then $f(X) = (X-\beta)^2 g(X)$ (for some $g \in \mathbf{C}[X]$) and hence the derivative $f'(X) = 2(X-\beta)g(X)+(X-\beta)^2 g'(X)$ would also have β as a root. By Lemma 1.12, f and f' would have a common factor over in $\mathbf{Q}[X]$. Since f' is not zero, this is contrary to the assumption that f is irreducible. □

Remarks. *The only place we used the fact that the underlying field is $\mathbf{Q}$ was when we asserted that f' is not zero. Indeed, if

$$f(X) = a_n X^n + \cdots + a_0 \tag{1.3}$$

then

$$f'(X) = n a_n X^{n-1} + \ldots \neq 0. \tag{1.4}$$

All we used about $\mathbf{Q}$ is that it has characteristic zero. By contrast, in $\mathbf{F}_p$ there do exist nonconstant polynomials, such as X^p, with zero derivative. It turns out that finite fields are nonetheless perfect (by a more elaborate argument). However there do exist nonperfect fields of positive characteristic.

Let us also remark that the derivative f' is a purely algebraic object – we are not doing calculus. We omit a detailed discussion, but the key point is that (1.4) can be taken as the *definition* of the derivative, and then one may prove key facts such as the product rule (which we used in the proof of Lemma 1.13) algebraically. When this is done, the derivative makes sense over an arbitrary field*.

As a consequence of Lemma 1.13, if $\alpha_1, \ldots, \alpha_n$ are the conjugates of α (including α, which by convention we take to be α_1) then

$$m_\alpha(X) = \prod_{j=1}^{n} (X - \alpha_j).$$

Note that m_α, since it is irreducible and satisfied by each α_j, is also the minimal polynomial for each of the conjugates α_j.

Now we discuss embeddings. Let K be a number field. Then an *embedding* is a field homomorphism $\sigma : K \to \mathbf{C}$ which preserves $\mathbf{Q}$ (pointwise). It is an easy exercise to see that σ must be injective (in fact, any field homomorphism mapping 0 to 0 and 1 to 1 is injective) and so K is isomorphic to $\sigma(K)$.

Proposition 1.14. *Let $K = \mathbf{Q}(\theta)$ be a number field of degree n. Then any embedding $\sigma : K \to \mathbf{C}$ maps θ to one of its conjugates θ_i. Conversely, for each i there is a unique embedding $\sigma_i : K \to \mathbf{C}$ with $\sigma(\theta) = \theta_i$. In particular, K has exactly n distinct embeddings.*

Proof. *Suppose that m_θ is the minimal polynomial of θ, thus $n = \deg m_\theta$. Let $\sigma : K \to \mathbf{C}$ be an embedding. Then

$$0 = \sigma(m_\theta(\theta)) = m_\theta(\sigma(\theta))$$

and so $\sigma(\theta)$ must be a root of m_θ, that is to say one of the θ_i.

It is also easy to see that if σ is an embedding then it is uniquely determined by its value on θ: indeed (if $c_0, \dots, c_{n-1} \in \mathbf{Q}$) then

$$\sigma(c_0 + c_1\theta + \cdots + c_{n-1}\theta^{n-1}) = c_0 + c_1\sigma(\theta) + \cdots + c_{n-1}\sigma(\theta)^{n-1}.$$

It follows that there are at most n embeddings from K to $\mathbf{C}$.

To see that these embeddings do exist, recall that m_θ is the minimal polynomial of each of the θ_i. Thus

$$\mathbf{Q}(\theta_i) \cong \mathbf{Q}[X]/(m_{\theta_i}) = \mathbf{Q}[X]/(m_{\theta_j}) \cong \mathbf{Q}(\theta_j).$$

Here, the isomorphism

$$\mathbf{Q}[X]/(m_{\theta_i}) \to \mathbf{Q}(\theta_i)$$

is given by evaluation, i.e. $f(X) \to f(\theta_i)$, and similarly for j. Thus the isomorphism $\mathbf{Q}(\theta_i) \cong \mathbf{Q}(\theta_j)$ maps θ_i to θ_j. By convention, we are taking $\theta = \theta_1$, so taking $i = 1$ gives the statement we need. □

Remarks. Just to be clear, although we used the primitive element θ in the proof, the notion of embedding depends only on K, and not on θ (which will, in general, be very far from unique). There is no canonical ordering of the σ_i, but it is usual to take σ_1 to be the identity.

Examples. When $K = \mathbf{Q}(i)$, the two embeddings are the identity map and complex conjugation.

When $K = \mathbf{Q}(\sqrt{2})$, the two embeddings are the identity map and the map which sends $\sqrt{2}$ to $-\sqrt{2}$, thus $\sigma(a + b\sqrt{2}) = a - b\sqrt{2}$.

More generally the same holds for any quadratic field $K = \mathbf{Q}(\sqrt{d})$ with d a squarefree integer.

When $K = \mathbf{Q}(2^{1/3})$, there are three embeddings: the identity σ_1, the map σ_2 defined by $\sigma_2(2^{1/3}) = \omega 2^{1/3}$, and the map $\sigma_3(2^{1/3}) = \omega^2 2^{1/3}$. Note in particular that for these embeddings (unlike the two quadratic examples) we do *not* have $\sigma(K) = K$. (The reason for this is that $K/\mathbf{Q}$ is not Galois.)

Norms. Let K be a number field of degree n, and let $\sigma_1, \ldots, \sigma_n : K \to \mathbf{C}$ be its embeddings. If $\alpha \in K$, we define the *norm*

$$N_{K/\mathbf{Q}}(\alpha) := \prod_{i=1}^{n} \sigma_i(\alpha). \tag{1.5}$$

Examples. If $K = \mathbf{Q}(i)$ then $N_{K/\mathbf{Q}}(a + ib) = (a + ib)(a - ib) = a^2 + b^2$.

If $K = \mathbf{Q}(\sqrt{d})$ then $N_{K/\mathbf{Q}}(a+b\sqrt{d}) = (a+b\sqrt{d})(a-b\sqrt{d}) = a^2-db^2$. An important thing to note here is that this will be nonnegative if $d < 0$ but *not* if $d > 0$. For instance when $K = \mathbf{Q}(\sqrt{2})$ we have $N_{K/\mathbf{Q}}(a+b\sqrt{2}) = a^2 - 2b^2$ which certainly takes negative values.

The following facts follow immediately from the fact that the embeddings σ_i are field homomorphisms preserving $\mathbf{Q}$:

$$N_{K/\mathbf{Q}}(\alpha\beta) = N_{K/\mathbf{Q}}(\alpha)N_{K/\mathbf{Q}}(\beta),$$
$$N_{K/\mathbf{Q}}(\gamma) = 0 \text{ if and only if } \gamma = 0;$$
$$N_{K/\mathbf{Q}}(q) = q^n \text{ for } q \in \mathbf{Q}.$$

This last point, though obvious, should be carefully noted: the norm of an algebraic integer α is not an absolute function of α, but depends on the field K in which α sits. When $K = \mathbf{Q}(\sqrt{2})$, $N_{K/\mathbf{Q}}(5 + \sqrt{2}) = 23$. When looking at Sheet 1, Q2, you might want to try calculating $N_{K/\mathbf{Q}}(5 + \sqrt{2})$ when K is the larger field $\mathbf{Q}(\sqrt{2}, \sqrt{3})$.

The following fact is not so obvious. We first give a (very much nonexaminable) proof using a little Galois theory; we will give a second proof later.

Lemma 1.15. *The norm $N_{K/\mathbf{Q}}$ takes values in $\mathbf{Q}$.*

Proof. *Let $K = \mathbf{Q}(\theta)$. Let $\tilde{K} \supseteq K$, $\tilde{K} \subseteq \mathbf{C}$ be the splitting field of θ, so $\tilde{K}/\mathbf{Q}$ is Galois. All the conjugates of θ lie in $\tilde{K}$ and so $\sigma_i(K) \subseteq \tilde{K}$ for all i. Thus if $\sigma \in \mathrm{Gal}(\tilde{K}/\mathbf{Q})$ we can define the composites $\sigma\sigma_i : K \to \tilde{K}$. These will all be embeddings of K, and they are distinct. Thus $\{\sigma\sigma_1, \dots, \sigma\sigma_n\}$ is a permutation of $\{\sigma_1, \dots, \sigma_n\}$. It follows that

$$\sigma(N_{K/\mathbf{Q}}(\alpha)) = \prod_{j=1}^{n} \sigma\sigma_j(\alpha) = \prod_{j'=1}^{n} \sigma_{j'}(\alpha) = N_{K/\mathbf{Q}}(\alpha).$$

Thus $N_{K/\mathbf{Q}}(\alpha)$ is invariant under the whole Galois group G and hence, by Galois theory, is rational. □

Example. *I recommend trying this out on a nontrivial example beyond the quadratic ones discussed above. For instance, when $K = \mathbf{Q}(2^{1/3})$ we have $\tilde{K} = \mathbf{Q}(2^{1/3}, \omega)$, where $\omega = e^{2\pi i/3}$, and a nontrivial element $\sigma \in \mathrm{Gal}(\tilde{K}/\mathbf{Q})$ is the one with $\sigma(2^{1/3}) = \omega 2^{1/3}$ and $\sigma(\omega) = \omega^2$. If σ_i is the embedding with $\sigma_i(2^{1/3}) = \omega^i 2^{1/3}$ $(i = 0, 1, 2)$ then we have $\sigma\sigma_0 = \sigma_1$, $\sigma\sigma_1 = \sigma_0$, $\sigma\sigma_2 = \sigma_2$.

Norms and determinants. Suppose that K is a number field and that $e_1, \dots, e_n$ is a basis for K over $\mathbf{Q}$. Then for various reasons it is natural[1] to introduce the matrix $M = M(e_1, \dots, e_n)$ whose (i, j)th entry is $M_{ij} = \sigma_i(e_j)$.

Lemma 1.16. *Suppose that $e'_1, \dots, e'_n$ is another basis for K over $\mathbf{Q}$ and that the change of basis is given by*

$$e'_j = \sum_k A_{kj} e_k, \tag{1.6}$$

where $A_{kj} \in \mathbf{Q}$. Let $M' = M(e'_1, \dots, e'_n)$. Then $M' = MA$.

Proof. Indeed, since σ_i is a field homomorphism fixing $\mathbf{Q}$ we have

$$M'_{ij} = \sigma_i(e'_j) = \sum_k A_{kj}\sigma_i(e_k) = \sum_k M_{ik}A_{kj} = (MA)_{ij}.$$

This concludes the proof. □

Lemma 1.17. *The matrix $M(e_1, \dots, e_n)$ is always nonsingular (if $e_1, \dots, e_n$ is a basis for K over $\mathbf{Q}$).*

Proof. By the preceding lemma, we need only find *one* basis for which this is so. Suppose $K = \mathbf{Q}(\theta)$, and take the basis $1, \theta, \cdots, \theta^{n-1}$, that is to say $e_j = \theta^{j-1}$. Then $M_{ij} = \sigma_i(\theta^{j-1}) = x_i^{j-1}$, where $x_i := \sigma_i(\theta)$. Note that the

[1] Note, however, that this is not canonically defined, since there is no natural ordering on the embeddings $\sigma_1, \dots, \sigma_n$. Different orderings permute the rows of the matrix.

x_i, being the conjugates of θ, are distinct by Lemma 1.13. The determinant $\det M$ is then what is known as a *Vandermonde* determinant, and its value is $\prod_{i<j}(x_i - x_j) \neq 0$. (The evaluation of the Vandermonde determinant is an exercise on Sheet X.) □

We may now give an alternative interpretation of the norm, as the determinant of the multiplication-by-α map, as a linear map from K to K as vector spaces over $\mathbf{Q}$. This gives a second proof that $N_{K/\mathbf{Q}}(\alpha) \in \mathbf{Q}$, not using any Galois theory.

Lemma 1.18. *Let $\alpha \in K$. Then $N_{K/\mathbf{Q}}(\alpha)$ is the determinant of the multiplication-by-α map from K to K, considered as a $\mathbf{Q}$-linear map.*

Proof. Let $e_1, \dots, e_n$ be some basis for K over $\mathbf{Q}$. Let $e'_j := \alpha e_j$, and suppose that

$$e'_j = \sum_k A_{kj} e_k \tag{1.7}$$

with $A_{kj} \in \mathbf{Q}$. Thus A is the matrix of the multiplication-by-α map, with respect to the basis $e_1, \dots, e_n$. Let $M = M(e_1, \dots, e_n)$ and $M' = M(e'_1, \dots, e'_n)$. Then, as we saw above,

$$M' = MA. \tag{1.8}$$

Note, however, that

$$M'_{ij} = \sigma_i(e'_j) = \sigma_i(\alpha)\sigma_i(e_j) = \sigma_i(\alpha) M_{ij},$$

and so

$$M' = DM \tag{1.9}$$

where D is the diagonal matrix with $D_{ii} = \sigma_i(\alpha)$. It follows, since M is nonsingular (by Lemma 1.17), that $A = M^{-1}DM$, and therefore

$$\det A = \det D = \prod_i \sigma_i(\alpha) = N_{K/\mathbf{Q}}(\alpha). \tag{1.10}$$

This concludes the proof. □

Examples. Let us first check a quadratic example. When $K = \mathbf{Q}(i)$, a basis for K over $\mathbf{Q}$ is $\{e_1, e_2\} = \{1, i\}$. Let $\alpha = 2 + i$. Then

$$e'_1 = (2+i)e_1 = 2 + i = 2e_1 + e_2,$$

$$e'_2 = (2+i)e_2 = (2+i)i = -e_1 + 2e_2.$$

Thus

$$\det A = \begin{vmatrix} 2 & -1 \\ 1 & 2 \end{vmatrix} = 5,$$

which does indeed conform with what we saw earlier.

Now let us look at a cubic example, where Lemma 1.18 actually makes the computation of the norm easier than using the definition in terms of conjugates. Suppose that $\alpha = a + b2^{1/3} + c2^{2/3}$ in $K = \mathbf{Q}(2^{1/3})$. Let $e_1 = 1$, $e_2 = 2^{1/3}$, $e_3 = 2^{2/3}$. Let $e_i' = \alpha e_i$. Then we can compute

$$\begin{aligned} e_1' &= ae_1 + be_2 + ce_3, \\ e_2' &= 2ce_1 + ae_2 + be_3, \\ e_3' &= 2be_1 + 2ce_2 + ae_3. \end{aligned}$$

Thus

$$N_{K/\mathbf{Q}}(\alpha) = \begin{vmatrix} a & b & c \\ 2c & a & b \\ 2b & 2c & a \end{vmatrix} = a^3 + 2b^3 + 4c^3 - 6abc.$$

Discriminants. In this section we introduce the notion of *discriminant*. We will use the word in two different ways in these notes. First, in this chapter, a discriminant is associated with an n-tuple of elements. In the next chapter we will use this notion to define the discriminant Δ_K of a number field, which is a single quantity associated to K and somehow measuring its "size".

Let K be a number field with embeddings $\sigma_1, \dots, \sigma_n$.

Definition 1.19. Let $e_1, \dots, e_n$ be a basis for K over $\mathbf{Q}$. Then we define the *discriminant* $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$ to be $(\det M)^2$, where $M = M(e_1, \dots, e_n)$, as above, is the matrix with $M_{ij} = \sigma_i(e_j)$.

It follows from Lemma 1.17 that $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) \neq 0$. An important alternative expression for $\mathrm{disc}_{K/\mathbf{Q}}$ involves the *trace*, which we define now.

Definition 1.20. Suppose that $\alpha \in K$. Then the *trace* $\mathrm{tr}_{K/\mathbf{Q}}(\alpha)$ is defined to be $\sum_i \sigma_i(\alpha)$, the sum being over all embeddings of K.

Lemma 1.21. *For all α we have $\mathrm{tr}_{K/\mathbf{Q}}(\alpha) \in \mathbf{Q}$.*

Proof. *As with the norm, a short proof may be given using Galois theory, and in fact the proof is almost exactly the same as for the norm: suppose $K = \mathbf{Q}(\theta)$, and let $\tilde{K}$ be the splitting field of θ, so that $\tilde{K}/\mathbf{Q}$ is Galois.

For $\sigma \in \mathrm{Gal}(\tilde{K}/\mathbf{Q})$ the embeddings $\sigma\sigma_1, \ldots, \sigma\sigma_n$ are a rearrangement of $\sigma_1, \ldots, \sigma_n$, and so

$$\sigma(\mathrm{tr}_{K/\mathbf{Q}}(\alpha)) = \sum_k \sigma\sigma_k(\alpha) = \sum_{k'} \sigma_{k'}(\alpha) = \mathrm{tr}_{K/\mathbf{Q}}(\alpha).$$

Thus $\mathrm{tr}_{K/\mathbf{Q}}(\alpha)$ is invariant under $\mathrm{Gal}(\tilde{K}/\mathbf{Q})$ and hence is rational*.

We may also note from the proof of Lemma 1.18 that $\mathrm{tr}_{K/\mathbf{Q}}(\alpha)$ is the trace of the multiplication-by-α map from K to K. Indeed (in the notation of that proof)

$$\mathrm{tr}(A) = \mathrm{tr}(M^{-1}DM) = \mathrm{tr}(D) = \sum_i \sigma_i(\alpha) = \mathrm{tr}_{K/\mathbf{Q}}(\alpha).$$

Either way, the proof is complete. □

The link between the discriminant and the trace is as follows. First note that

$$\mathrm{disc}_{K/\mathbf{Q}}(e_1, \ldots, e_n) = (\det M)^2 = \det(M^T M).$$

However, $M^T M$ has (i, j)-entry $\sum_k \sigma_k(e_i)\sigma_k(e_j) = \sum_k \sigma_k(e_i e_j) = \mathrm{tr}_{K/\mathbf{Q}}(e_i e_j)$, thus

$$\mathrm{disc}_{K/\mathbf{Q}}(e_1, \ldots, e_n) = \det(\mathrm{tr}_{K/\mathbf{Q}}(e_i e_j)_{i,j}).$$

From this and Lemma 1.21, the following is immediate.

Lemma 1.22. *We have* $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \ldots, e_n) \in \mathbf{Q}$. □

Remark. The discriminant, whilst being rational and the square of something ($\det M$), is not necessarily positive. For instance,

$$\mathrm{disc}_{\mathbf{Q}(i)/\mathbf{Q}}(1, i) = \begin{vmatrix} 1 & i \\ 1 & -i \end{vmatrix}^2 = -4.$$

The following fact about how discriminants fare under base change is immediate from the corresponding fact for M, namely Lemma 1.16.

Lemma 1.23. *Suppose that $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_n \in K$ are related by $e'_j = \sum_k A_{kj} e_k$, where the matrix A has rational entries. Then*

$$\mathrm{disc}_{K/\mathbf{Q}}(e'_1, \ldots, e'_n) = (\det A)^2 \, \mathrm{disc}_{K/\mathbf{Q}}(e_1, \ldots, e_n).$$ □

We conclude by remarking that, regarding notation for the discriminant, there is not absolute consistency in the literature, or indeed in past exam questions. Sometimes people write Δ instead of M, and the discriminant becomes Δ^2. For us, the notation M is an auxillary one which is used to establish basic properties of the norm and discriminant.

2. Algebraic integers

Algebraic integers. In the last section we defined algebraic numbers. The notion of an algebraic *integer* is crucial in this course.

Definition 2.1. Suppose that $\alpha \in \overline{\mathbf{Q}}$ is an algebraic number. Then α is an *algebraic integer* if it satisfies a *monic* polynomial in $\mathbf{Z}[X]$.

Examples. A rational number is an algebraic integer if and only if it is an integer. The algebraic integers in $\mathbf{Q}(i)$ are $\{a + bi : a, b \in \mathbf{Z}\}$, and the algebraic integers in $\mathbf{Q}(\sqrt{2})$ are $\{a + b\sqrt{2}, a, b \in \mathbf{Z}\}$. We caution that the obvious generalization of this pattern to $\mathbf{Q}(\sqrt{d})$ fails. Indeed, the golden ratio $\frac{1}{2}(1 + \sqrt{5})$ is an algebraic integer, because it satisfies $X^2 - X - 1 = 0$. We will study the integers in quadratic fields in full generality later on.

The set of algebraic integers is denoted by $\mathcal{O}$. Note that the traditional integers $\mathbf{Z}$ are all algebraic integers. Usually, we will just call these "integers", but occasionally we will call them *rational integers* if there is a danger of confusion. Similarly, by *rational prime* we mean a prime in $\mathbf{Z}$.

Lemma 2.2. *Let α be an algebraic number. Then α is an algebraic integer if and only if its minimal polynomial m_α has integer coefficients. In particular, a rational number is an algebraic integer if and only if it is an integer, that is to say $\mathcal{O} \cap \mathbf{Q} = \mathbf{Z}$.*

Proof. The "if" direction is trivial. The "only if" direction follows from Gauss's lemma (see Appendix C): Suppose that $f \in \mathbf{Z}[X]$ is the monic integer polynomial of minimal degree satisfied by α. If f is not already the minimal polynomial of α, then $f(X)$ is reducible in $\mathbf{Q}[X]$, and hence in $\mathbf{Z}[X]$, contrary to the minimality assumption. □

Shortly (in Proposition 2.4 below) we are going to prove that the algebraic integers form a ring. The following lemma is very useful in that regard.

Lemma 2.3. *Let K be a number field. Then $\alpha \in K$ is an algebraic integer if and only if there is a nonzero finitely-generated $\mathbf{Z}$-module $V \subseteq K$ such that $\alpha V \subseteq V$.*

Proof. First suppose that α is an algebraic integer. Then we have $\alpha^d = \sum_{i=0}^{d-1} a_i \alpha^i$ for some rational integers a_i. Thus α^d is in the $\mathbf{Z}$-module generated by $1, \alpha, \dots, \alpha^{d-1}$, which therefore has the required property.

Conversely, suppose that $V \subset K$ is a finitely-generated $\mathbf{Z}$ module, with generating set $e_1, \dots, e_n$, and that $\alpha V \subseteq V$.

Then

$$\alpha e_j = \sum_k A_{kj} e_k$$

for some *integers* $A_{kj} \in \mathbf{Z}$. This means that the column vector $(e_1, \dots, e_n)$ lies in the kernel of the $n \times n$ matrix $A - \alpha I$, which therefore has zero determinant. That is, $\det(A - \alpha I) = 0$, which provides a monic polynomial with integer coefficients, satisfied by α. $\square$

Proposition 2.4. *The algebraic integers $\mathcal{O}$ form a ring.*

Proof. Suppose that $\alpha, \beta \in \mathcal{O}$. Then by Lemma 2.3 we can find finitely generated $\mathbf{Z}$-modules V (generated by $e_1, \dots, e_n$) and W (generated by $f_1, \dots, f_m$) such that $\alpha V \subseteq V$ and $\beta W \subseteq W$. Let VW be the $\mathbf{Z}$-module generated by the products vw. This is finitely generated, by the $e_i f_j$. Moreover,

$$(\alpha + \beta) VW \subseteq (\alpha V)W + V(\beta W) \subseteq VW,$$

and similarly

$$(\alpha\beta) VW \subseteq (\alpha V)(\beta W) \subseteq VW.$$

By the other direction of Lemma 2.3, both $\alpha + \beta$ and $\alpha\beta$ are algebraic integers. This completes the proof. $\square$

We finish this section with an easy lemma which is sometimes useful.

Lemma 2.5. *Suppose that $\alpha \in \overline{\mathbf{Q}}$. Then some integer multiple of α is an algebraic integer.*

Proof. Suppose that α satisfies the equation

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0,$$

where $a_0, \dots, a_{n-1} \in \mathbf{Q}$. Then, for any $m \in \mathbf{Z}$, $m\alpha$ satisfies the equation

$$(m\alpha)^n + m a_{n-1} (m\alpha)^{n-1} + \cdots + m^n a_0 = 0.$$

By choosing m suitably, we may clear all the denominators and ensure that all of $m a_{n-1}, m^2 a_{n-2}, \dots, m^n a_0$ are all integers. $\square$

A particular consequence of this is that every element of K is a ratio of two elements of $\mathcal{O}_K$. Therefore K is (isomorphic to) the field of fractions of $\mathcal{O}_K$.

Another consequence, of this and the primitive element theorem, is the following.

Proposition 2.6. *Every number field is of the form $K = \mathbf{Q}(\theta)$ with θ an algebraic integer. In particular, $1, \theta, \theta^2, \dots, \theta^{n-1}$ is a basis for K over $\mathbf{Q}$ consisting of algebraic integers.*

The ring of integers of a number field. If $K \subset \overline{\mathbf{Q}}$ is a number field, we write $\mathcal{O}_K := K \cap \mathcal{O}$ for the algebraic integers which lie in K. This is invariably called the *ring of integers* of K, this being justifiable as a consequence of Proposition 2.4. Let us record some key general facts about $\mathcal{O}_K$.

Lemma 2.7. *Let K be a number field and let $\sigma_1, \dots, \sigma_n \to \mathbf{C}$ be its embeddings. Suppose that $\alpha \in \mathcal{O}_K$. Then $\sigma_i(\alpha)$ is an algebraic integer.*

Proof. Let f be a monic integer polynomial satisfied by α. Then $\sigma_i(f(\alpha)) = f(\sigma_i(\alpha)) = 0$, since σ_i fixes $\mathbf{Q}$ and hence $\mathbf{Z}$. Thus f is also satisfied by $\sigma_i(\alpha)$. This concludes the proof. □

Corollary 2.8. *If $\alpha \in \mathcal{O}_K$ then $N_{K/\mathbf{Q}}(\alpha) \in \mathbf{Z}$ and $\mathrm{tr}_{K/\mathbf{Q}}(\alpha) \in \mathbf{Z}$.*

Proof. Recall the definition of norm, $N_{K/\mathbf{Q}}(\alpha) = \prod_i \sigma_i(\alpha)$. By Lemma 2.7 and the fact that $\mathcal{O}$ is a ring, $N_{K/\mathbf{Q}}(\alpha) \in \mathcal{O}$. However, we have already seen in Lemma 1.15 that $N_{K/\mathbf{Q}}(\alpha) \in \mathbf{Q}$. It follows that $N_{K/\mathbf{Q}}(\alpha) \in \mathcal{O} \cap \mathbf{Q} = \mathbf{Z}$.

The proof for the trace is essentially identical. □

Corollary 2.9. *Suppose that $e_1, \dots, e_n \in \mathcal{O}_K$. Then $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) \in \mathbf{Z}$.*

Proof. We have already shown (just with the assumption that the e_i lie in K) that $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) \in \mathbf{Q}$. Recall that the definition of discriminant was $(\det M)^2$, where the (i, j)-entry of M is $\sigma_i(e_j)$. By Lemma 2.7, each of these entries is an algebraic integer. Therefore (since $\mathcal{O}$ is a ring) $(\det M)^2 \in \mathcal{O}$. Hence $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) \in \mathcal{O} \cap \mathbf{Q} = \mathbf{Z}$. □

Units. Let K be a number field, and $\mathcal{O}_K$ its ring of integers. Note that $\mathcal{O}_K$ (being contained in a field) is an integral domain. A *unit* is an element u for which there is $v \in \mathcal{O}_K$ with $uv = 1$. Equivalently, the inverse u^{-1} (in the field K) in fact lies in $\mathcal{O}_K$. It is easy to see that the units form a group under multiplication.

We will sometimes write $U(\mathcal{O}_K)$ for the group of units in $\mathcal{O}_K$.

Example. The units in $\mathbf{Q}$ are ± 1. The units in $\mathbf{Q}(i)$ are $\{\pm 1, \pm i\}$. However, $\mathbf{Q}(\sqrt{3})$ has infinitely many units, and they can be very large (in the Euclidean norm on $\mathbf{R}$). Indeed, $7+4\sqrt{3}$ is a unit since $(7+4\sqrt{3})(7-4\sqrt{3}) = 1$, and hence so is any power $(7+4\sqrt{3})^n$.

Lemma 2.10. *$u \in \mathcal{O}_K$ is a unit if and only if $N_{K/\mathbf{Q}}(u) = \pm 1$.*

Proof. The only if direction is easy: if $uv = 1$ then $N_{K/\mathbf{Q}}(u)N_{K/\mathbf{Q}}(v) = N_{K/\mathbf{Q}}(uv) = 1$. But $N_{K/\mathbf{Q}}(u), N_{K/\mathbf{Q}}(v)$ are both integers, so must be ± 1.

Conversely, suppose that $N_{K/\mathbf{Q}}(u) = \pm 1$. Set $v := \pm\sigma_2(u)\cdots\sigma_n(u)$. Then $uv = \pm N_{K/\mathbf{Q}}(u) = 1$. Now $u \in \mathcal{O}_K$ is an algebraic integer and hence so are all the conjugates $\sigma_i(u)$, by Lemma 2.7. (Note however that they are not necessarily in K.) Since $\mathcal{O}$ is a ring, $v \in \mathcal{O}$. However, since $v = u^{-1}$, we also have $v \in K$, and so $v \in \mathcal{O} \cap K = \mathcal{O}_K$. Therefore u is a unit. $\square$

***Dirichlet's units theorem.** The schedules of this course do not call for a discussion of the structure of the group of units in general. However, I feel it would be remiss not to mention the main theorem in this regard.

Let K be a number field of degree n, with embeddings $\sigma_1, \dots, \sigma_n : K \to \mathbf{C}$. Some of these, say r of them, will be *real embeddings*, which means that $\sigma_i(K) \subset \mathbf{R}$. The other embeddings are called *complex*, and they must come in conjugate pairs since if $\sigma_i : K \to \mathbf{C}$ is an embedding then so is $\overline{\sigma}_i : K \to \mathbf{C}$, since complex conjugation is an automorphism of $\mathbf{C}$ preserving $\mathbf{Q}$. Suppose there are s complex conjugate pairs; thus $r + 2s = n$.

Theorem 2.11 (Dirichlet's Units Theorem)**.** *Suppose that K is a number field with r real embeddings and s pairs of complex conjugate embeddings. Then the group of units $U(\mathcal{O}_K)$ is isomorphic, as a multiplicative group, to a finite group (the roots of unity in $\mathcal{O}_K$) times $\mathbf{Z}^{r+s-1}$.*

We will not give the proof in this course.

Let us conclude by remarking that the only case in which $r + s - 1 = 0$ is when $r = 0$ and $s = 1$, in which case K is an imaginary quadratic field

$\mathbf{Q}(\sqrt{d})$ with $d < 0$. Thus only in this case are there finitely many units. See Sheet 1, Q4 for a complete description of the units in this case.

Integral bases. Let K be a number field with ring of integers $\mathcal{O}_K$. Since $\mathcal{O}_K$ is a ring containing $\mathbf{Z}$, $\mathcal{O}_K$ is certainly a $\mathbf{Z}$-module. The main result of this section is that this has a particularly nice structure.

Theorem 2.12 (Integral bases)**.** *Suppose K has degree n. Then $\mathcal{O}_K$ is a* free abelian group of rank *n, by which we mean that there are $e_1, \dots, e_n$ such that $\mathcal{O}_K = \bigoplus_{i=1}^n \mathbf{Z}e_i$ (that is, the e_i lie in $\mathcal{O}_K$ and every element of $\mathcal{O}_K$ is an integer combination of the e_i in precisely one way). In this situation, $e_1, \dots, e_n$ is called an* integral basis *for $\mathcal{O}_K$.*

Observe that if $e_1, \dots, e_n$ are an integral basis then they are also a basis for K as a vector space over $\mathbf{Q}$. This is because in any nontrivial $\mathbf{Q}$-relation $q_1e_1 + \cdots + q_ne_n = 0$ we may clear denominators to get a $\mathbf{Z}$-relation, which cannot exist by the definition of integral basis. Thus $e_1, \dots, e_n$ are n $\mathbf{Q}$-linearly independent elements of K, and must therefore be a basis.

Example. $\{1, i\}$ gives an integral basis for $K = \mathbf{Q}(i)$, since $\mathcal{O}_K = \{a + bi : a, b \in \mathbf{Z}\} = \mathbf{Z} \oplus \mathbf{Z}i$. We will specify integral bases for quadratic fields in general in the next section. For cubic and higher fields, it can be rather difficult to compute integral bases, although there are algorithms which are guaranteed to produce them. We will suggest some strategies shortly.

Proof. [Proof of Theorem 2.12.] First observe that there is some $\mathbf{Q}$-basis for K consisting of elements of $\mathcal{O}_K$. This follows by taking an arbitrary basis and multiplying up each element to get an element of $\mathcal{O}_K$, using Lemma 2.5. If $e_1, \dots, e_n$ is such a basis then $\operatorname{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$ is a non-zero integer, by Corollary 2.9 and Lemma 1.17. Suppose that $e_1, \dots, e_n \in \mathcal{O}_K$ are chosen so that $|\operatorname{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)|$ is minimal (subject to being non-zero). We claim that $e_1, \dots, e_n$ is then an integral basis.

Suppose this is not the case. Then (subtracting integer multiples of the e_i) there is some element $\sum_i c_ie_i \in \mathcal{O}_K$ with, for some i, $0 < |c_i| < 1$. Without loss of generality, $i = 1$. Set $e_1' := \sum_i c_ie_i$. Then $e_1', e_2, \dots, e_n$ is a basis for K as a vector space over $\mathbf{Q}$, all of whose elements lie in $\mathcal{O}_K$. Its base change matrix A relative to $e_1, \dots, e_n$ is given by $A_{j1} = c_j$ and $A_{ji} = \delta_{ij}$ when $i \geqslant 2$. Thus $\det(A) = c_1$ and so by Lemma 1.23

$$\operatorname{disc}_{K/\mathbf{Q}}(e_1', e_2, \dots, e_n) = c_1^2 \operatorname{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n).$$

Since $0 < c_1^2 < 1$, this contradicts the supposed minimality. $\square$

Integral bases are not unique. Let $e_1, \dots, e_n$ and $e'_1, \dots, e'_n$ be two bases for K over $\mathbf{Q}$. Then the sums $\bigoplus \mathbf{Z}e_i$ and $\bigoplus \mathbf{Z}e'_i$ are indeed both direct sums. If the base change matrix is given by $e'_i = \sum_j A_{ji}e_j$ then it is easy to see that $\bigoplus \mathbf{Z}e'_i \subseteq \bigoplus \mathbf{Z}e_i$ if, and only if, $A \in \mathrm{Mat}_n(\mathbf{Z})$, the $n \times n$ integer matrices. Similarly $\bigoplus \mathbf{Z}e_i \subseteq \bigoplus \mathbf{Z}e'_i$ if, and only if, $A^{-1} \in \mathrm{Mat}_n(\mathbf{Z})$ is an integer matrix. This implies the following.

Proposition 2.13. *Suppose that $e_1, \dots, e_n$ is an integral basis, and suppose $e'_1, \dots, e'_n$ are elements of K given by $e'_i = \sum_j A_{ji}e_j$. Then $e'_1, \dots, e'_n$ is an integral basis for $\mathcal{O}_K$ if and only if both $A, A^{-1} \in \mathrm{Mat}_n(\mathbf{Z})$.*

A matrix A with this property is called *unimodular*.

Lemma 2.14. *Suppose that $A \in \mathrm{Mat}_n(\mathbf{Z})$. Then A is unimodular if and only if* $\det A = \pm 1$.

Proof. The only if direction is easy: we have $1 = (\det A)(\det A^{-1})$, and if A is unimodular then both $\det A$ and $\det A^{-1}$ are integers.

The if direction requires some nontrivial linear algebra, specifically Cramer's formula for the inverse of a matrix, that is to say $1/\det A$ times the adjoint matrix. This formula makes it clear that if $A \in \mathrm{Mat}_n(\mathbf{Z})$ and $\det A = \pm 1$ then $A^{-1} \in \mathrm{Mat}_n(\mathbf{Z})$. $\square$

As a consequence, the unimodular matrices form a group. It is the double cover of $\mathrm{SL}_n(\mathbf{Z}) = \{A \in \mathrm{Mat}_n(\mathbf{Z}) : \det A = 1\}$. Even when $n = 2$ this group is certainly infinite. For instance, $\begin{pmatrix} 5 & 3 \\ 13 & 8 \end{pmatrix}$ is unimodular.

Corollary 2.15. *Suppose that $e_1, \dots, e_n$ and $e'_1, \dots, e'_n$ are two integral bases for $\mathcal{O}_K$. Then*

$$\mathrm{disc}_{K/\mathbf{Q}}(e'_1, \dots, e'_n) = \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n).$$

Proof. By Proposition 2.13 and Lemma 2.14 we have $e'_i = \sum_j A_{ji}e_j$ with $\det A = \pm 1$. By Lemma 1.23,

$$\begin{aligned} \mathrm{disc}_{K/\mathbf{Q}}(e'_1, \dots, e'_n) &= (\det A)^2 \,\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) \\ &= \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n). \end{aligned}$$

This concludes the proof. $\square$

Corollary 2.15 allows us to make the following definition.

Definition 2.16 (Discriminant of a field)**.** Let K be a number field. Then its discriminant Δ_K is defined to be $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$, where $e_1, \dots, e_n$ is any integral basis for $\mathcal{O}_K$.

We have layered many definitions on top of one another. For the moment one should, roughly thinking, imagine that Δ_K describes the "size" or "density" of the ring of integers $\mathcal{O}_K$. This interpretation will become a little clearer in Section 10.

Quadratic fields. Let us work through some of the concepts just discussed for quadratic fields $\mathbf{Q}(\sqrt{d})$, $d \neq 1$ a squarefree integer.

Proposition 2.17. *Let $K = \mathbf{Q}(\sqrt{d})$, $d \neq 1$ squarefree. Then an integral basis for K is given by*

- 1 *and* $\sqrt{d}$ *if* $d \equiv 2, 3 \pmod 4$*;*
- 1 *and* $\frac{1}{2}(1 + \sqrt{d})$ *if* $d \equiv 1 \pmod 4$*.*

The discriminant Δ_K is given as follows:

- $4d$ *if* $d \equiv 2, 3 \pmod 4$*;*
- d *if* $d \equiv 1 \pmod 4$*.*

Proof. Suppose that $a + b\sqrt{d} \in \mathcal{O}_K$, where $a, b \in \mathbf{Q}$. Then (by Lemma 2.7) $a - b\sqrt{d} \in \mathcal{O}_K$. In particular $(a + b\sqrt{d}) + (a - b\sqrt{d}) = 2a$ (i.e., the trace) lies in $\mathcal{O}_K$, which means that $a = \frac{\ell}{2}$ for some rational integer ℓ. Also, $(a + b\sqrt{d}) - (a - b\sqrt{d}) = 2b\sqrt{d}$ lies in $\mathcal{O}_K$ and hence so does its square $4b^2 d$. Since d is squarefree, the only denominator b could have is 2. Thus we also have $b = \frac{m}{2}$ for some $m \in \mathbf{Z}$. Thus everything in $\mathcal{O}_K$ is, up to adding elements of $\mathbf{Z} \oplus \mathbf{Z}\sqrt{d}$, an element of the set $S := \{0, \frac{1}{2}, \frac{\sqrt{d}}{2}, \frac{1}{2}(1 + \sqrt{d})\}$. The middle two elements of S are easily seen not to be algebraic integers, so we need only decide whether or not $\alpha = \frac{1}{2}(1 + \sqrt{d}) \in \mathcal{O}$. The minimal polynomial $m_\alpha(X)$ is $X^2 - X + \frac{1-d}{4}$, so this is so if and only if $d \equiv 1 \pmod 4$.

The discriminants may now be calculated by simply evaluating 2×2 determinants – we leave this to the reader. $\square$

It follows from Proposition 2.17 that quadratic fields are *monogenic*, meaning that $\mathcal{O}_K = \mathbf{Z}[\alpha]$ for some α. (Sometimes this is called a "power integral basis"). Whilst many fields share this property, it is not universal. On the example sheets, we give an example of a cubic field which is not monogenic.

Computing an integral basis. We managed to compute an integral basis for quadratic fields by hand. For larger fields, this quickly gets difficult. In this section, we give a couple of lemmas which can be helpful in this regard.

Lemma 2.18. *Let K be a number field and suppose that $e_1, \dots, e_n \in \mathcal{O}_K$ are such that $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$ is nonzero and squarefree. Then $e_1, \dots, e_n$ is an integral basis.*

Proof. Let $e'_1, \dots, e'_n$ be some integral basis. Let the base change matrix from the e'_i to the e_i be A, thus $A \in \mathrm{Mat}_n(\mathbf{Z})$. Then by Lemma 1.23 we have $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) = (\det A)^2 \, \mathrm{disc}_{K/\mathbf{Q}}(e'_1, \dots, e'_n)$, and so

$$(\det A)^2 \mid \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n).$$

Since $\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$ is squarefree it follows that $\det A = \pm 1$, and so A is unimodular. By Proposition 2.13, it follows that $e_1, \dots, e_n$ is an integral basis. $\square$

Remark. The converse is not true, so this lemma is by no means universally applicable. One can already see this for quadratic fields since $\Delta_{\mathbf{Q}(i)}$ is divisible by 4.

Lemma 2.21 below is of more general applicability. In the proof we will need the following result about abelian groups.

Lemma 2.19. *Suppose that $e_1, \dots, e_n$ and $e'_1, \dots, e'_n$ are linearly independent tuples in $\mathcal{O}_K$, and that $e'_i = \sum_j A_{ji} e_j$, where $A \in \mathrm{Mat}_n(\mathbf{Z})$. Set $M := \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n$ and $M' := \mathbf{Z}e'_1 \oplus \cdots \oplus \mathbf{Z}e'_n$, thus $M' \subseteq M$. Then $[M : M']$, the index of M' as an additive subgroup of M, is equal to $|\det A|$.*

Proof. See Appendix A. $\square$

Corollary 2.20. *Suppose that $e'_1, \dots, e'_n \in \mathcal{O}_K$ are linearly independent over $\mathbf{Q}$. Write $M' = \mathbf{Z}e'_1 \oplus \cdots \oplus \mathbf{Z}e'_n$. Then*

$$\mathrm{disc}_{K/\mathbf{Q}}(e'_1, \dots, e'_n) = [\mathcal{O}_K : M']^2 \Delta_K.$$

Remark. This is tautologous (given what we have already proven) when $e'_1, \dots, e'_n$ is an integral basis. The point, of course, is that it applies more generally.

Proof. Let $e_1, \dots, e_n$ be an integral basis for $\mathcal{O}_K$, and let A be the base-change matrix expressing the e'_i in terms of the e_i. Then, by Lemma 1.23

and the definition of Δ_K,

$$\mathrm{disc}_{K/\mathbf{Q}}(e_1', \dots, e_n') = (\det A)^2 \, \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n) = (\det A)^2 \Delta_K.$$

However, since $M = \mathcal{O}_K$, it follows from Lemma 2.19 that

$$[\mathcal{O}_K : M'] = [M : M'] = \det A.$$

The result follows. $\square$

Finally, we come to the lemma which is actually useful for computing integral bases in practice.

Lemma 2.21. *Suppose that K is a number field and that $e_1, \dots, e_n$ are elements of $\mathcal{O}_K$, independent over $\mathbf{Q}$, which do not form an integral basis. Then there exists a prime p with $p^2 \mid \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$ and integers $m_1, \dots, m_n \in \{0, \dots, p-1\}$, not all zero, such that $\frac{1}{p}(m_1 e_1 + \dots + m_n e_n) \in \mathcal{O}_K$.*

Proof. Let $M = \mathbf{Z}e_1 \oplus \dots \oplus \mathbf{Z}e_n$. By assumption, $M \neq \mathcal{O}_K$. Therefore there is some prime p dividing $[\mathcal{O}_K : M]$; by Corollary 2.20, $p^2 \mid \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$. By Cauchy's theorem from finite group theory, the additive group $\mathcal{O}_K/M$ has an element of order p. The lift of this in $\mathcal{O}_K$ must be of the form $\frac{1}{p}(m_1 e_1 + \dots + m_n e_n)$, with $m_i \in \mathbf{Z}$ and not all divisible by p. By subtracting elements of M, we may then ensure that all of the m_i lie in $\{0, 1, \dots, p-1\}$, and they are not all zero. $\square$

Suppose that, in the conclusion of Lemma 2.21, $m_1 \neq 0$. By the proof of Proposition 2.12, if we replace e_1 by $e_1' = \frac{1}{p}(m_1 e_1 + \dots + m_n e_n)$, then

$$0 < |\,\mathrm{disc}_{K/\mathbf{Q}}(e_1', e_2, \dots, e_n)| < |\,\mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)|.$$

This allows us to give an algorithm for computing an integral basis which, although potentially painful, is guaranteed to terminate in finite time. The algorithm goes as follows:

- Start with elements $e_1, \dots, e_n$ of $\mathcal{O}_K$ spanning K as a vector space over $\mathbf{Q}$ (for example, one might start with a power basis $1, \theta, \dots, \theta^{n-1}$, the existence of which is guaranteed by Proposition 2.6).
- For each prime p with $p^2 \mid \mathrm{disc}_{K/\mathbf{Q}}(e_1, \dots, e_n)$, test all $\frac{1}{p}(m_1 e_1 + \dots + m_n e_n)$, $0 \leqslant m_i < p$, not all m_i zero, to see if they lie in $\mathcal{O}_K$.
- If none do, $e_1, \dots, e_n$ is an integral basis.

- Suppose that $\frac{1}{p}(m_1e_1 + \cdots + m_ne_n) \in \mathcal{O}_K$, with (say) $m_1 \neq 0$. Set $e_1' := \frac{1}{p}(m_1e_1 + \cdots + m_ne_n)$, then replace $e_1, \ldots, e_n$ with $e_1', e_2, \ldots, e_n$ and return to the start.

Let us additionally remark that we can save a factor of roughly p in the time taken for the second step by observing that if there is some $\frac{1}{p}(m_1e_1 + \cdots + m_ne_n) \in \mathcal{O}_K$ with $p \nmid m_i$, then we can find such an element with $m_i \equiv 1 (\operatorname{mod} p)$, by multiplying up by the inverse of $m_i (\operatorname{mod} p)$. Then we may reduce so that all the m_i lie between 0 and $p-1$, and in particular $m_i = 1$.

3. Irreducibles and factorisation

Basic concepts. Most of the rest of the course is about the multiplicative structure of $\mathcal{O}_K$. As you have known for a long time, when $K = \mathbf{Q}$ (thus $\mathcal{O}_K = \mathbf{Z}$) there is a very nice multiplicative structure: unique decomposition into primes.

Although, at school, you learn that a "prime" is a number with no factors other than itself and ± 1, we will instead call numbers with this property *irreducible*. As you know, $\mathbf{Z}$ has unique factorisation into irreducibles. Let us give the formal definition of what this means. We state the next couple of definitions in the context of arbitrary integral domains R, but you can always think of $R = \mathcal{O}_K$, which is the case of relevance in this course.

Definition 3.1. Let R be an integral domain. An element $x \in R$ is *irreducible* if it is not a unit and if, whenever $x = yz$ with $y, z \in R$, then one of y, z is a unit.

Definition 3.2 (UFD)**.** Let R be an integral domain. Then R is a *unique factorisation domain* (UFD) if it enjoys unique factorisation into irreducibles. More precisely, we have the following.

(i) If $r \in R$ is not zero or a unit, then r can be written as a (finite) product of irreducibles.

(ii) There is essentially a unique way of doing this: if

$$r = x_1 \cdots x_m = y_1 \cdots y_n$$

with x_i, y_j irreducible then $m = n$ and, after relabelling, x_i equals y_i up to a unit, in the sense that there is a unit u_i such that $x_i = y_iu_i$.

Remark. One often says that if x and y differ by a unit then they are *associate*. Thus, in a UFD, factorisations into irreducibles exist and are unique up to reorderings and associates.

We start with the good news, which is that when $R = \mathcal{O}_K$ factorisation into irreducibles does always exist.

Lemma 3.3. *Let $\mathcal{O}_K$ be the ring of integers of a number field. Then every $x \in \mathcal{O}_K$ may be written, in at least one way, as a product of irreducibles.*

Proof. We proceed by induction on the absolute value of the norm $|N_{K/\mathbf{Q}}(x)|$ which, by Lemma 2.8, is a natural number. If x is itself irreducible, we are done. Otherwise, we have $x = yz$ with neither y nor z a unit. Taking norms, we have $N_{K/\mathbf{Q}}(x) = N_{K/\mathbf{Q}}(y)N_{K/\mathbf{Q}}(z)$. Since neither y nor z is a unit, $N_{K/\mathbf{Q}}(y), N_{K/\mathbf{Q}}(z) \neq \pm 1$. (Here we used Lemma 2.10.) It follows that $|N_{K/\mathbf{Q}}(y)|, |N_{K/\mathbf{Q}}(z)| < N_{K/\mathbf{Q}}(x)$, and so by induction y, z admit decompositions into irreducibles. Hence so does x. $\square$

Remark. This lemma holds in any commutative noetherian ring, a concept you may wish to read up on.

There is more good news: the rings of integers in many small number fields such as $\mathbf{Q}(i), \mathbf{Q}(\sqrt{2})$ and $\mathbf{Q}(\sqrt{-2})$ are UFDs. These facts were (probably) proven in *Rings and Modules* by showing that these fields are *Euclidean domains*. We will not be saying very much about Euclidean domains in this course. However, the fact that these examples are UFDs may also be proven using the techniques we develop in this course. We do this explicitly for $\mathbf{Q}(i)$ in Section 11.

Failure in $\mathbf{Q}[\sqrt{-5}]$. However, there is bad news - it is not hard to come up with an example where $\mathcal{O}_K$ does not admit unique factorisation into irreducibles.

Lemma 3.4. *When $K = \mathbf{Q}(\sqrt{-5})$, $\mathcal{O}_K$ is not a UFD.*

Proof. First note that, by Lemma 2.17, $\mathcal{O}_K = \mathbf{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbf{Z}\}$. Now observe that

$$6 = 2 \times 3 = (1 + \sqrt{-5}) \times (1 - \sqrt{-5}).$$

We claim that $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are all irreducible, and that neither 2 nor 3 are associate to $1 \pm \sqrt{-5}$.

To see this, we use norms. Note that

$$N_{K/\mathbf{Q}}(a+b\sqrt{-5}) = (a+b\sqrt{-5})(a-b\sqrt{-5}) = a^2+5b^2.$$

Thus the possible values of the norm are

$$1, 4, 5, 6, 9, \ldots. \tag{3.1}$$

Note that

$$N_{K/\mathbf{Q}}(2) = 4,\ N_{K/\mathbf{Q}}(3) = 9,\ N_{K/\mathbf{Q}}(1 \pm \sqrt{-5}) = 6.$$

None of these numbers $4, 6, 9$ factors as a product of two smaller numbers in the sequence (3.1), and so $2, 3, 1 \pm \sqrt{-5}$ are all irreducible. Indeed, if we had $2 = xy$ with neither x nor y a unit then, taking norms, we would have $N_{K/\mathbf{Q}}(2) = N_{K/\mathbf{Q}}(x)N_{K/\mathbf{Q}}(y)$, with neither $N_{K/\mathbf{Q}}(x), N_{K/\mathbf{Q}}(y)$ being 1 by Lemma 2.10.

Neither 2 nor 3 is associate to $1 \pm \sqrt{-5}$, because associate elements have the same norm. □

The usefulness of unique factorisation. We will be spending most of the rest of the course discussing unique factorisation. As justification for this, let us see how to use unique factorisation in $\mathbf{Z}[\sqrt{-2}]$ (proven in Rings and Modules, or provable using the techniques we will develop below) to solve the equation $y^2 + 2 = x^3$ mentioned in the introduction.

Theorem 3.5. *The only integer solutions to $y^2+2 = x^3$ are $x = 3$, $y = \pm 5$.*

Proof. Factor the equation as

$$(y+\sqrt{-2})(y-\sqrt{-2}) = x^3. \tag{3.2}$$

We claim that the two factors on the left are coprime (the only integers in $\mathbf{Z}[\sqrt{-2}]$ dividing both of them are units). Suppose, to the contrary, that some irreducible α divides both factors. Then α divides $(y+\sqrt{-2})-(y-\sqrt{-2}) = 2\sqrt{-2} = -(\sqrt{-2})^3$. Now $\sqrt{-2}$ is irreducible in $\mathbf{Z}[\sqrt{-2}]$, since it has norm 2, so if it factors into two elements of $\mathbf{Z}[\sqrt{-2}]$, one of them must have norm 1 and hence be a unit. Therefore, by unique factorisation into irreducibles, α is an associate of $\sqrt{-2}$. Modifying α by a unit, we can assume that $\alpha = \sqrt{-2}$.

Thus $\sqrt{-2}|(y+\sqrt{-2})$, and so $\sqrt{-2}|y$. Taking norms, we see that $2|y^2$, and hence $2|y$. But then, returning to the original equation $y^2+2 = x^3$, we see that $2|x$, and hence $y^2 \equiv 6 \pmod 8$. This is impossible, and so indeed the two factors on the left in (3.2) are coprime.

Using unique factorisation again, it follows that both $y \pm \sqrt{-2}$ are associates of cubes in $\mathbf{Z}[\sqrt{-2}]$. Since the only units in $\mathbf{Z}[\sqrt{-2}]$ are ± 1, and -1 is a cube, both $y \pm \sqrt{-2}$ *are* cubes. Suppose that

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3,$$

where $a, b \in \mathbf{Z}$. Expanding out and comparing coefficients of $\sqrt{-2}$, we obtain

$$1 = b(3a^2 - 2b^2).$$

This is a very easy equation to solve over the integers. We must have either $b = -1$, in which case $3a^2 - 2 = -1$, which is impossible, or else $b = 1$, in which case $3a^2 - 2 = 1$ and so $a = \pm 1$. This leads to $y + \sqrt{-2} = (\pm 1 + \sqrt{-2})^3$ and so $y = \pm 5$. $\square$

We conclude with a historical note. According to [2] and the references linked there, Fermat considered this equation but is not thought to have had a valid proof. Euler gave the argument above, but did not understand the fact that he was using unique factorisation, or what notions such as "coprime" mean. Thus he also did not have a valid proof.

4. Ideals and their basic properties

In the next few chapters we come to the main theme of the course: whilst $\mathcal{O}_K$ is not necessarily a UFD, we may recover a theory of unique factorisation by working in the enlarged world of *ideals*.

The notion of an ideal should be familiar from *Rings and Modules* (we will, however, recall it below).

First a word on notation. In previous iterations of this course in Oxford, capital letters such as I, J, P, Q have been used for ideals in $\mathcal{O}_K$. However, it is rather standard to use *fraktur* letters $\mathfrak{a}, \mathfrak{b}, \mathfrak{p}, \mathfrak{q}$. This is what is done in the recommended book [1], as well as in many (but not all) other sources. We will follow this convention too, both in the course and the exam (this does make things a little trickier at the board). In particular, $\mathfrak{p}$ and $\mathfrak{q}$ will *always* denote prime ideals (we will recall the definition in the next section).

Ideals and principal ideals. Let us first recall the basic definitions, adapted to the notation of this course.

Definition 4.1 (Ideals, principal ideals)**.** An ideal $\mathfrak{a}$ in $\mathcal{O}_K$ is a subset which is a subgroup under addition, and which is closed under multiplication by elements of $\mathcal{O}_K$: if $x \in \mathfrak{a}$ and $\alpha \in \mathcal{O}_K$ then $\alpha x \in \mathfrak{a}$. We will sometimes write $\text{Ideals}(\mathcal{O}_K)$ for the set of ideals in $\mathcal{O}_K$. Given $x \in \mathcal{O}_K$, we may form the *principal ideal*

$$(x) := \{\alpha x : \alpha \in \mathcal{O}_K\}.$$

Given elements $x_1, \ldots, x_r \in \mathcal{O}_K$, the ideal generated by the x_i is

$$(x_1, \ldots, x_r) := \{\alpha_1 x_1 + \cdots + \alpha_r x_r : \alpha_1, \ldots, \alpha_r \in \mathcal{O}_K\}.$$

The map $\iota : \mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$ which associates $x \in \mathcal{O}_K$ to the principal ideal (x) is "an embedding up to units". (More precisely, ι induces an injective map $\mathcal{O}_K/U(\mathcal{O}_K) \to \text{Ideals}(\mathcal{O}_K)$.) Indeed if $(x) = (y)$ then there must be some u, v such that $x = uy$ and $y = xv$, but then $x = xuv$ and so $uv = 1$; conversely, if x and y are associates (differ up to units) then $(x) = (y)$.

Sometimes, ι will be surjective.

Definition 4.2 (PID)**.** If the map $\iota : \mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$ is surjective, that is to say if every ideal is a principal ideal, then $\mathcal{O}_K$ is said to be a principal ideal domain (PID).

You have seen in *Rings and Modules* that a PID is a UFD, not just for rings of integers $\mathcal{O}_K$ but for general integral domains. Indeed, when showing that a Euclidean domain is a UFD, one first shows that it is a PID and then one shows that all PIDs are UFDs.

The converse is not true in general: for instance $\mathbf{Z}[X, Y]$ is a UFD (because a polynomial ring over a UFD is a UFD) but it is not a PID since, for example, the ideal (X, Y) is not principal.

We will show later on that the converse is true in number fields.

Theorem 4.3. *Let $\mathcal{O}_K$ be the ring of integers of a number field. Suppose that $\mathcal{O}_K$ is a UFD. Then $\mathcal{O}_K$ is a PID.*

Proof. See Chapter 6. As we have remarked, this is not true for arbitrary integral domains and so we must rely on properties at least somewhat specific to number fields. □

The picture we have at the moment (not all proven!) is as follows. We have a map $\mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$. This is surjective if and only if $\mathcal{O}_K$ is a

UFD. Our plan is to show that unique factorisation can *always* be recovered by working in the larger world Ideals($\mathcal{O}_K$).

A nonprincipal ideal. Let us pause to check that we are indeed building a nonempty theory, by giving an example of a nonprincipal ideal. But the remarks above, to find such an ideal we need to look in some K where $\mathcal{O}_K$ is not a UFD. We have already discussed such an example, $K = \mathbf{Q}(\sqrt{-5})$.

Lemma 4.4. *Let $K = \mathbf{Q}(\sqrt{-5})$. Then the ideal $\mathfrak{a} = (2, 1+\sqrt{-5})$ generated by 2 and $1+\sqrt{-5}$ is not principal.*

Proof. First note that
$$(2) \subsetneq \mathfrak{a};$$
the inclusion is strict since $\frac{1+\sqrt{-5}}{2} \notin \mathcal{O}_K$. Second, note that
$$\mathfrak{a} \subsetneq (1).$$
Indeed if $1 \in \mathfrak{a}$ then we would have $1 = 2(a+b\sqrt{-5})+(1+\sqrt{-5})(c+d\sqrt{-5})$ for some integers a, b, c, d. Comparing coefficients gives $1 = 2a + c - 5d$, so $c + d \equiv 1 (\mathrm{mod}\, 2)$, and $2b + c + d = 0$, so $c + d \equiv 0 (\mathrm{mod}\, 2)$. This is a contradiction.

It follows that if $\mathfrak{a} = (\alpha)$ were principal then $1 < N_{K/\mathbf{Q}}(\alpha) < 4$ (in fact, that $N_{K/\mathbf{Q}}(\alpha) = 2$). However, recalling that $N_{K/\mathbf{Q}}(a + b\sqrt{-5}) = a^2 + 5b^2$, we see that there is no such element. □

Basic properties of ideals. Let us record some simple properties of ideals, somewhat specific to the number field case.

Lemma 4.5. *Let $\mathfrak{a}$ be a non-zero ideal in $\mathcal{O}_K$. Then $\mathfrak{a}$ contains a non-zero rational integer a, and thus the principal ideal (a) is contained in $\mathfrak{a}$.*

Proof. Let $\alpha \in \mathfrak{a}$ be some nonzero element. Since $\alpha \in \mathcal{O}_K$, it is an algebraic integer and therefore satisfies some equation $\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_0 = 0$ with $c_0, \dots, c_{n-1} \in \mathbf{Z}$, and with $c_0 \neq 0$ (otherwise divide through by α). Rearranging gives $c_0 = -\alpha(c_1 + \cdots + c_{n-1}\alpha^{n-2} + \alpha^{n-1})$, and therefore c_0 is a multiple of α, and hence lies in $\mathfrak{a}$. □

Lemma 4.6. *Let $\mathfrak{a}$ be a nonzero ideal. Then the quotient ring $\mathcal{O}_K/\mathfrak{a}$ is finite.*

Proof. First note that if $\mathfrak{b} \subseteq \mathfrak{a}$ then there is a natural surjective map from $\mathcal{O}_K/\mathfrak{b}$ to $\mathcal{O}_K/\mathfrak{a}$. Therefore it suffices to prove the statement for any nonzero ideal $\mathfrak{b}$ contained in $\mathfrak{a}$. By Lemma 4.5, it suffices to prove that $\mathcal{O}_K/(a)$ is finite, for any non-zero rational integer a. Switching a to $-a$ if necessary, we may assume $a > 0$. Let $e_1, \ldots, e_n$ be an integral basis for $\mathcal{O}_K$. Then

$$(a) = \{m_1 e_1 + \cdots + m_n e_n | m_i \in \mathbf{Z}, a | m_i\}.$$

Therefore the quotient $\mathcal{O}_K/(a)$ is isomorphic to $(\mathbf{Z}/a\mathbf{Z})^n$, which is clearly finite. □

In particular (forgetting the ideal structure), $\mathfrak{a}$ is a finite-index $\mathbf{Z}$-submodule of $\mathcal{O}_K$.

Norms. Integral basis for an ideal.

Definition 4.7 (Norm of an ideal)**.** Let $\mathfrak{a}$ be a nonzero ideal in $\mathcal{O}_K$. Then we define the norm $N(\mathfrak{a})$ to be $|\mathcal{O}_K/\mathfrak{a}|$.

It follows from Lemma 4.6 that $N(\mathfrak{a})$ is finite, provided $\mathfrak{a} \neq \{0\}$.

As we have seen, $\mathcal{O}_K$ is a free abelian group of rank n (that is, it has an integral basis). It is a general fact (see Appendix A) that any finite index subgroup of a free abelian group of rank n is also free abelian of rank n. Thus $\mathfrak{a}$ is free abelian of rank n, or in other words $\mathfrak{a}$ has an integral basis, that is to say

$$\mathfrak{a} = \bigoplus_{i=1}^{n} \mathbf{Z} e_i'$$

for some $e_i' \in \mathcal{O}_K$. This could also be proven by mimicing the proof that $\mathcal{O}_K$ has an integral basis (Theorem 2.12).

Moreover, the following is a consequence of Proposition A.1.

Lemma 4.8. *Suppose that $e_1, \ldots, e_n$ is an integral basis for $\mathcal{O}_K$. Let $\mathfrak{a}$ be an ideal with integral basis $e_1', \ldots, e_n'$, and suppose that $e_i' = \sum_j A_{ji} e_j$ for some matrix A. Then $N(\mathfrak{a}) = |\det A|$.*

In the course of the proof of Lemma 4.6, we showed that if a is a positive rational integer then $\mathcal{O}_K/(a) \cong (\mathbf{Z}/a\mathbf{Z})^n$, and so $N((a)) = a^n$ (where n is the degree of K). We also have $N_{K/\mathbf{Q}}(a) = a^n$, and so $N((a)) = N_{K/\mathbf{Q}}(a)$ for $a \in \mathbf{Z} \setminus \{0\}$. In fact this generalises to all principal ideals.

Lemma 4.9. *Suppose that $\mathfrak{a} = (\alpha)$ is a principal ideal, for some $\alpha \in \mathcal{O}_K \setminus \{0\}$. Then $N(\mathfrak{a}) = |N_{K/\mathbf{Q}}(\alpha)|$.*

Proof. Let $e_1, \dots, e_n$ be an integral basis for $\mathcal{O}_K$. Then an integral basis for (α) is $e'_1, \dots, e'_n$, where $e'_i = \alpha e_i$. We have already seen, in Lemma 1.18, that if A is the matrix of the multiplication-by-α map, that is if $e'_i = \sum_j A_{ji} e_j$, then $\det A = N_{K/\mathbf{Q}}(\alpha)$. The result follows immediately from Lemma 4.8. $\square$

In other words, the absolute value of the norm respects the embedding $\mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$, and generalises the notion of (absolute value of) norm on $\mathcal{O}_K$ to a notion on $\text{Ideals}(\mathcal{O}_K)$.

Multiplying ideals. Prime ideals. Our next task is to embed the multiplicative structure of $\mathcal{O}_K$ into a multiplicative structure on $\text{Ideals}(\mathcal{O}_K)$ by defining the notion of the product of two ideals.

Definition 4.10. Let $\mathfrak{a}, \mathfrak{b}$ be ideals in $\mathcal{O}_K$. Then we define the product $\mathfrak{a}\mathfrak{b}$ to consist of all finite sums $\sum_{i=1}^k a_i b_i$ with $a_i \in \mathfrak{a}$ and $b_i \in \mathfrak{b}$.

We leave it as an exercise to check that $\mathfrak{a}\mathfrak{b}$ is an ideal. Since $\mathcal{O}_K$ is commutative, the product operation on ideals is commutative too. It is very important to note that the definition of product does *not* say that $\mathfrak{a}\mathfrak{b}$ consists of the products ab with $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$; one would not expect that to be closed under addition. Observe also that

$$\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a}, \mathfrak{b}.$$

Also, $\mathcal{O}_K = (1)$ is itself an ideal and

$$\mathfrak{a} \cdot (1) = \mathfrak{a}.$$

If $\mathfrak{a} = (x)$ and $\mathfrak{b} = (y)$ with $x, y \in \mathbf{Z}$ then $\mathfrak{a}\mathfrak{b} = (xy)$. In particular, the embedding (up to units) of $\mathcal{O}_K$ in $\text{Ideals}(\mathcal{O}_K)$ respects this multiplicative structure.

Remark. Though it is possible to define the *sum* of two ideals $\mathfrak{a} + \mathfrak{b} = \{a + b : a \in \mathfrak{a}, b \in \mathfrak{b}\}$, this does *not* respect the additive structure on $\mathcal{O}_K$ under the map $\mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$. (For instance, if $\mathfrak{a} = (1) = \mathfrak{b}$ then $\mathfrak{a} + \mathfrak{b} = (1) \neq (1 + 1) = (2)$).

Now we have a notion of multiplication of ideals, it is very simple to give a definition of divisor.

Definition 4.11. Let $\mathfrak{a}, \mathfrak{b}$ be two ideals in $\mathcal{O}_K$. Then we say that $\mathfrak{b} \mid \mathfrak{a}$ if there is an ideal $\mathfrak{c}$ such that $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$.

Note that if $\mathfrak{b} \mid \mathfrak{a}$ then $\mathfrak{a} \subseteq \mathfrak{b}$. That is, *division implies containment.* Remarkably, the converse is also true, but much harder to prove (Theorem 5.2). However, we strongly suggest the reader keep this fact in mind when reading what follows.

Prime ideals. The notion of prime ideal is the standard one from ring theory, specialised to the setting of number fields.

Definition 4.12. An ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is prime if it is not $\mathcal{O}_K = (1)$, and if $xy \in \mathfrak{p}$ implies that either x or y lies in $\mathfrak{p}$.

Let us record the following equivalent description of prime ideal.

Lemma 4.13. *An ideal $\mathfrak{p}$ is prime if and only if the following is true: whenever $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{p}$, either $\mathfrak{a} \subseteq \mathfrak{p}$ or $\mathfrak{b} \subseteq \mathfrak{p}$.*

Proof. Suppose first that $\mathfrak{p}$ is prime, that $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{p}$, and that $\mathfrak{a}$ is not contained in $\mathfrak{p}$. Let $x \in \mathfrak{a} \setminus \mathfrak{p}$, and let $y \in \mathfrak{b}$ be arbitrary.

Then $xy \in \mathfrak{a}\mathfrak{b} \subseteq \mathfrak{p}$ and hence $xy \in \mathfrak{p}$. But $\mathfrak{p}$ is prime, so either x or y lies in $\mathfrak{p}$. Since $x \notin \mathfrak{p}$ we must have $y \in \mathfrak{p}$. Therefore $\mathfrak{b} \subseteq \mathfrak{p}$.

Conversely, suppose that $\mathfrak{p}$ is not prime, and find $x, y \notin \mathfrak{p}$ with $xy \in \mathfrak{p}$. Then if we take $\mathfrak{a} = (x)$ and $\mathfrak{b} = (y)$ we see that $\mathfrak{a}\mathfrak{b} = (xy) \subset \mathfrak{p}$, but neither $\mathfrak{a}$ nor $\mathfrak{b}$ is contained in $\mathfrak{p}$. □

In number fields, we do not introduce the notion of maximal ideal, since in $\mathcal{O}_K$ all prime ideals are maximal. Let us recall from *Rings and Modules* that the quotient R/I is an integral domain (resp. field) if I is prime (resp. maximal).

Lemma 4.14. *In $\mathcal{O}_K$, all prime ideals are maximal. In particular, if $\mathfrak{p}$ and $\mathfrak{q}$ are two prime ideals with $\mathfrak{p} \subseteq \mathfrak{q}$, then $\mathfrak{p} = \mathfrak{q}$.*

Proof. If $\mathfrak{p}$ is prime then $\mathcal{O}_K/\mathfrak{p}$ is an integral domain. It is also finite, by Lemma 4.6. However, all finite integral domains are fields, since any nonzero element x has $x^n = 1$ for some n, which means that x^{n-1} is an inverse for x. Therefore $\mathcal{O}_K/\mathfrak{p}$ is a field, which is equivalent to $\mathfrak{p}$ being maximal. □

5. Unique factorisation into prime ideals

The main theorem of this chapter, and one of the main theorems of the course, is the following.

Theorem 5.1. *Let K be a number field with ring of integers $\mathcal{O}_K$. Then any non-zero proper ideal $\mathfrak{a}$ admits a unique factorisation $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_k$ into prime ideals.*

Remark. This statement is actually cleaner than the statement of unique factorisation over the integers, because there is no ambiguity up to multiplication by units. Indeed if x and y are associates then the ideals (x) and (y) are *the same*.

During the proof of Theorem 5.1, we will establish two facts of independent interest. First, we will prove that containment of ideals is equivalent to division:

Proposition 5.2. *Suppose that $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero ideals in $\mathcal{O}_K$. Then $\mathfrak{a} \subseteq \mathfrak{b}$ if and only if $\mathfrak{b} \mid \mathfrak{a}$.*

Second, we will show that prime ideals behave like prime numbers in the following sense.

Lemma 5.3. *Let $\mathfrak{p}$ be a prime ideal, and suppose that $\mathfrak{p} \mid \mathfrak{a}\mathfrak{b}$. Then $\mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$.*

Once these results are proven, one can easily establish analogues of facts familiar from elementary number theory. For instance, we can say that two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are *coprime* if there is no prime ideal $\mathfrak{p}$ dividing both of them. Using unique factorisation one may then show that if $\mathfrak{a}$ and $\mathfrak{b}$ are coprime ideals dividing a third ideal $\mathfrak{c}$, then $\mathfrak{a}\mathfrak{b} \mid \mathfrak{c}$.

Prime factors. We turn now to the proof of Theorem 5.1, starting with some basic preliminary facts.

Lemma 5.4. *Let $\mathfrak{a}$ be a proper ideal in $\mathcal{O}_K$. Then there is a prime ideal $\mathfrak{p}$ with $\mathfrak{a} \subseteq \mathfrak{p}$.*

Proof. If $\mathfrak{a}$ is maximal, then it is itself prime. Otherwise, find an ideal $\mathfrak{b}$ with $\mathfrak{a} \subsetneq \mathfrak{b} \subsetneq \mathcal{O}_K$. Note that $N(\mathfrak{b}) = |\mathcal{O}_K/\mathfrak{b}| < |\mathcal{O}_K/\mathfrak{a}| = N(\mathfrak{a})$. Thus this process can only continue for finitely many steps before we reach a maximal (and hence prime) ideal. $\square$

Remark. In fact, in any ring with 1, every ideal is contained in a maximal (and hence prime) ideal; this is a standard application of Zorn's lemma (and hence relies on the axiom of choice). The proof of Lemma 5.4 uses the fact that the index of nonzero ideals in $\mathcal{O}_K$ is finite to give a more down-to-earth proof in this case.

Lemma 5.5. *Let $\mathfrak{a}$ be a nonzero ideal in $\mathcal{O}_K$. Then there are prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_k$ such that $\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \mathfrak{a}$.*

Proof. Suppose the result is false. Then there is a counterexample $\mathfrak{a}$ with minimal norm. Clearly $\mathfrak{a}$ is not itself prime, and therefore we may find $x, y \in \mathcal{O}_K$ with $xy \in \mathfrak{a}$ but $x, y \notin \mathfrak{a}$. The ideals $\mathfrak{a}' := \mathfrak{a} + (x)$ and $\mathfrak{a}'' := \mathfrak{a} + (y)$ strictly contain $\mathfrak{a}$. It is immediate from the definition of norm that $N(\mathfrak{a}'), N(\mathfrak{a}'') < N(\mathfrak{a})$, and hence by minimality we have

$$\mathfrak{p}'_1 \cdots \mathfrak{p}'_{k'} \subseteq \mathfrak{a}',$$
$$\mathfrak{p}''_1 \cdots \mathfrak{p}''_{k''} \subseteq \mathfrak{a}''.$$

Finally, observe that $\mathfrak{a}'\mathfrak{a}'' \subset \mathfrak{a}$, since $\mathfrak{a}$ is an ideal and $xy \in \mathfrak{a}$. $\square$

Remark. What we are really using is the fact that $\mathcal{O}_K$ is noetherian, that is to say there is no infinite ascending chain of ideals. This property follows immediately from the fact that nonzero ideals have finite index, which is (of course) the main ingredient in the proof of Lemma 5.5.

Finding an inverse. The key ingredient in the proof of Theorem 5.1 is the following, which is a far less obvious result than the ones we have established so far.

Proposition 5.6. *Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$. Then there is an ideal $\mathfrak{b}$ such that $\mathfrak{a}\mathfrak{b}$ is principal.*

Remarks. The title of the section comes from the fact that $\mathfrak{b}$ is indeed an inverse to $\mathfrak{a}$ in the ideal class group, which we shall introduce later.

Before proving Proposition 5.6, we assemble some lemmas. Here is the first of them.

Lemma 5.7. *Suppose that $\mathfrak{a}$ is a nonzero proper ideal (thus it is not all of $\mathcal{O}_K$). Then there is some $\theta \in K \setminus \mathcal{O}_K$ such that $\theta\mathfrak{a} \subseteq \mathcal{O}_K$.*

Proof. Let x be a nonzero element of $\mathfrak{a}$. Thus $(x) \subseteq \mathfrak{a}$. By Lemma 5.5 there are prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r \subseteq (x).$$

Assume that r is minimal with this property.

By Lemma 5.4 there is a prime ideal $\mathfrak{p}$ such that $\mathfrak{a} \subseteq \mathfrak{p}$. Thus, putting everything together,

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r \subseteq (x) \subseteq \mathfrak{a} \subseteq \mathfrak{p}, \tag{5.1}$$

so $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subseteq \mathfrak{p}$.

Since $\mathfrak{p}$ is prime, by Lemma 4.13 there is some i, without loss of generality $i = 1$, such that $\mathfrak{p}_1 \subseteq \mathfrak{p}$. Since prime ideals are maximal (specifically, by Lemma 4.14) we in fact have $\mathfrak{p} = \mathfrak{p}_1$, and so by (5.1)

$$\mathfrak{a} \subseteq \mathfrak{p}_1. \tag{5.2}$$

Now by the minimality of r, we do *not* have $\mathfrak{p}_2 \cdots \mathfrak{p}_r \subseteq (x)$. Let $y \in \mathfrak{p}_2 \cdots \mathfrak{p}_r \setminus (x)$. Take $\theta := \frac{y}{x}$. Then $\theta \in K \setminus \mathcal{O}_K$.

Finally, note that

$$\begin{aligned}
\theta\mathfrak{a} &= \frac{y}{x}\mathfrak{a} \\
&\subseteq \frac{1}{x}\mathfrak{p}_2 \cdots \mathfrak{p}_k \mathfrak{a} && \text{since } y \in \mathfrak{p}_2 \cdots \mathfrak{p}_k \\
&\subseteq \frac{1}{x}\mathfrak{p}_1 \cdots \mathfrak{p}_k && \text{since } \mathfrak{a} \subseteq \mathfrak{p}_1, \text{ by (5.2)} \\
&\subseteq \frac{1}{x}(x) && \text{since } \mathfrak{p}_1 \cdots \mathfrak{p}_r \subseteq (x), \text{ by (5.1)} \\
&= \mathcal{O}_K.
\end{aligned}$$

This concludes the proof. □

Here is the second preparatory lemma for the proof of Proposition 5.6.

Lemma 5.8. *Suppose that $\mathfrak{a}$ is an ideal in $\mathcal{O}_K$, and that $\theta \in K$ is such that $\theta\mathfrak{a} \subseteq \mathfrak{a}$. Then $\theta \in \mathcal{O}_K$.*

Proof. This is a special case of Lemma 2.3, since $\mathfrak{a}$ is a $\mathbf{Z}$-module. (Recall the proof: Let $e_1, \dots, e_n$ be an integral basis for $\mathfrak{a}$. Certainly $\theta e_i \in \mathfrak{a}$ for all i, and so for some integer matrix A we have $\theta e_i = \sum_j A_{ji} e_j$, for all i. Thus the column vector $(e_1, \dots, e_n)^T$ lies in the kernel of $A - \theta I$, which is therefore singular, and so $\det(A - \theta I) = 0$. This is a monic polynomial with integer coefficients, satisfied by θ.) □

With these two preparatory lemmas in hand, we may prove Proposition 5.6 itself. In fact we will show more: that for any nonzero $x \in \mathfrak{a}$ there is an ideal $\mathfrak{b}$ such that $\mathfrak{a}\mathfrak{b} = (x)$.

Define
$$\mathfrak{b} := \{y \in \mathcal{O}_K : y\mathfrak{a} \subseteq (x)\}.$$
That is, $\mathfrak{b}$ is the biggest ideal for which $\mathfrak{a}\mathfrak{b} \subseteq (x)$. To complete the proof we need to show that $\mathfrak{a}\mathfrak{b}$ is not *properly* contained in (x).

Define $\mathfrak{c} := \frac{1}{x}\mathfrak{a}\mathfrak{b}$. Then $\mathfrak{c}$ is an ideal in $\mathcal{O}_K$, and we want to show that $\mathfrak{c}$ is in fact all of $\mathcal{O}_K$. Suppose, as a hypothesis for contradiction, that this is not the case. By our first preparatory lemma, Lemma 5.7, there is some $\theta \in K \setminus \mathcal{O}_K$ such that $\theta\mathfrak{c} \subseteq \mathcal{O}_K$. Since $x \in \mathfrak{a}$, $\mathfrak{b} = \frac{1}{x}(x)\mathfrak{b} \subset \frac{1}{x}\mathfrak{a}\mathfrak{b} = \mathfrak{c}$, that is to say $\mathfrak{b} \subseteq \mathfrak{c}$. Therefore $\theta\mathfrak{b} \subseteq \mathcal{O}_K$.

Also, $\theta\mathfrak{b}\mathfrak{a} = \theta\mathfrak{c}(x) \subseteq \mathcal{O}_K(x) = (x)$. It therefore follows from the definition of $\mathfrak{b}$ that $\theta\mathfrak{b} \subseteq \mathfrak{b}$.

From Lemma 5.8, θ is an algebraic integer. This is a contradiction, since $\theta \in K \setminus \mathcal{O}_K$. This concludes the proof.

Cancellation, divisibility and prime ideals. The proof of Proposition 5.6 was quite involved. However, now we have it in hand, we can reach a number of pleasant consequences quite quickly.

Corollary 5.9 (Cancellation)**.** *Suppose that $\mathfrak{a}\mathfrak{c} = \mathfrak{a}\mathfrak{c}'$. Then $\mathfrak{c} = \mathfrak{c}'$.*

Proof. By Proposition 5.6 there is $\mathfrak{b}$ such that $\mathfrak{a}\mathfrak{b} = (x)$ is principal. Multiplying through by $\mathfrak{b}$, we see that $\mathfrak{c}(x) = \mathfrak{c}'(x)$, and then it is clear that $\mathfrak{c} = \mathfrak{c}'$. □

Proposition 5.2 is also a quick corollary. We recall the statement.

Proposition 5.2. *Suppose that $\mathfrak{a} \subseteq \mathfrak{b}$. Then there is some $\mathfrak{c}$ such that $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$. In other words, $\mathfrak{b}|\mathfrak{a}$ if and only if $\mathfrak{a} \subseteq \mathfrak{b}$.*

Proof. By Proposition 5.6 there is $\mathfrak{d}$ so that $\mathfrak{b}\mathfrak{d} = (x)$ is principal. Multiplying the hypothesis through by $\mathfrak{d}$ gives $\mathfrak{a}\mathfrak{d} \subseteq \mathfrak{b}\mathfrak{d} = (x)$. Let $\mathfrak{c} = \frac{1}{x}\mathfrak{d}\mathfrak{a}$, which is an ideal in $\mathcal{O}_K$. Then $\mathfrak{b}\mathfrak{c} = \frac{1}{x}\mathfrak{b}\mathfrak{d}\mathfrak{a} = \frac{1}{x}(x)\mathfrak{a} = \mathfrak{a}$. □

Recall Lemma 4.13: this stated that if $\mathfrak{p}$ is a prime ideal and $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{p}$ then either $\mathfrak{a} \subseteq \mathfrak{p}$ or $\mathfrak{b} \subseteq \mathfrak{p}$. In the light of Proposition 5.2, this may be rephrased in the following much more suggestive form.

Lemma 5.10. *Let $\mathfrak{p}$ be a prime ideal, and suppose that $\mathfrak{p} \mid \mathfrak{a}\mathfrak{b}$. Then $\mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$.*

As we shall shortly see, Lemma 5.10 implies unique factorisation into prime ideals quite easily.

Proof of unique factorsation. We may now proceed to the proof of unique factorisation, which is quite straightforward now that we have prepared the ground. Let us recall the statement.

Theorem 5.1. *Let K be a number field with ring of integers $\mathcal{O}_K$. Then any non-zero proper ideal $\mathfrak{a}$ admits a unique factorisation $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_k$ into prime ideals.*

Proof. We first show existence of *some* factorisation into prime ideals. This we do by induction on $N(\mathfrak{a})$. We know from Lemma 5.4 that there is some prime ideal $\mathfrak{p}$ with $\mathfrak{a} \subseteq \mathfrak{p}$ or, (as we now know) $\mathfrak{p} \mid \mathfrak{a}$. Let $\mathfrak{b}$ be such that $\mathfrak{a} = \mathfrak{p}\mathfrak{b}$. Then $\mathfrak{a} \subseteq \mathfrak{b}$. Moreover, $\mathfrak{a}$ is a *proper* subset of $\mathfrak{b}$, since if not we would have $\mathfrak{b}\mathfrak{p} = \mathfrak{b}$ which, by cancellation, would imply $\mathfrak{p} = \mathcal{O}_K$. It follows that $N(\mathfrak{b}) < N(\mathfrak{a})$, and so by induction $\mathfrak{b}$ is a product of primes. (Once again, what we are really using here is the fact that $\mathcal{O}_K$ is noetherian, that is to say has no infinite ascending chain of ideals.)

To prove uniqueness, we use Lemma 5.10 repeatedly, in a manner entirely analogous to the proof of unique factorisation in $\mathbf{Z}$. Suppose that

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k = \mathfrak{q}_1 \cdots \mathfrak{q}_m.$$

Then, by Lemma 5.10, $\mathfrak{p}_1$ divides some $\mathfrak{q}_i$, say $\mathfrak{p}_1 \mid \mathfrak{q}_1$. Thus $\mathfrak{q}_1 \subseteq \mathfrak{p}_1$, which means, by Lemma 4.14, that $\mathfrak{p}_1 = \mathfrak{q}_1$.

Applying the cancellation property, Corollary 5.9, we see that

$$\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_m.$$

One may now proceed inductively. $\square$

Students interested in some further reading may want to look up the concept of *Dedekind domain*, which is the "correct" general context for proving unique factorisation into prime ideals.

Finding the prime ideals. A key feature of ideal arithmetic is that the prime ideals in a number field may be found 'above' the ideals corresponding to rational primes (that is, integers that are prime).

Proposition 5.11. *Every prime ideal $\mathfrak{p}$ occurs as the prime factor of a unique (p), where p is some rational prime.*

Proof. By Lemma 4.5, $\mathfrak{p}$ contains some rational integer m. Thus $(m) \subseteq \mathfrak{p}$, that is to say $\mathfrak{p} \mid (m)$. Factoring m into (rational) primes p_i and using Lemma 5.10 repeatedly, we then see that $\mathfrak{p} \mid (p_i)$ for some i.

For uniqueness, note that if $\mathfrak{p} \mid (p_1), (p_2)$ with $p_1 \neq p_2$ then $p_1, p_2 \in \mathfrak{p}$. However, by the Euclidean algorithm there are $a, b \in \mathbf{Z}$ such that $ap_1 + bp_2 = 1$ and hence $1 \in \mathfrak{p}$, which means that $\mathfrak{p} = \mathcal{O}_K$. This, of course, is not the case. □

If $\mathfrak{p}$ divides (p) then we say that $\mathfrak{p}$ "lies above" p.

The important thing to note is that (p) is not generally a prime ideal, even if p is a (rational) prime. For instance, in $\mathbf{Q}(i)$ we have $(5) = (2-i)(2+i)$, so 5 *splits* in $\mathbf{Q}(i)$. We will study splitting in much greater depth later on.

6. Irreducibles and factorisation, revisited

In this brief chapter we prove Theorem 4.3: that is, if $\mathcal{O}_K$ is a UFD, then it is a PID. Recall that this fails for general rings (for example $\mathbf{Q}[X, Y]$) and so we must use some specific properties of $\mathcal{O}_K$. The key fact we will use is Lemma 5.10: if $\mathfrak{p}$ is a prime ideal in $\mathcal{O}_K$, and if $\mathfrak{p} \mid \mathfrak{a}\mathfrak{b}$, then $\mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$.

Irreducibles and primes. Most of this material is in *Rings and Modules* but there is certainly no harm in refreshing our memory.

Let R be an integral domain (such as $\mathcal{O}_K$). Recall that $x \in R$ is *prime* if $x \mid yz$ implies that $x \mid y$ or $x \mid z$.

Lemma 6.1. *Primes are always irreducible.*

Proof. Suppose that x is prime and that $x = ab$. Then either $x \mid a$ or $x \mid b$, without loss of generality the former. Then $a = xv$ for some v. Thus $x = (xv)b$ and so $1 = vb$, which means that b is a unit. □

The converse is *not* true: irreducibles need not be prime. However, this is true when R is a UFD. (In fact, this characterises UFDs, but we do not need this fact here.)

Lemma 6.2. *Let R be a UFD. The all irreducibles $x \in R$ are prime.*

Proof. Suppose x is irreducible and that $x \mid yz$. Then $xv = yz$ for some v. Factor v, y, z into irreducibles, obtaining $xv_1 \cdots v_n = y_1 \cdots y_k z_1 \cdots z_m$. By uniqueness of this factorisation, x must be one of the y_i (say) up to a unit, which means that $x \mid y$. □

The notion of a prime in $\mathcal{O}_K$ behaves well under the map $\mathcal{O}_K \to \text{Ideals}(\mathcal{O}_K)$. This is almost a tautology:

Lemma 6.3. *Let $x \in \mathcal{O}_K$ be prime. Then the principal ideal $(x) \in \mathrm{Ideals}(\mathcal{O}_K)$ is prime. Conversely, suppose the principal ideal (x) is prime; then x is prime.*

Proof. Suppose that $x \in \mathcal{O}_K$ is a prime element. Suppose that $yz \in (x)$. Then $x \mid yz$, and so either $x \mid y$ or $x \mid z$, which means that either $y \in (x)$ or $z \in (x)$. Thus (x) is a prime ideal.

Conversely suppose that (x) is a prime ideal. Suppose that $x \mid yz$. Then $yz \in (x)$, which means that either $y \in (x)$ or $z \in (x)$, and so either $x \mid y$ or $x \mid z$. Thus x is a prime element. □

UFDs and PIDs. We can now prove Theorem 4.3, that is to say if $\mathcal{O}_K$ is a UFD then it is also a PID.

Every ideal can be factored into prime ideals. Therefore it is enough to show that if $\mathcal{O}_K$ is a UFD then all prime ideals $\mathfrak{p}$ in $\mathcal{O}_K$ are principal.

Let $\mathfrak{p}$ be a prime ideal. Let $\alpha \in \mathfrak{p}$, so that $\mathfrak{p} \mid (\alpha)$. Let $\alpha = \alpha_1 \cdots \alpha_k$ be the (essentially unique) factorisation of α into irreducibles in $\mathcal{O}_K$. By Lemma 6.2, the α_i are all primes in $\mathcal{O}_K$. By Lemma 6.3, all of the (α_i) are prime ideals.

Therefore the factorisation of (α) into prime ideals is $(\alpha_1) \cdots (\alpha_k)$. Since $\mathfrak{p} \mid (\alpha)$, it follows from Lemma 5.10 that $\mathfrak{p}$ *is* one of the (α_i), and therefore it is principal. This concludes the proof.

7. More on norms of ideals

So far, we have made very limited use of the concept of the norm of an ideal. We have used the fact that $|\mathcal{O}_K/\mathfrak{a}|$ is finite to avoid Zorn's lemma (in the proof of Lemma 5.4) and (essentially) to prove that $\mathcal{O}_K$ is noetherian (in the proof of Lemma 5.5, and again in final part of the proof of Theorem 5.1 itself).

Now that we have Theorem 5.1 in hand, we can revisit the notion of norm of an ideal and establish some important further facts about it.

Norm of a product. The main result of this section is the following very useful fact.

Proposition 7.1. *For any two ideals $\mathfrak{a}$ and $\mathfrak{b}$ we have $N(\mathfrak{a}\mathfrak{b}) = N(\mathfrak{a})N(\mathfrak{b})$.*

We say that two ideals $\mathfrak{a}$ and $\mathfrak{b}$ are *coprime* if they do not have any prime (ideal) factors in common.

Lemma 7.2. *If $\mathfrak{a}$ and $\mathfrak{b}$ are coprime then $\mathfrak{a} \cap \mathfrak{b} = \mathfrak{a}\mathfrak{b}$.*

Proof. It is always the case that $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b}$, thus $\mathfrak{a} \cap \mathfrak{b} \mid \mathfrak{a}\mathfrak{b}$. In the other direction, note that $\mathfrak{a} \cap \mathfrak{b} \subseteq \mathfrak{a}$ and so $\mathfrak{a} \mid \mathfrak{a} \cap \mathfrak{b}$. Similarly $\mathfrak{b} \mid \mathfrak{a} \cap \mathfrak{b}$. Thus, since $\mathfrak{a}, \mathfrak{b}$ do not share any prime factors, $\mathfrak{a}\mathfrak{b} \mid \mathfrak{a} \cap \mathfrak{b}$. The result follows. $\square$

Proposition 7.1 in the coprime case is now an immediate consequence of the Chinese remainder theorem and the definition of norm:

$$N(\mathfrak{a}\mathfrak{b}) = |\mathcal{O}_K/\mathfrak{a}\mathfrak{b}| = |\mathcal{O}_K/(\mathfrak{a} \cap \mathfrak{b})| = |(\mathcal{O}_K/\mathfrak{a}) \oplus (\mathcal{O}_K/\mathfrak{b})| = N(\mathfrak{a})N(\mathfrak{b}).$$

By factoring into prime ideals, Proposition 7.1 is therefore a consequence of the special case in which $\mathfrak{a}, \mathfrak{b}$ are prime powers, that is to say the following.

Lemma 7.3. *Let $\mathfrak{p}$ be a prime ideal and t an integer. Then $N(\mathfrak{p}^t) = N(\mathfrak{p})^t$.*

We isolate a lemma from the proof.

Lemma 7.4. *Let $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$, and let i be an integer. Then $|\mathfrak{p}^i/\mathfrak{p}^{i+1}| = N(\mathfrak{p})$.*

Remark. Here, when writing the quotient $\mathfrak{p}^i/\mathfrak{p}^{i+1}$, we are ignoring the ideal structure and taking the quotient as abelian groups.

Proof. By the cancellation lemma for ideals, $\mathfrak{p}^{i+1}$ is strictly contained in $\mathfrak{p}^i$. Therefore we may pick some $\alpha \in \mathfrak{p}^i \setminus \mathfrak{p}^{i+1}$. Note that

$$\mathfrak{p}^{i+1} \subsetneq (\alpha) + \mathfrak{p}^{i+1} \subseteq \mathfrak{p}^i.$$

By unique factorisation of prime ideals, we can only have

$$(\alpha) + \mathfrak{p}^{i+1} = \mathfrak{p}^i. \tag{7.1}$$

Define a homomorphism

$$\pi : \mathcal{O}_K \to \mathfrak{p}^i/\mathfrak{p}^{i+1}$$

by

$$\pi(x) := x\alpha + \mathfrak{p}^{i+1}.$$

By (7.1), π is surjective.

We claim that $\ker \pi = \mathfrak{p}$. Write $(\alpha) = \mathfrak{p}^i\mathfrak{a}$, where $\mathfrak{a}$ is coprime to $\mathfrak{p}$. Now

$$x \in \ker \pi \;\Leftrightarrow\; x\alpha \in \mathfrak{p}^{i+1} \;\Leftrightarrow\; \mathfrak{p}^{i+1} \mid (x)(\alpha) \;\Leftrightarrow\; \mathfrak{p} \mid (x)\mathfrak{a} \;\Leftrightarrow\; \mathfrak{p} \mid (x) \;\Leftrightarrow\; x \in \mathfrak{p}.$$

The claim follows.

Consequently,

$$\mathfrak{p}^i/\mathfrak{p}^{i+1} \cong \mathcal{O}_K/\ker \pi = \mathcal{O}_K/\mathfrak{p},$$

from which Lemma 7.4 is immediate. □

Lemma 7.3 now follows almost immediately by a telescoping product argument:

$$N(\mathfrak{p}^t) = |\mathcal{O}_K/\mathfrak{p}^t| = |\mathcal{O}_K/\mathfrak{p}||\mathfrak{p}/\mathfrak{p}^2|\cdots|\mathfrak{p}^{t-1}/\mathfrak{p}^t| = N(\mathfrak{p})^t.$$

Here, we used the tower law for indices of abelian groups, that is to say $[G_1 : G_2] = [G_1 : G_2][G_2 : G_3]$ if $G_3 \leqslant G_2 \leqslant G_1$.

The following is an immediate (and useful) corollary of Proposition 7.1.

Corollary 7.5. *Let $\mathfrak{a}$ be an ideal for which $N(\mathfrak{a})$ is prime. Then $\mathfrak{a}$ is prime.*

Ideals divide their norms. We have already seen in Lemma 4.5 that every ideal $\mathfrak{a}$ contains some rational integer a, so that $(a) \subseteq \mathfrak{a}$. We now know that this means $\mathfrak{a} \mid (a)$. That is, every ideal divides the ideal generated by some rational integer. (The same result follows from Proposition 5.11 and the fact that $\mathfrak{a}$ factors into primes.)

Here is a more precise version of the same fact, which will be useful when bounding class numbers later on.

Lemma 7.6. *For any ideal $\mathfrak{a}$ we have $\mathfrak{a} \mid (N(\mathfrak{a}))$.*

Proof. Let $m := N(\mathfrak{a})$. By the definition of norm, $|\mathcal{O}_K/\mathfrak{a}| = m$. Therefore the $\times m$ map is trivial on the additive group $\mathcal{O}_K/\mathfrak{a}$, and so in particular $m \in \mathfrak{a}$. This is precisely what it means for $\mathfrak{a}$ to divide (m). □

A corollary of this, and unique factorisation into prime ideals, is there are only finitely many ideals of a given norm.

Automorphisms. *In this section we record a small lemma, Lemma 7.7, which is not really important in the theoretical development but is occasionally useful in computations, as we shall see in the next chapter.

Suppose that K is a number field and that $\sigma = \sigma_i : K \to \mathbf{C}$ is an embedding which fixes $\mathbf{Q}$. That is, $\sigma : K \to K$ is a field automorphism fixing $\mathbf{Q}$. By Lemma 2.7, σ maps $\mathcal{O}_K$ to itself.

Lemma 7.7. *Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$. Then*

(i) *$\mathfrak{a}^\sigma := \{\sigma(x) : x \in \mathfrak{a}\}$ is an ideal;*
(ii) *If $\mathfrak{p}$ is a prime ideal, $\mathfrak{p}^\sigma$ is also prime;*
(iii) *$N(\mathfrak{a}) = N(\mathfrak{a}^\sigma)$.*

Proof. We leave (i) and (ii) as exercises. For (iii), note that there is a bijection $\mathcal{O}_K/\mathfrak{a} \to \mathcal{O}_K/\mathfrak{a}^\sigma$ given by

$$t + \mathfrak{a} \mapsto \sigma(t) + \mathfrak{a}^\sigma,$$

thus

$$N(\mathfrak{a}) = |\mathcal{O}_K/\mathfrak{a}| = |\mathcal{O}_K/\mathfrak{a}^\sigma| = N(\mathfrak{a}^\sigma).$$

This completes the proof. □

8. $\mathbf{Q}(\sqrt{-5})$ REVISITED

At this point, it is extremely instructive to revisit the example given in Chapter 3, which we are now in a position to "explain" in terms of what we know about ideals.

Recall that we were working in $\mathbf{Q}(\sqrt{-5})$, and we observed that

$$6 = 2 \times 3 = (1 + \sqrt{-5}) \times (1 - \sqrt{-5}), \tag{8.1}$$

with all of $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ being irreducible.

Let $\mathfrak{p}_1 = (2, 1 + \sqrt{-5})$, $\mathfrak{p}_2 = (2, 1 - \sqrt{-5})$, $\mathfrak{q}_1 = (3, 1 + \sqrt{-5})$, $\mathfrak{q}_2 = (3, 1 - \sqrt{-5})$.

We claim that $\mathfrak{p}_1\mathfrak{p}_2 = (2)$. To see this, note that (by definition of the product of ideals and the fact that $(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6$) we have $\mathfrak{p}_1\mathfrak{p}_2 = (4, 2 + 2\sqrt{-5}, 2 - 2\sqrt{-5}, 6)$. Clearly all four generators are contained in (2), so $\mathfrak{p}_1\mathfrak{p}_2 \subseteq (2)$. In the other direction, $2 = 6 - 4$ lies in $\mathfrak{p}_1\mathfrak{p}_2$, so $(2) \subseteq \mathfrak{p}_1\mathfrak{p}_2$.

We leave it to the reader to check, in similar fashion, that $\mathfrak{q}_1\mathfrak{q}_2 = (3)$.

There is an automorphism $\sigma : \mathbf{Q}(\sqrt{-5}) \to \mathbf{Q}(\sqrt{-5})$ with $\sigma(\sqrt{-5}) = -\sqrt{-5}$. We have $\mathfrak{p}_2 = \mathfrak{p}_1^\sigma$, and so by Lemma 7.7 we have $N(\mathfrak{p}_1) = N(\mathfrak{p}_2)$. Since $N(\mathfrak{p}_1)N(\mathfrak{p}_2) = N(\mathfrak{p}_1\mathfrak{p}_2) = N((2)) = 4$, it follows that $N(\mathfrak{p}_1) = N(\mathfrak{p}_2) = 2$. As a consequence of Corollary 7.5, both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are prime.

It follows from Lemma 4.9 that neither $\mathfrak{p}_1$ nor $\mathfrak{p}_2$ are principal, since the norm of any element $\alpha = a + b\sqrt{-5}$ is $a^2 + 5b^2$, which does not take the value 2.

Similarly, $N(\mathfrak{q}_1) = N(\mathfrak{q}_2) = 3$, both $\mathfrak{q}_1$ and $\mathfrak{q}_2$ are prime, and neither of them are principal.

Evidently we have

$$6 = 2 \times 3 = (\mathfrak{p}_1\mathfrak{p}_2)(\mathfrak{q}_1\mathfrak{q}_2).$$

By unique factorisation into prime ideals, we must be able to find the other factorisation in (8.1) here too.

To this end, observe that $(1+\sqrt{-5}) \subseteq \mathfrak{p}_1, \mathfrak{q}_1$ and so $\mathfrak{p}_1\mathfrak{q}_1 \mid (1+\sqrt{-5})$ (note that, since $\mathfrak{p}_1, \mathfrak{q}_1$ have different norms, they are different ideals and hence coprime). Since $N(1+\sqrt{-5}) = 6 = N(\mathfrak{p}_1\mathfrak{q}_1)$, we in fact have $\mathfrak{p}_1\mathfrak{q}_1 = (1+\sqrt{-5})$. Similarly, $\mathfrak{p}_2\mathfrak{q}_2 = (1-\sqrt{-5})$.

Hence,

$$6 = (1+\sqrt{-5})(1-\sqrt{-5}) = (\mathfrak{p}_1\mathfrak{q}_1)(\mathfrak{p}_2\mathfrak{q}_2).$$

Finally, we remark (and you should check) that in fact $\mathfrak{p}_1 = \mathfrak{p}_2$, but $\mathfrak{q}_1$ and $\mathfrak{q}_2$ are distinct. (Later, we will introduce some terminology for this: 2 is "ramified" in $\mathbf{Q}(\sqrt{-5})$, but 3 is not.)

9. Factoring into prime ideals in practice

In this chapter we will examine some strategies for factoring ideals into prime ideal factors. We begin with the case of rational prime ideals (p), where there is a useful tool – Dedekind's lemma. At the end of the chapter we indicate a general strategy for reducing to this case.

Splitting of rational primes. Let p be a rational prime. We wish to factor (p) as a product of prime ideals in $\mathcal{O}_K$. (Recall from Section 5 that *all* prime ideals occur this way). Dedekind's lemma, stated in Theorem 9.2 below, is a very useful tool for this problem.

Such a factorisation will, of course, have the form

$$(p) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} \tag{9.1}$$

for *distinct* prime ideals $\mathfrak{p}_i$ and positive integer exponents e_i, called the *ramification index* of $\mathfrak{p}_i$.

Taking norms, we see that each $N(\mathfrak{p}_i)$ must equal some power p^{f_i} of p; the number f_i is called the *degree* of $\mathfrak{p}_i$. Taking norms of both sides of (9.1) yields

$$n = \sum_{i=1}^{r} e_i f_i. \tag{9.2}$$

There are bits of language to describe various extreme situations. For instance,

- If $r = n$ (so all the e_i, f_i are equal to 1), p is said to *split completely* in K.
- If $e_i > 1$ for some i then p is said to *ramify.*

- If $r = 1$ and $e_1 = n$ (so $f_1 = 1$) then p is said to be *totally ramified* in K.
- If $r = 1$ and $e_1 = 1$ (so $f_1 = n$) then p is said to be *inert* in K. In this case (p) is itself a prime ideal.

There are also notions such as *wild* and *tame* ramification, which have to do with the possibility that p divides e_i; these are not relevant in this course.

Irreducibility over $\mathbf{Z}$ and mod p. Let $f(X) \in \mathbf{Z}[X]$, and let $\overline{f}(X) \in \mathbf{F}_p[X]$ be its reduction mod p. If f is reducible, then so is $\overline{f}$. However, the converse is not true: $X^2 + 1$ is irreducible in $\mathbf{Z}[X]$, but factors as $(X+1)^2$ in $\mathbf{F}_2[X]$.

The main tool in the proof of Dedekind's lemma is the following result about this situation. This is perhaps a little subtle and the proof is even less examinable than many of the others in the course.

Lemma 9.1. *Suppose that $\alpha \in \mathcal{O}$ has minimal polynomial $m(X) \in \mathbf{Z}[X]$. Let $\overline{m}(X) \in \mathbf{F}_p[X]$ be the reduction of m mod p (here identifying $\mathbf{Z}/p\mathbf{Z}$ and $\mathbf{F}_p$), and let $\overline{g}(X)$ be any monic irreducible factor of $\overline{m}(X)$. Let $\overline{\alpha}$ be a root of $\overline{g}$ (in the algebraic closure of $\mathbf{F}_p$). Then*

(i) *There is a natural ring homomorphism $\pi : \mathbf{Z}[\alpha] \to \mathbf{F}_p[\overline{\alpha}]$ given by $\pi(f(\alpha)) = \overline{f}(\overline{\alpha})$;*
(ii) $\ker \pi = (p, g(\alpha))$*;*
(iii) $(p, g(\alpha))$ *is a maximal ideal in $\mathbf{Z}[\alpha]$ of index $p^{\deg \overline{g}}$.*
(iv) *If $\overline{g}_1$, $\overline{g}_2$ are different irreducible factors of $\overline{m}$, the corresponding ideals $(p, g_1(\alpha))$ and $(p, g_2(\alpha))$ are distinct.*

Remark. Here, $g(X) \in \mathbf{Z}[X]$ is any polynomial whose reduction in $\mathbf{F}_p[X]$ is $\overline{g}(X)$; the ideal $(p, g(\alpha))$ is insensitive to which such "lift" we choose.

Proof. *(i) It needs to be checked that π is well defined, in other words that if $f(\alpha) = 0$ then $\overline{f}(\overline{\alpha}) = 0$. However, if $f(\alpha) = 0$ then $m(X) \mid f(X)$, thus $f(X) = m(X)q(X)$ for some $q \in \mathbf{Z}[X]$. Reducing mod p, we see that $\overline{m}(X) \mid \overline{f}(X)$, and hence certainly $\overline{g}(X) \mid \overline{f}(X)$. Since $\overline{g}(\overline{\alpha}) = 0$, it follows that $\overline{f}(\overline{\alpha}) = 0$.

(ii) It is clear that $\pi(p) = \pi(g(\alpha)) = 0$, so certainly $(p, g(\alpha)) \subseteq \ker \pi$.

For the other direction, suppose that $\pi(f(\alpha)) = 0$, or in other words that $\overline{f}(\overline{\alpha}) = 0$. Now note that $\overline{g}$ is irreducible in $\mathbf{F}_p[X]$ and is satisfied by $\overline{\alpha}$, and hence it is the minimal polynomial of $\overline{\alpha}$ (over $\mathbf{F}_p$). It follows that $\overline{g} \mid \overline{f}$, that is to say $\overline{f}(X) = \overline{g}(X)\overline{q}(X)$ for some $\overline{q}(X) \in \mathbf{F}_p[X]$. Lifting (arbitrarily) to

$\mathbf{Z}[X]$, we have $f(X) = g(X)q(X)$ up to some multiple of p, and so indeed $f(\alpha) \in (p, g(\alpha))$.

(iii) The map π is clearly surjective, and so

$$\mathbf{F}_p[\overline{\alpha}] \cong \mathbf{Z}[\alpha]/\ker \pi.$$

By Lemma 1.5, $\mathbf{F}_p[\overline{\alpha}]$ is a field; this implies that $\ker \pi$ is a maximal ideal. Moreover the degree $[\mathbf{F}_p[\overline{\alpha}] : \mathbf{F}_p]$ is $\deg \overline{g}$, so in particular it has size $p^{\deg \overline{g}}$.

(iv) As a consequence of the first three parts, $\mathbf{Z}[\alpha]/(p, g(\alpha))$ is a field extension of $\mathbf{F}_p$, and α maps under the quotient to a root of $\overline{g}$. Thus if we did have $(p, g_1(\alpha)) = (p, g_2(\alpha))$ then $\overline{g}_1, \overline{g}_2$ would have a common root in some extension of $\mathbf{F}_p$. By Lemma 1.12, $\overline{g}_1, \overline{g}_2$ would then have a common factor in $\mathbf{F}_p[X]$, which is a contradiction since $\overline{g}_1, \overline{g}_2$ are distinct irreducible polynomials.

This completes the proof*. $\square$

Dedekind's lemma. Now we come to Dedekind's Lemma itself.

Theorem 9.2 (Dedekind's Lemma)**.** *Let K be a number field of degree n. Suppose that $\mathcal{O}_K = \mathbf{Z}[\alpha]$ for some α. Let $m(X) \in \mathbf{Z}[X]$ be the minimal polynomial of α. Let $\overline{m}(X) \in \mathbf{F}_p[X]$ be the reduction of m mod p, and suppose that this factors into distinct irreducible polynomials (over $\mathbf{F}_p$) as $\overline{g}_1(X)^{e_1} \cdots \overline{g}_r(X)^{e_r}$, where the $\overline{g}_i(X)$ are distinct. Then the factorisation of (p) into distinct prime ideals is $\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, where $\mathfrak{p}_i = (p, g_i(\alpha))$, and here g_i is an arbitrary lift of $\overline{g}_i$ to $\mathbf{Z}[X]$. Moreover, $N(\mathfrak{p}_i) = p^{\deg \overline{g}_i}$.*

Proof. Much follows immediately from Lemma 9.1. Indeed, from (iii) of that Lemma, $\mathfrak{p}_i$ is prime, and

$$N(\mathfrak{p}_i) = |\mathcal{O}_K/\mathfrak{p}_i| = [\mathbf{Z}(\alpha) : \mathfrak{p}_i] = p^{\deg \overline{g}_i}.$$

From (iv) of that lemma, the $\mathfrak{p}_i$ are distinct.

Now observe that

$$\mathfrak{p}_i^{e_i} = (p, g_i(\alpha))^{e_i} \subseteq (p, g_i(\alpha)^{e_i}),$$

and so

$$\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} \subseteq (p, g_1(\alpha)^{e_1} \cdots g_r(\alpha)^{e_r}) = (p, m(\alpha)) = (p). \tag{9.3}$$

However, the norm of the left-hand side of (9.3) is

$$N(\mathfrak{p}_1)^{e_1} \cdots N(\mathfrak{p}_r)^{e_r} = p^{e_1 \deg \overline{g}_1 + \cdots + e_r \deg \overline{g}_r} = p^{\deg \overline{m}} = p^{\deg m} = p^n,$$

which is the norm of the right-hand side. It follows that the inclusion (9.3) is in fact an equality. $\square$

Remarks. We have imposed the condition that K is *monogenic*, that is to say that $\mathcal{O}_K = \mathbf{Z}[\alpha]$ for some α. As we have seen on the example sheets, this is not a universal property, but it does hold for quadratic and cyclotomic fields, as well as many cubic fields.

One can prove a version of Dedekind's Lemma with the weaker assumption that $K = \mathbf{Q}(\alpha)$ and that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. This gives a version of Dedekind's theorem applicable to *all* number fields K, albeit with finitely many exceptional primes p for each K. Though this is not vastly more difficult to prove, we do not give it here.

Example: Splitting of primes in $\mathbf{Q}(i)$. As an example, we study the splitting of primes in the Gaussian field $\mathbf{Q}(i)$.

Proposition 9.3. *Rational primes p split in $\mathbf{Q}(i)$ as follows:*

- 2 *is ramified;*
- *If p is odd and $p \equiv 1 (\mathrm{mod}\, 4)$, p splits completely as a product of two ideals of norm p;*
- *If p is odd and $p \equiv 3 (\mathrm{mod}\, 4)$ then (p) is a prime ideal.*

Proof. This is a simple exercise in the application of Dedekind's criterion. Certainly the criterion applies, since $\mathcal{O}_K = \mathbf{Z}[i]$. The minimal polynomial of i is $X^2 + 1$. Over $\mathbf{F}_p$, this may be irreducible, or it may factor into two linear factors. The second possibility occurs precisely when -1 is a quadratic residue mod p, which (from *Part A Number Theory*) we know occurs precisely when $p = 2$ or p is an odd prime $\equiv 1 (\mathrm{mod}\, 4)$).

When $p = 2$, $X^2 + 1 = (X + 1)^2$ in $\mathbf{F}_2[X]$, and so by Dedekind's criterion $(2) = (2, 1 + i)^2$ is the factorisation of (2) into prime ideals.

When p is an odd prime $\equiv 1 (\mathrm{mod}\, 4)$, there are two distinct square roots of -1 modulo p, $\pm\gamma$ (say). Then $X^2 + 1 = (X + \gamma)(X - \gamma)$ and Dedekind tells us that $(p) = (p, i + \gamma)(p, i - \gamma)$. For instance, $X^2 + 1 = (X + 2)(X - 2)$ in $\mathbf{F}_5[X]$ and so $(5) = (5, 2 + i)(5, -2 + i)$.

When p is an odd prime $\equiv 3 (\mathrm{mod}\, 4)$, $X^2 + 1$ is irreducible and so Dedekind tells us that $(p) = (p, i^2 + 1) = (p)$ is prime. $\square$

Factoring a general ideal. One fairly commonly finds the need to factor an arbitrary ideal $\mathfrak{a} \subseteq \mathcal{O}_K$ into prime ideals. This can be a little tedious, but here is a general strategy which will always work. Things can often be sped up with *ad hoc* observations.

- Begin by finding a rational integer $m \in \mathfrak{a}$. To do this, first pick $\alpha \in \mathfrak{a}$, and then find a polynomial $f \in \mathbf{Z}[X]$, $f(X) = c_n\alpha^n + \cdots + c_0$ satisfied by α (a good choice is the minimal polynomial). Then $c_0 = -\alpha(c_1 + c_2\alpha + \cdots + c_n\alpha^{n-1})$ lies in $\mathfrak{a}$.
- We have $\mathfrak{a}|(m)$. Factor m into rational primes p_i. We may then apply Dedekind to each (p_i).
- We now have a list of all possible prime ideal factors of $\mathfrak{a}$. Note they may occur with multiplicity. To find out which of them actually *are* prime factors of $\mathfrak{a}$, we need to be able to test when $\mathfrak{b} \mid \mathfrak{a}$, or in other words when $\mathfrak{a} \subseteq \mathfrak{b}$. This can often be done in an *ad hoc* way; if necessary, one can explicitly see if each generator of $\mathfrak{a}$ is in the $\mathcal{O}_K$ span of the generators of $\mathfrak{b}$ by writing everything in terms of an integral basis and then solving the resulting system of equations by putting everything in Smith normal form, but in examples we will see this is not generally necessary.

Example. Let $K = \mathbf{Q}(\sqrt{-29})$. Find the prime factorisation of $\mathfrak{a} = (6, 1 + \sqrt{-29})$ into prime ideals in $\mathcal{O}_K$.

Solution. Since $\mathfrak{a}|(6) = (2)(3)$, we first factor (2) and (3). We have $\mathcal{O}_K = \mathbf{Z}[\sqrt{-29}]$, and the minimal polynomial of $\sqrt{-29}$ is $X^2 + 29$. Modulo 2, this factors as $(X+1)^2$, so $(2) = \mathfrak{p}^2$ where $\mathfrak{p} = (2, 1 + \sqrt{-29})$. Modulo 3, this factors as $(X-1)(X+1)$ and so $(3) = \mathfrak{q}_1\mathfrak{q}_2$ where $\mathfrak{q}_1 = (3, 1 + \sqrt{-29})$ and $\mathfrak{q}_2 = (3, -1 + \sqrt{-29})$.

We need to work out which of these divide $\mathfrak{a}$. We do not have $\mathfrak{p}^2 \mid \mathfrak{a}$, since $\mathfrak{p}^2 = (2)$ and $\frac{1}{2}(1 + \sqrt{-29}) \notin \mathcal{O}$. However, it is clear that $\mathfrak{a} \subseteq \mathfrak{p}$, that is to say $\mathfrak{p} \mid \mathfrak{a}$. In particular, $\mathfrak{a} \neq \mathcal{O}_K = (1)$.

Turning to the $\mathfrak{q}$'s, it is clear that $\mathfrak{a} \subseteq \mathfrak{q}_1$ and so $\mathfrak{q}_1 \mid \mathfrak{a}$. However,the ideal $\mathfrak{a} + \mathfrak{q}_2$ generated by $\mathfrak{a}, \mathfrak{q}_2$ contains $(1 + \sqrt{-29}) - (-1 + \sqrt{-29}) = 2$, as well as 3, and hence contains 1; this means that $\mathfrak{a} \not\subseteq \mathfrak{q}_2$ and so $\mathfrak{q}_2 \nmid \mathfrak{a}$. Alternatively, we could try and see whether $1 + \sqrt{-29} \in \mathfrak{q}_2$ by writing things in an integral basis, as suggested (as a last resort!) above: if

$$(1 + \sqrt{-29}) = 3(a + b\sqrt{-29}) + (c + d\sqrt{-29})(-1 + \sqrt{-29})$$

then, comparing coefficients, we get $3a - c - 29d = 3b + c - d = 1$. Adding gives $3(a + b - 10d) = 2$, a contradiction. One could be more systematic using Smith normal form if desired.

10. The class group

Basic definitions. Suppose that $\mathfrak{a}, \mathfrak{b}$ are ideals in $\mathcal{O}_K$. We write $\mathfrak{a} \sim \mathfrak{b}$ if there are principal ideals $(x), (y)$ such that $\mathfrak{a}(x) = \mathfrak{b}(y)$. It is easy to check that $\sim$ is an equivalence relation. The *ideal class group* $\mathrm{Cl}(K)$ is then defined to be the quotient $\mathrm{Ideals}(\mathcal{O}_K)/\sim$, that is to say the set of ideals up to equivalence. Equivalence classes are denoted by square brackets $[\mathfrak{a}]$, and these are called *ideal classes*. Note that all principal ideals lie in the same class.

It is easy to check that if $\mathfrak{a} \sim \mathfrak{b}$ and $\mathfrak{a}' \sim \mathfrak{b}'$ then $\mathfrak{a}\mathfrak{a}' \sim \mathfrak{b}\mathfrak{b}'$. This means that the product operation on ideals descends to give a well-defined product on ideal classes, thus $[\mathfrak{a}] \cdot [\mathfrak{b}] = [\mathfrak{a}\mathfrak{b}]$. This operation has an identity (the class consisting of principal ideals) and inverses exist by Proposition 5.6. Therefore $\mathrm{Cl}(K)$ is indeed a group, called the *ideal class group* of K.

Note that $\mathrm{Cl}(K)$ is trivial (that is, has size 1) if and only if $\mathcal{O}_K$ is a PID. Indeed, if $\mathfrak{a} \sim (1)$ then there are $x, y \in \mathcal{O}_K$ so that $\mathfrak{a}(x) = (y)$. This means that $x \mid y$ (indeed, $y = ax$ for some $a \in \mathfrak{a}$) and so $\mathfrak{a} = (\frac{y}{x})$ is principal.

***Fractional ideals.** The class group looks more natural if we introduce the notion of a fractional ideal. This is a subset of K of the form

$$x^{-1}\mathfrak{a} := \{x^{-1}a : a \in \mathfrak{a}\} \subseteq K,$$

for some ideal $\mathfrak{a}$ in $\mathcal{O}_K$ and some $x \in K$.

Note that fractional ideals are $\mathcal{O}_K$-modules, and in fact it is easy to show that the fractional ideals are precisely the finitely-generated $\mathcal{O}_K$-submodules of K. (One may "clear denominators", picking x so that if $e_1, \dots, e_r$ generate the fractional ideal then each xe_i lies in $\mathcal{O}_K$.)

One may develop the basic theory of fractional ideals in much the same way as for ideals, for example defining products and principal fractional ideals $\{(x) = x\alpha : \alpha \in \mathcal{O}_K\}$ for all $x \in K$.

Unlike the ideals, however, the non-zero fractional ideals form a group under multiplication. This follows from Proposition 5.6 and the fact that every non-zero principal fractional ideal is invertible, since $(x)(x^{-1}) = (1)$. This group is often denoted by $\mathrm{Div}(\mathcal{O}_K)$.

The ideal class group $\mathrm{Cl}(K)$ is then isomorphic to the quotient of $\mathrm{Div}(\mathcal{O}_K)$ by the subgroup of principal ideals.

Minkowski bound. Finiteness of the class group. In this section we will state, and set up the proof of, the most important theorem about the ideal class group. This is the fact that it is a finite group. We establish this together with additional information, the *Minkowski bound*, which can be used to calculate the group in practice (we will present several examples in the next chapter). The key statement is Theorem 10.3 below.

The proof is by no means trivial. It involves tools from the geometry of numbers (see Section 10 for a brief introduction, and Appendix B for proofs) as well as quite a number of other nontrivial ideas. Because the proof is quite hard, we will present the imaginary quadratic case (which is conceptually easier) first, in Section 10, and then the general case in Section 10. The arguments of Section 10 are probably the most highly non-examinable in the course (they are in absolutely no sense examinable), and I will only lecture them if time allows.

The Minkowski constant M_K. Let K be a number field with embeddings $\sigma_1, \dots, \sigma_n : K \to \mathbf{C}$. It is (somewhat[2]) standard to write r_1 for the number of real embeddings $\sigma_i : K \to \mathbf{C}$, and r_2 the number of pairs of conjugate complex embeddings $\sigma_i \to \mathbf{C}$. (An embedding is deemed real if its image is contained in $\mathbf{R}$, and complex otherwise). Note that $r_1 + 2r_2 = n$.

Definition 10.1 (Minkowski constant)**.** Suppose that K is a number field of degree n with r_1 real embeddings and r_2 pairs of conjugate complex embeddings. Let Δ_K be the discriminant of K. Then we define the *Minkowski constant*

$$M_K := \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n} \sqrt{|\Delta_K|}. \tag{10.1}$$

Almost all (but not all) applications of the Minkowski bound you are likely to see in a first course such as this are to quadratic fields $\mathbf{Q}(\sqrt{d})$, so let us pause to record the values of M_K in this case explicitly. There are two possibilities:

(i) Real quadratic fields ($d > 0$), where $r_1 = 2$ and $r_2 = 0$. Then $M_K = \frac{1}{2}\sqrt{|\Delta_K|}$;

(ii) Imaginary quadratic fields ($d < 0$), where $r_1 = 0$ and $r_2 = 1$. Then $M_K = \frac{2}{\pi}\sqrt{|\Delta_K|}$.

[2] It is also (somewhat) standard to write r, s instead of r_1, r_2.

In fact, combining this with Proposition 2.17, we can be even more explicit, as follows.

Lemma 10.2. *Let $\mathbf{Q}(\sqrt{d})$, $d \neq 1$ a squarefree integer, be a quadratic field. Then M_K is given as follows:*

(i) *If $d > 0$ and $d \equiv 2, 3 \pmod 4$, $M_K = \sqrt{d}$;*
(ii) *If $d > 0$ and $d \equiv 1 \pmod 4$, $M_K = \frac{1}{2}\sqrt{d}$;*
(iii) *If $d < 0$ and $d \equiv 2, 3 \pmod 4$, $M_K = \frac{4}{\pi}\sqrt{|d|}$;*
(iv) *If $d < 0$ and $d \equiv 1 \pmod 4$, $M_K = \frac{2}{\pi}\sqrt{|d|}$.*

Now we state the key result, the Minkowski bound.

Theorem 10.3 (Minkowski bound)**.** *Let K be a number field with Minkowski constant M_K. Then*

(i) *the class group $\mathrm{Cl}(K)$ is finite;*
(ii) *every class in $\mathrm{Cl}(K)$ contains an ideal $\mathfrak{a}$ with $N(\mathfrak{a}) \leqslant M_K$;*
(iii) *$\mathrm{Cl}(K)$ is generated by (the identity and) the prime ideals $\mathfrak{p}$ dividing the principal ideals (p), where p is a rational prime of size at most M_K.*

Remark. (ii) is the key statement; the others follow almost immediately from it. Indeed, recall Lemma 7.6, which states that $\mathfrak{a} | (N(\mathfrak{a}))$. Then (ii) implies that the (ideal) divisors of the ideals (a), with a a rational integer $\leqslant M_K$, represent every class in $\mathrm{Cl}(K)$. (i) follows immediately. Factoring each such a into rational primes, (iii) also follows straight away.

Definition 10.4. The size of $\mathrm{Cl}(K)$ is called the *class number* of K and it is denoted h_K.

Elements with small norm. In this section we give an initial reduction toward the proof of Theorem 10.3, showing that it is a consequence of the following result, which states that every ideal $\mathfrak{a}$ contains an element of small norm (relative to the norm of $\mathfrak{a}$).

Proposition 10.5 (Elements of small norm)**.** *Let K be a number field and let $\mathfrak{a}$ be a nonzero ideal in $\mathcal{O}_K$. Then there is some $x \in \mathfrak{a}$ with $|N_{K/\mathbf{Q}}(x)| \leqslant M_K N(\mathfrak{a})$.*

This proposition contains all the real difficulties in the proof of Theorem 10.3 and occupies the last few sections of this chapter. To conclude this section, we deduce Theorem 10.3 from it.

Proof. [Proof of Theorem 10.3, assuming Proposition 10.5.] It is enough to prove Theorem 10.3 (ii); as we observed, the other statements follow quickly from this.

Take some ideal class in $\mathrm{Cl}(K)$, and let $\mathfrak{b}$ be an (arbitrary) ideal in it. Let $\mathfrak{c}$ be an inverse of $\mathfrak{b}$ in the class group, so that $\mathfrak{b}\mathfrak{c} = (x)$ principal. By Proposition 10.5, $\mathfrak{c}$ contains an element y with $|N_{K/\mathbf{Q}}(y)| \leqslant M_K N(\mathfrak{c})$. Now $(y) \subseteq \mathfrak{c}$, that is to say $\mathfrak{c}$ divides (y), and so there is $\mathfrak{a}$ with $\mathfrak{c}\mathfrak{a} = (y)$. In the ideal class group, we have $[\mathfrak{b}] = [\mathfrak{c}]^{-1} = [\mathfrak{a}]$. Taking norms, and using Lemma 4.9, we have

$$N(\mathfrak{a})N(\mathfrak{c}) = N((y)) = |N_{K/\mathbf{Q}}(y)| \leqslant M_K N(\mathfrak{c}),$$

and so $N(\mathfrak{a}) \leqslant M_K$. The result is proven. $\square$

The remaining (much more substantial) task is to prove Proposition 10.5.

Geometry of numbers. In the proof of Proposition 10.5, we will use the *geometry of numbers*, which can be roughly defined as the study of when convex bodies intersect lattices.

A *lattice* Λ in $\mathbf{R}^n$ is the free abelian group generated by n linearly independent vectors $v_1, \dots, v_n$, that is to say $\Lambda = \mathbf{Z}v_1 \oplus \mathbf{Z}v_2 \oplus \cdots \oplus \mathbf{Z}v_d$. The determinant $\det(\Lambda)$ is $|\det(v_1, \dots, v_n)|$, which can be interpreted as the volume of the fundamental parallelepiped $\bigoplus_{i=1}^d [0,1]v_i$; it turns out to depend only on Λ, and not on the choice of integral basis $v_1, \dots, v_n$ (this essentially follows as in Proposition 2.13 – see Lemma A.4). For more on lattices, see Appendix A.

The result from the geometry of numbers that we shall need is the following result, known as Minkowski's first theorem.

Theorem 10.6 (Minkowski I)**.** *Suppose that $\Lambda \subseteq \mathbf{R}^n$ is a lattice, and that $B \subseteq \mathbf{R}^n$ is a centrally symmetric (that is, if $x \in B$ then $-x \in B$), compact, convex body. Suppose that $\mathrm{vol}(B) \geqslant 2^n \det(\Lambda)$. Then B contains a nonzero point of Λ.*

The proof of this is not especially difficult. See Appendix B.

Elements with small norm: imaginary quadratic fields. We turn now to the proof of Proposition 10.5. We will first give the proof in the imaginary quadratic case $K = \mathbf{Q}(\sqrt{d})$, $d < 0$, as it is rather easier to understand than the general case, and also most of the examples we will consider will be of this form. The reason this case is easier to understand is that (with the usual identification of $\mathbf{C}$ and $\mathbf{R}^2$) the ring of integers $\mathcal{O}_K$ is a lattice. Moreover, if $x \in K$ then $N_{K/\mathbf{Q}}(x)$ is simply the Euclidean norm of x, squared.

Let e_1, e_2 be an integral basis of $\mathcal{O}_K$. Then (considered as a subset of $\mathbf{R}^2$), $\mathcal{O}_K$ is a lattice with fundamental parallelepiped spanned by e_1, e_2, and the determinant of this lattice is $|\det N|$, where $N := \left(\begin{smallmatrix} \operatorname{Re} e_1 & \operatorname{Re} e_2 \\ \operatorname{Im} e_1 & \operatorname{Im} e_2 \end{smallmatrix}\right)$. Recall (see Definition 1.19) that the discriminant Δ_K is $(\det M)^2$, where $M := \left(\begin{smallmatrix} e_1 & e_2 \\ \overline{e_1} & \overline{e_2} \end{smallmatrix}\right)$. One may easily check using elementary row operations that $|\det N| = \frac{1}{2}|\det M|$, and so we arrive at the conclusion that the determinant of $\mathcal{O}_K$, considered as a lattice in $\mathbf{R}^2$, is $\frac{1}{2}\sqrt{|\Delta_K|}$. (We caution that in general talking about 'the determinant of $\mathcal{O}_K$' has no meaning; here we are thinking of $\mathcal{O}_K$ as embedded in $\mathbf{R}^2$ via the standard identification of $\mathbf{C}$ with $\mathbf{R}^2$.)

Now let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$. It may also be considered as a lattice in $\mathbf{R}^2$, and since $[\mathcal{O}_K : \mathfrak{a}] = N(\mathfrak{a})$, it follows from Lemma A.6 that (considered as a lattice) it has determinant $\frac{1}{2}N(\mathfrak{a})\sqrt{|\Delta_K|}$.

The Euclidean ball of radius r, where $r^2 := M_K N(\mathfrak{a})$, has area $\pi r^2 = 2N(\mathfrak{a})\sqrt{|\Delta_K|}$. By Minkowski's first theorem, this ball contains a nonzero point of $\mathfrak{a}$, and therefore $\mathfrak{a}$ has an element of norm at most r^2.

This concludes the proof of Proposition 10.5 in the imaginary quadratic case.

***Elements with small norm: general case.** Let us give the generalisation of the argument of the preceding section to an arbitrary number field. The basic form of the argument is the same, but there are two moderately serious issues (and some LaTeX difficulties). We give the proof as a response to these issues.

Serious issue 1. In general, $\mathcal{O}_K \subset \mathbf{C}$ does not resemble a lattice. Indeed, this is already the case for *real* quadratic fields $K = \mathbf{Q}(\sqrt{d})$, $d > 0$. In this case, $\mathcal{O}_K$ will in fact be a dense subset of the real line. Equally, since lattices in $\mathbf{C}$ are two-dimensional, it makes no sense to try and think of $\mathcal{O}_K$ as a lattice in $\mathbf{C}$ when $[K : \mathbf{Q}] > 2$.

Solution. The trick is to use the embeddings $\sigma_i : K \to \mathbf{C}$ to embed $\mathcal{O}_K$ in an n-dimensional Euclidean space in which it *is* a lattice. To do this, suppose that $\sigma_1, \ldots, \sigma_{r_1}$ are the real embeddings and that $\sigma_{r_1+1}, \ldots, \sigma_{r_1+r_2}$ are mutually non-conjugate complex embeddings (thus, if we include a complex embedding σ, we do not include $\overline{\sigma}$). Now consider the map

$$\Phi : K \to \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \cong \mathbf{R}^n$$

given by

$$\Phi(x) = (\sigma_1(x), \dots, \sigma_{r_1}(x), \sigma_{r_1+1}(x), \dots, \sigma_{r_1+r_2}(x)).$$

To spell it out,

$$\begin{aligned}\Phi(x) =&(\sigma_1(x), \dots, \sigma_{r_1}(x),\\ &\operatorname{Re}\sigma_{r_1+1}(x), \operatorname{Im}\sigma_{r_1+1}(x), \dots, \operatorname{Re}\sigma_{r_1+r_2}(x), \operatorname{Im}\sigma_{r_1+r_2}(x)).\end{aligned}$$

In the imaginary quadratic case, Φ just corresponds to the usual identification of $\mathbf{C}$ with $\mathbf{R}^2$, as discussed above. In general one should probably think of $\Phi(K)$ as "$K \otimes_{\mathbf{Q}} \mathbf{R}$" but I will not elaborate on this comment.

Example. Suppose that $K = \mathbf{Q}(\sqrt{2})$. Then

$$\Phi(a + b\sqrt{2}) = (a + b\sqrt{2}, a - b\sqrt{2}).$$

Note in particular that

$$\Phi(\mathcal{O}_K) = \{a(1,1) + b(\sqrt{2}, -\sqrt{2}) : a, b \in \mathbf{Z}\}$$

is a lattice in $\mathbf{R}^2$.

This, it turns out, is a general feature, and moreover we have the following lemma, which generalises the observations we made in the imaginary quadratic case.

Lemma 10.7. *$\Phi(\mathcal{O}_K)$ is a lattice in $\mathbf{R}^n$, and*

$$\det(\Phi(\mathcal{O}_K)) = \frac{1}{2^{r_2}}\sqrt{|\Delta_K|}. \tag{10.2}$$

Proof. Certainly Φ is an additive homomorphism. Thus, if $e_1, \dots, e_n$ is an integral basis for $\mathcal{O}_K$, $\Phi(\mathcal{O}_K)$ is the $\mathbf{Z}$-module generated by $\Phi(e_1), \dots, \Phi(e_n)$. Thus $\det(\Phi(e_1), \dots, \Phi(e_n))$ is $\det N$, where N is the transpose of

$$\begin{pmatrix} \sigma_1(e_1) & \dots & \sigma_{r_1}(e_1) & \operatorname{Re}\sigma_{r_1+1}(e_1) & \operatorname{Im}\sigma_{r_1+1}(e_1) & \dots & \operatorname{Im}\sigma_{r_1+r_2}(e_1) \\ \vdots & & & & & & \vdots \\ \sigma_1(e_n) & \dots & \sigma_{r_1}(e_n) & \operatorname{Re}\sigma_{r_1+1}(e_n) & \operatorname{Im}\sigma_{r_1+1}(e_n) & \dots & \operatorname{Im}\sigma_{r_1+r_2}(e_n) \end{pmatrix}.$$

On the other hand, recall (from Chapter 2) that Δ_K is $(\det M)^2$, where

$$M^T := \begin{pmatrix} \sigma_1(e_1) & \dots & \sigma_{r_1}(e_1) & \sigma_{r_1+1}(e_1) & \overline{\sigma_{r_1+1}}(e_1) & \dots & \overline{\sigma_{r_1+r_2}}(e_1) \\ \vdots & & & & & & \vdots \\ \sigma_1(e_n) & \dots & \sigma_{r_1}(e_n) & \sigma_{r_1+1}(e_n) & \overline{\sigma_{r_1+1}}(e_n) & \dots & \overline{\sigma_{r_1+r_2}}(e_n) \end{pmatrix}$$

Here, we have arranged the embeddings of K in complex conjugate pairs.

Now by the alternating multilinearity of the determinant,

$$\begin{aligned}\det(\dots,\operatorname{Re} v,\operatorname{Im} v,\dots) &= \det(\dots,\frac{1}{2}(v+\overline{v}),\frac{1}{2}(v-\overline{v}),\dots)\\ &= -\frac{1}{2}\det(\dots,v,\overline{v},\dots).\end{aligned}$$

Using this r_2 times, it follows that $|\det N| = 2^{-r_2}|\det M|$, which implies (10.2). In particular, $\det N \neq 0$ so $\Phi(e_1),\dots,\Phi(e_n)$ are independent, and $\Phi(\mathcal{O}_K)$ *is* a lattice. □

Since $\Phi(\mathfrak{a})$ is a subgroup of $\Phi(\mathcal{O}_K)$ of index $N(\mathfrak{a})$, the following is a consequence of Lemma A.6.

Corollary 10.8. *Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$. Then $\Phi(\mathfrak{a})$ is a lattice in $\mathbf{R}^n$, and*

$$\det(\Phi(\mathfrak{a})) = \frac{1}{2^{r_2}}N(\mathfrak{a})\sqrt{|\Delta_K|}.$$

Serious issue 2. The set $\{\Phi(x) : x \in K, |N_{K/\mathbf{Q}}(x)| \leqslant R\}$ is not naturally contained in a convex set. Indeed, $|N_{K/\mathbf{Q}}(x)| \leqslant R$ if and only if $\Phi(x)$ belongs to the set

$$\begin{aligned}B := \{(x_1,\dots,x_{r_1},&z_{r_1+1},\dots,z_{r_1+r_2}) \in \mathbf{R}^{r_1}\times\mathbf{C}^{r_2} :\\ &|x_1|\cdots|x_{r_1}||z_{r_1+1}|^2\cdots|z_{r_1+r_2}|^2 \leqslant R\}.\end{aligned}$$

This is generally not convex (although, as we saw in the last section, it is convex in the imaginary quadratic case, when $r_1 = 0$ and $r_2 = 1$).

Solution. B contains a relatively large convex set B', and we can use this instead. Indeed, set

$$\begin{aligned}B' := \{(x_1,\dots,x_{r_1},&z_{r_1+1},\dots,z_{r_1+r_2}) \in \mathbf{R}^{r_1}\times\mathbf{C}^{r_2} :\\ &|x_1|+\dots+|x_{r_1}|+2(|z_{r_1+1}|+\dots+|z_{r_1+r_2}|) \leqslant nR^{1/n}\}.\end{aligned}$$

It is quite easy to check that B' is convex. The fact that $B' \subseteq B$ is an instance of the arithmetic-geometric means inequality:

$$\begin{aligned}\Big(\frac{|x_1|+\dots+|x_{r_1}|+2(|z_{r_1+1}|+\dots+|z_{r_1+r_2}|)}{n}&\Big)^n\\ &\geqslant |x_1|\cdots|x_{r_1}||z_{r_1+1}|^2\cdots|z_{r_1+r_2}|^2.\end{aligned}$$

In particular,

$$\text{If } \Phi(x) \in B', \text{ then } |N_{K/\mathbf{Q}}(x)| \leqslant R. \tag{10.3}$$

Now we have
$$\mathrm{vol}(B') = \frac{1}{n!}2^{r_1}(\frac{\pi}{2})^{r_2}(nR^{1/n})^n. \tag{10.4}$$
(this is a multivariable integration calculation, which I have put on Sheet X).

Using Lemma 10.8 and (10.4), a short computation now confirms that $\mathrm{vol}(B') \geqslant 2^n \det(\Phi(\mathfrak{a}))$ if and only if
$$R \geqslant \frac{n!}{n^n}(\frac{4}{\pi})^{r_2} N(\mathfrak{a})\sqrt{|\Delta_K|},$$
that is to say if and only if
$$R \geqslant M_K N(\mathfrak{a}).$$
If R does satisfy this inequality, Minkowski's Theorem (Theorem 10.6) tells us that B' contains a point of $\Phi(\mathfrak{a})$ which, by (10.3), implies that $\mathfrak{a}$ contains an element of norm at most R.

The proof of Proposition 10.5 in the general case is now finished.

11. Example class group calculations

In this chapter we compute the class groups of some example imaginary quadratic fields K. The general procedure is always

(i) Observe the basic features of K (ring of integers, integral basis, discriminant etc) and write down the Minkowski bound M_K. By Theorem 10.3, generators for $\mathrm{Cl}(K)$ may be found amongst the prime divisors of (p), $p \leqslant M_K$.

(ii) Factor all of the ideals (p), where $p \leqslant M_K$ is a rational prime, using Dedekind's theorem. This will give an explicit list of prime ideals generating $\mathrm{Cl}(K)$.

(iii) Figure out what relations there are, in the ideal class group, between the prime ideals generated in (ii).

Items (i) and (ii) are purely formulaic, but there is a little bit of an art to (iii), at least as we shall do things in this course. However, in the imaginary quadratic case there is a *key trick* available: one can easily list the elements of $\mathcal{O}_K$ (if any) of a given norm, since the norm takes only positive values.

If $\mathfrak{a} = (\alpha)$ is principal then (Lemma 4.9) $N(\mathfrak{a}) = |N_{K/\mathbf{Q}}(\alpha)| = N_{K/\mathbf{Q}}(\alpha)$. Thus one can test whether or not an ideal $\mathfrak{a}$ is principal by writing down all the elements $\alpha \in \mathcal{O}_K$ with $N_{K/\mathbf{Q}}(\alpha) = N(\mathfrak{a})$ and then testing whether $\mathfrak{a} = (\alpha)$ or not, which in practice is pretty straightforward. In particular, if

$N(\mathfrak{a})$ is not the norm of some element, $\mathfrak{a}$ cannot be principal. (However, the converse is not true.)

We will work through four examples according to the scheme detailed above. In all cases, the basic features of K have already been worked out in Propositions 2.17 (integral bases) and 10.2 (Minkowski constant), which the reader should recall now.

$\mathbf{Q}(i)$ and sums of squares. Let us begin by giving a new proof of the following fact from *Rings and Modules*.

Lemma 11.1. *The class group of $K = \mathbf{Q}(i)$ is trivial. In particular, $\mathcal{O}_K = \mathbf{Z}[i]$ is a PID.*

Proof. By Lemma 10.2 (part (iii)), $M_K = \frac{4}{\pi} < 2$. Since there are no primes less than 2, Theorem 10.3 (ii) immediately implies that $\mathrm{Cl}(K)$ is trivial. $\square$

Corollary 11.2. *Let p be an odd prime with $p \equiv 1 (\mathrm{mod}\, 4)$. Then p is a sum of two squares.*

Proof. Let $K = \mathbf{Q}(i)$. Recall Proposition 9, which details the manner in which rational primes split in $\mathcal{O}_K = \mathbf{Z}[i]$. If $p \equiv 1 (\mathrm{mod}\, 4)$ then (p) splits as $\mathfrak{p}_1\mathfrak{p}_2$ in $\mathcal{O}_K$, where $\mathfrak{p}_1, \mathfrak{p}_2$ have norm p. Since (as we now know) $\mathcal{O}_K$ is a PID, $\mathfrak{p}_1$ is principal, say $\mathfrak{p}_1 = (a + ib)$ for some $a, b \in \mathbf{Z}$. Taking norms, we see that

$$p = N(\mathfrak{p}_1) = N((a + ib)) = N_{K/\mathbf{Q}}(a + ib) = a^2 + b^2.$$

This completes the proof. $\square$

Of course, this is an *if and only if*: if $p \equiv 3 (\mathrm{mod}\, 4)$, then it follows immediately by working mod 4 that p is not the sum of two squares. You could deduce this from the machinery above if you really wanted to.

$\mathbf{Q}(\sqrt{-5})$. We have already said a lot about this field, but let us revisit it in the light of our new techniques.

(i) Since $d \equiv 3 (\mathrm{mod}\, 4)$, $\mathcal{O}_K = \mathbf{Z}[\sqrt{-5}]$. By Lemma 10.2 (iii), $M_K = \frac{4}{\pi}\sqrt{5} < 3$ (to check this without resorting to a calculator, square up both sides to see that it is enough to show that $\pi^2 > 80/9$, which is obvious since $\pi > 3$). It follows from the Minkowski bound, Theorem 10.3, that generators of $\mathrm{Cl}(K)$ may be found amongst the (ideal) prime factors of (2).

(ii) The minimal polynomial $m(X)$ for $\sqrt{-5}$ is X^2+5. Over $\mathbf{F}_2$, this factors as $(X+1)^2$. By Dedekind's lemma we therefore have $(2)=\mathfrak{p}^2$ where $\mathfrak{p}=(2,1+\sqrt{-5})$ is a prime ideal of norm 2.

(iii) Since $N_{K/\mathbf{Q}}(a+b\sqrt{-5})=a^2+5b^2$, there is no element of $\mathcal{O}_K$ of norm 2. Therefore $\mathfrak{p}$ is not principal.

The only conclusion now is that $\mathrm{Cl}(K)$ is a cyclic group of order two, generated by $[\mathfrak{p}]$. In particular, $h_K=2$.

$\mathbf{Q}(\sqrt{-29})$. (i) Since $d\equiv 3(\mathrm{mod}\,4)$, $\mathcal{O}_K=\mathbf{Z}[\sqrt{-29}]$. By Lemma 10.2 (iii), $M_K=\frac{4}{\pi}\sqrt{29}<7$. Thus, by the Minkowski bound, generators of $\mathrm{Cl}(K)$ may be found amongst the (ideal) prime factors of (2), (3) and (5).

(ii) The minimal polynomial $m(X)$ for $\sqrt{-29}$ is X^2+29.

Over $\mathbf{F}_2$ this factors as $(X+1)^2$, so by Dedekind $(2)=\mathfrak{p}^2$ where $\mathfrak{p}=(2,1+\sqrt{-29})$ has norm 2.

Over $\mathbf{F}_3$ this factors as $(X+1)(X-1)$, so by Dedekind $(3)=\mathfrak{q}_3\mathfrak{q}_3'$ where $\mathfrak{q}_3=(3,1+\sqrt{-29})$, $\mathfrak{q}_3'=(3,-1+\sqrt{-29})$ are distinct prime ideals of norm 3.

Over $\mathbf{F}_5$ this factors as $(X+1)(X-1)$, so by Dedekind $(5)=\mathfrak{q}_5\mathfrak{q}_5'$ where $\mathfrak{q}_5=(5,1+\sqrt{-29})$, $\mathfrak{q}_5'=(5,-1+\sqrt{-29})$ are distinct prime ideals of norm 5.

(iii) Since $[\mathfrak{q}_3']=[\mathfrak{q}_3]^{-1}$, $[\mathfrak{q}_5']=[\mathfrak{q}_5]^{-1}$, the class group is generated by $\mathfrak{p},\mathfrak{q}_3,\mathfrak{q}_5$. However, we need to do quite a lot more work to determine it completely. We make the following preliminary observations.

- None of $\mathfrak{p},\mathfrak{q}_3,\mathfrak{q}_5$ is principal, since $\mathcal{O}_K$ does not have elements of norm $2,3$ or 5 (the norm is $N(a+b\sqrt{-29})=a^2+29b^2$).
- $\mathfrak{q}_3^2$ is not principal. Indeed, the only elements of $\mathcal{O}_K$ of norm 9 are ± 3, so if $\mathfrak{q}_3^2$ was principal we would have $\mathfrak{q}_3^2=(3)=\mathfrak{q}_3\mathfrak{q}_3'$ and thus $\mathfrak{q}_3=\mathfrak{q}_3'$, contrary to what we learned from Dedekind (namely that these ideals are *distinct*).
- $\mathfrak{q}_3^3$ is not principal, since there is no element in $\mathcal{O}_K$ of norm 27.
- $\mathfrak{q}_5^2$ is not principal, for essentially the same reason that $\mathfrak{q}_3^2$ is not.
- There *is* an element of $\mathcal{O}_K$ of norm 125, namely $3+2\sqrt{-29}$. We need to find the prime factorisation of $\mathfrak{a}:=(3+2\sqrt{-29})$. A very helpful observation here is that $\mathfrak{q}_5\nmid\mathfrak{a}$. Indeed, $2+2\sqrt{-29}\in\mathfrak{q}_5$, so if $\mathfrak{a}\subseteq\mathfrak{q}_5$ we would have $1\in\mathfrak{q}_5$, which is absurd. Now $\mathfrak{a}\mid(N(\mathfrak{a}))=(125)=(5)^3$. Thus all prime factors of $\mathfrak{a}$ are $\mathfrak{q}_5$ or $\mathfrak{q}_5'$, and hence they must all be the latter. Comparing norms gives $\mathfrak{a}=\mathfrak{q}_5'^3$. Thus

$\mathfrak{q}_5'^3$ is principal. By the same reasoning (or taking conjugates) so is $\mathfrak{q}_5^3$. Thus $[\mathfrak{q}_5]$ has order 3 in $\mathrm{Cl}(K)$.

The above are at least somewhat scientific, but we got stuck with $\mathfrak{q}_3$, and to finish the job it really helps to "observe" the relation

$$(2)(3)(5) = (30) = (1+\sqrt{-29})(1-\sqrt{-29}).$$

The prime factorisation of the left-hand side is of course $\mathfrak{p}^2\mathfrak{q}_3\mathfrak{q}_3'\mathfrak{q}_5\mathfrak{q}_5'$, and the two (principal) ideals on the right hand side both have norm 30. Thus $(1+\sqrt{-29})$ must be one of $\mathfrak{p}\mathfrak{q}_3\mathfrak{q}_5$, $\mathfrak{p}\mathfrak{q}_3\mathfrak{q}_5'$, $\mathfrak{p}\mathfrak{q}_3'\mathfrak{q}_5$, $\mathfrak{p}\mathfrak{q}_3'\mathfrak{q}_5'$. Whichever holds, we see that $[\mathfrak{q}_3]$ is in the group generated by $[\mathfrak{p}]$ and $[\mathfrak{q}_5]$. (For instance, if $(1+\sqrt{-29}) = \mathfrak{p}\mathfrak{q}_3'\mathfrak{q}_5$ then $[\mathfrak{p}][\mathfrak{q}_3]^{-1}[\mathfrak{q}_5]$ is the identity).

We are now done: $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}]$, which has order 2, and $[\mathfrak{q}_5]$, which has order 3, and therefore $\mathrm{Cl}(K)$ is cyclic of order 6. (It is easy to conclude from all this that in fact $[\mathfrak{q}_3]$ has order 6, which explains why it was troublesome to analyse!)

Here is another way in which we could have finished the argument, once we found elements of order 2 and 3 in the class group. By Theorem 10.3 (ii), every ideal class contains an ideal $\mathfrak{a}$ with $N(\mathfrak{a}) \leqslant M_K < 7$. However, the distinct ideals of norm less than or equal to 6 are (1), $\mathfrak{p}$, $\mathfrak{q}_3$, $\mathfrak{q}_3'$, (2), $\mathfrak{q}_5$, $\mathfrak{q}_5'$, $\mathfrak{p}\mathfrak{q}_3$ and $\mathfrak{p}\mathfrak{q}_3'$. Thus the class group has size at most 9, and the only such group with elements of order 2 and 3 is $\mathbf{Z}/6\mathbf{Z}$.

$\mathbf{Q}(\sqrt{-163})$ and the Rabinowitch Phenomenon.

Proposition 11.3. *Let $a \geqslant 2$ be an integer. Let $A := 4a-1$. Then the following three statements are equivalent:*

(i) *x^2+x+a is prime for $0 \leqslant x \leqslant \frac{2}{\pi}\sqrt{a}$;*
(ii) *x^2+x+a is prime for $0 \leqslant x \leqslant a-2$;*
(iii) *$h_{\mathbf{Q}(\sqrt{-A})} = 1$.*

Remark. At first sight[3], the implication (i) $\Rightarrow$ (ii) seems completely remarkable.

Proof. We will show (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii).

To show (i) $\Rightarrow$ (iii), we will try to evaluate the class number h_K, where $K = \mathbf{Q}(\sqrt{-A})$, in the same manner that we did for the examples in Chapter 11. We have $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{-A}}{2}]$, since $-A \equiv 1 \pmod 4$. Thus we also have

[3]Perhaps somewhat disappointingly, a proof *can* be phrased in completely elementary terms, though this is not trivial. See IMO 1987 Question 6.

$\mathcal{O}_K = \mathbf{Z}[\frac{-1+\sqrt{-A}}{2}]$. By Lemma 10.2 (iv) the Minkowski constant M_K is $\frac{2}{\pi}\sqrt{A} < \frac{4}{\pi}\sqrt{a}$. Thus generators of $\mathrm{Cl}(K)$ may be found amongst the (ideal) prime factors of the principal ideals (p), where $p \leqslant \frac{4}{\pi}\sqrt{a}$ is a rational prime.

Let p be such a prime. The minimal polynomial $m(X)$ of $\frac{-1+\sqrt{-A}}{2}$ is $m(X) = X^2 + X + a$. If this has a root $x(\operatorname{mod} p)$ then the other root is $-1 - x \equiv p - 1 - x(\operatorname{mod} p)$, since the sum of the roots is $-1(\operatorname{mod} p)$. Thus $m(X)$, if it has a root mod p, has a root in the range $0, 1, 2, \dots, \frac{1}{2}(p-1)$. Note that $\frac{1}{2}(p-1) < \frac{2}{\pi}\sqrt{a}$. Since we are assuming (i), it follows that x^2+x+a is prime for $x = 0, 1, 2, \dots, \frac{1}{2}(p-1)$, and so the only way it can be $0(\operatorname{mod} p)$ for one of these x is if it equals exactly p. But this is impossible, since $x^2 + x + a \geqslant a$ whilst $p < \frac{4}{\pi}\sqrt{a}$. It follows that $m(X)$ is irreducible $(\operatorname{mod} p)$ and so Dedekind tells us that (p) is inert. That is, all ideals (p) with $p \leqslant \frac{4}{\pi}\sqrt{a}$ are principal and so indeed $\mathrm{Cl}(K)$ is trivial, and so (iii) holds.

Now we show that (iii) $\Rightarrow$ (ii). For this, we more-or-less reverse the above argument. Suppose that $x^2 + x + a$ is *not* prime for some $0 \leqslant x \leqslant a - 2$. On this range, $x^2 + x + a \leqslant (a-2)^2 + (a-2) + a = (a-1)^2 + 1 < a^2$, so $x^2 + x + a$ has a prime factor p with $p < a$. Thus $m(X)$ has a root $(\operatorname{mod} p)$ and so by Dedekind's lemma, (p) splits in $\mathcal{O}_K$ as a product of two ideals of norm p. Since $\mathrm{Cl}(K)$ is trivial, these ideals must be principal. Thus there is some $\alpha \in \mathcal{O}_K$ with $N_{K/\mathbf{Q}}(\alpha) = N((\alpha)) = p$. Suppose that $\alpha = x + y\frac{1+\sqrt{-A}}{2}$, with $x, y \in \mathbf{Z}$. Then $p = N_{K/\mathbf{Q}}(\alpha) = x^2 + xy + ay^2$. Obviously p is not a square, and so $y \neq 0$. Therefore

$$p = x^2 + xy + ay^2 = (x + \frac{y}{2})^2 + A(\frac{y}{2})^2 \geqslant \frac{A}{4} > a - 1.$$

But $p < a$, and so this is a contradiction. $\square$

It is now rather easy to check (using (i)) that $h_{\mathbf{Q}(\sqrt{-A})} = 1$ for the following values of A: $A = 11, 19, 43, 67, 163$. The last of these implies (by (ii)) the famous fact, observed by Euler, that x^2+x+41 is prime for $x = 0, 1, \dots, 39$.

A much deeper fact (the solution of the so-called "class number one problem") is that there are no larger values of A with this property.

12. An elliptic curve

We look at an example of how to use the ideas of the course to solve a specific diophantine equation, specifically to find all the integral points on a certain cubic curve (elliptic curve). The example is somewhat similar to the equation $y^2 + 2 = x^3$ considered by Fermat and Euler, which we solved in Theorem 3.5. However, in this example unique factorisation fails.

Proposition 12.1. *There are no integer solutions to $y^2 + 37 = x^3$.*

Proof. Let $K = \mathbf{Q}(\sqrt{-37})$. It turns out that $h_K = 2$; this is a question on Sheet 4. In particular, $\mathcal{O}_K$ does not have unique factorisation.

The argument closely parallels the proof of Theorem 3.5, but we cannot use unique factorsation.

The equation factors in $\mathcal{O}_K$ as $(y + \sqrt{-37})(y - \sqrt{-37}) = x^3$. We do not have unique factorisation into elements of $\mathcal{O}_K$, only into ideals, so we think of this as an equation

$$(y + \sqrt{-37})(y - \sqrt{-37}) = (x)^3 \tag{12.1}$$

of *ideals*.

We are going to prove that the two ideals on the left are coprime. Suppose some prime ideal $\mathfrak{p}$ divides both terms on the LHS. Then $y + \sqrt{-37}, y - \sqrt{-37} \in \mathfrak{p}$, and so, taking the difference, $2\sqrt{-37} \in \mathfrak{p}$. Therefore $\mathfrak{p} \mid (2\sqrt{-37})$. (Here, of course, we are using the fact that containment and division of ideals are the same thing, Theorem 5.2.)

Taking norms, we have

$$N(\mathfrak{p}) \mid N(2\sqrt{-37}) = 2^2 \cdot 37. \tag{12.2}$$

Also, since $\mathfrak{p} \mid (y + \sqrt{-37})$, we have $\mathfrak{p} \mid (x)^3$ and so

$$N(\mathfrak{p}) \mid N((x)^3) = x^6. \tag{12.3}$$

We claim that neither 2 nor 37 divides x.

If $2 \mid x$ then $8 \mid x^3$, so $y^2 = x^3 - 37 \equiv 3 \pmod 4$, a contradiction.

If $37 \mid x$ then $37 \mid y$, and so $37^2 \mid x^3 - y^2 = 37$. This is also a contradiction.

From these facts and (12.2), (12.3) we have $N(\mathfrak{p}) = 1$, which is impossible; therefore we are forced to conclude that $\mathfrak{p}$ does not exist, so the ideals $(y + \sqrt{-37}), (y - \sqrt{-37})$ are indeed coprime.

Now we return to (12.1). By unique factorisation of ideals, both $(y + \sqrt{-37})$ and $(y - \sqrt{-37})$ are cubes of ideals. Suppose that $(y + \sqrt{-37}) = \mathfrak{a}^3$. In particular, $[\mathfrak{a}]^3$ is trivial in the class group. However, we know that $h_K = 2$, that is to say the class group has order 2. Therefore $[\mathfrak{a}]$ must itself be trivial, or in other words $\mathfrak{a}$ is a principal ideal. Thus we have an equation

$$(y + \sqrt{-37}) = (a + b\sqrt{-37})^3$$

for some $a, b \in \mathbf{Z}$. This means that

$$y + \sqrt{-37} = u(a + b\sqrt{-37})^3$$

in $\mathcal{O}_K$, where u is a unit. The only units are ± 1; by replacing a, b with $-a, -b$ if necessary, we may in fact assume that $u = 1$. Expanding out and comparing coefficients of $\sqrt{-37}$ (which, of course, is irrational) we obtain

$$y = a(a^2 - 111b^2), \quad b(3a^2 - 37b^2) = 1.$$

The second of these implies that $b = \pm 1$ and hence that $3a^2 - 37 = \pm 1$, which is obviously impossible. This concludes the proof. □

Remarks. This was an exam question in 2005, and the fact that $h_K = 2$ was given. In addition to the questions on the example sheets you may wish to try using similar techniques to find all solutions to $y^2 + 54 = x^3$. Unlike the example we went over in detail, this equation does have some solutions.

13. The case $n = 3$ of Fermat's last theorem

Our aim in this chapter is to prove the following famous result.

Theorem 13.1 (Euler)**.** *There is no nontrivial integer solution to the equation*

$$x^3 + y^3 + z^3 = 0. \tag{13.1}$$

That is, every solution to this equation has $xyz = 0$.

We begin with some preliminary comments. First of all, let $\omega := e^{2\pi i/3}$ be a primitive third root of unity. Then the equation factors as

$$(x + y)(x + \omega y)(x + \omega^2 y) = (-z)^3, \tag{13.2}$$

and therefore it is not very surprising that we will be working in the field $\mathbf{Q}(\omega)$. Observe that in fact $\omega = \frac{1}{2}(-1+\sqrt{-3})$, so $K = \mathbf{Q}(\omega)$ is the quadratic field $\mathbf{Q}(\sqrt{-3})$ and the ring of integers is $\mathbf{Z}[\omega]$. We will show the more general result that (13.1) has no nontrivial solutions in $\mathbf{Z}[\omega]$.

We begin by assembling some basic facts about $\mathbf{Z}[\omega]$. We leave it to the reader to check using the methods of Chapter 11 that the class number h_K is one (in fact, this is easier than all of the examples presented there; since $\mathcal{O}_K$ is also a Euclidean domain, you may also have done this in *Rings and Modules*). Thus $\mathbf{Z}[\omega]$ is a unique factorisation domain. In particular, primes and irreducibles are the same thing. We remark that there are six units in $\mathbf{Z}[\omega]$, namely $\{\pm 1, \pm\omega, \pm\omega^2\}$: this is easily seen by noting that $N_{K/\mathbf{Q}}(a+b\omega) = a^2 - ab + b^2$.

Next we introduce an important element of $\mathbf{Z}[\omega]$, namely $\lambda = \sqrt{-3}$. In the argument, we will be working "mod λ", where $\lambda = \sqrt{-3}$. Note that λ is prime, since $N_{K/\mathbf{Q}}(\lambda) = 3$ is prime. The main reason for this is that cubes have very special behaviour modulo powers of λ, as the following lemma (which generalises the fact that $m^3 \in \{0, \pm 1\} \pmod 9$ for $m \in \mathbf{Z}$) shows.

Lemma 13.2. *Suppose that $x \in \mathbf{Z}[\omega]$ is coprime to λ. Then $x^3 \equiv \pm 1 \pmod 9$.*

Proof. We work modulo λ. Note that $9 = \lambda^4$. Since $N((\lambda)) = N_{K/\mathbf{Q}}(\lambda) = 3$, the quotient $\mathbf{Z}[\omega]/(\lambda)$ has size three. The three equivalence classes are represented by $0, 1, -1$, which are mutually incongruent mod λ. Thus $x \equiv \pm 1 \pmod \lambda$. Suppose $x = \pm 1 + \lambda a$ for some $a \in \mathbf{Z}[\omega]$. Then

$$x^3 = \pm 1 - a\lambda^3 \mp a^2\lambda^4 + a^3\lambda^3 \equiv \pm 1 + (a^3 - a)\lambda^3 \pmod 9.$$

However, $a^3 \equiv a \pmod \lambda$, since a is congruent to one of $0, \pm 1 \pmod \lambda$. The result follows. □

Proof. [Proof of Theorem 13.1]. Suppose there is a nontrivial solution to (13.1), with $x, y, z \in \mathbf{Z}[\omega]$. We may divide out by common factors and thereby assume that x, y, z have no common factor. This means that x, y, z must in fact be pairwise coprime, since if some prime γ were to divide x, y (say) then γ would divide $z^3 = -x^3 - y^3$ and hence z. Note also that at least one (and hence precisely one) of x, y, z must be divisible by the prime λ: indeed, working mod λ and applying Lemma 13.2, we see that if this were not the case then $x^3 + y^3 + z^3 \in \{\pm 1, \pm 3\} \pmod 9$. Without loss of generality, $\lambda | z$. We may remove the factors of λ from z to get a nontrivial solution to the equation

$$x^3 + y^3 + \lambda^{3n} z^3 = 0, \tag{13.3}$$

where now x, y, z are pairwise coprime and none is divisible by λ, and $n \geqslant 1$. Consider the slightly more general equation

$$x^3 + y^3 = u\lambda^{3n} z^3, \tag{13.4}$$

where u is one of the six units in $\mathbf{Z}[\omega]$. Let $P(n)$ denote the statement that this equation has no solution in coprime elements $x, y, z \in \mathbf{Z}[\omega]$. By the above discussion, if we know $P(n)$ for all $n \geqslant 1$ then Theorem 13.1 follows. We will now show $P(1)$, and that $P(n-1) \Rightarrow P(n)$. As the reader will see, the argument requires us to work with (slightly) more general equation (13.4), rather than just (13.3).

Proof of $P(1)$. Again, we work modulo λ. By Lemma 13.2, $x^3 + y^3 \in \{0, \pm 2\} (\mathrm{mod}\, \lambda^4)$, thus the power of λ dividing $x^3 + y^3$ is either 0 or at least 4. However, the power of λ dividing $u\lambda^3 z^3$ is 3. This is a contradiction.

The inductive step. Suppose now that $n \geqslant 2$, and suppose we have established $P(n-1)$. Suppose $P(n)$ is false, thus (13.4) has a solution in coprime elements $x, y, z \in \mathbf{Z}[\omega]$. Finally we use the factorisation of the LHS of (13.4), so the equation becomes

$$(x+y)(x+\omega y)(x+\omega^2 y) = u\lambda^{3n} z^3. \tag{13.5}$$

Evidently, this means that λ divides one of the factors on the LHS. However, if it divides one of them, then it divides all of them: this is because $1-\omega$ and $1-\omega^2$ are associates of λ (in fact, $\lambda = \omega(1-\omega) = (-\omega^2)(1-\omega^2)$). Moreover, λ is the *only* common factor of each pair of factors on the LHS of (13.5). For instance, if δ divides $x+y$ and $x+\omega y$ then it also divides $(\omega - 1)y = (x+\omega y) - (x+y)$ and $(1-\omega)x = (x+\omega y) - \omega(x+y)$. Since x and y are coprime, we have $\delta \mid \omega - 1$ and so $\delta \mid \lambda$. Thus (13.5) becomes

$$(\frac{x+y}{\lambda})(\frac{x+\omega y}{\lambda})(\frac{x+\omega^2 y}{\lambda}) = u\lambda^{3n-3} z^3,$$

with the three factors on the left being coprime elements of $\mathbf{Z}[\omega]$.

The power λ^{3n-3} still divides the LHS. Since the three factors on the LHS are coprime, it divides one of them. Replacing y with ωy or $\omega^2 y$ if necessary, we may assume that $\lambda^{3n-3} \mid \frac{x+y}{\lambda}$, and so our equation now becomes

$$(\frac{x+y}{\lambda^{3n-2}})(\frac{x+\omega y}{\lambda})(\frac{x+\omega^2 y}{\lambda}) = uz^3,$$

with the three terms on the left being coprime elements of $\mathbf{Z}[\omega]$.

Using the fact that $\mathbf{Z}[\omega]$ is a UFD, and considering prime factorisations, this implies that we have

$$x + y = \lambda^{3n-2} u_1 z_1^3, \quad x + \omega y = \lambda u_2 z_2^3, \quad x + \omega^2 y = \lambda u_3 z_3^3,$$

where the u_i are units and the z_i are coprime elements of $\mathbf{Z}[\omega]$, none divisible by λ (and $u_1 u_2 u_3 = u$, $z_1 z_2 z_3 = z$, but we will not need this). Since $(x+y) + \omega(x+\omega y) + \omega^2(x+\omega^2 y) = 0$, we have (after a little rearrangement)

$$(x')^3 + \mu(y')^3 = u'\lambda^{3(n-1)}(z')^3, \tag{13.6}$$

where $x' = z_2$, $y' = z_3$, $\mu = \omega u_3/u_2$, $z' = z_1$ and $u' = -u_1/\omega u_2$.

This is almost of the form (13.4), with n replaced by $n-1$, except for the unit μ. To say more about μ, we again work mod λ. The RHS of (13.6) is divisible by λ^3 (since $n \geqslant 2$) whereas, by Lemma 13.2, the LHS is $\pm 1 \pm \mu (\operatorname{mod} \lambda^3)$. It follows that $\mu \equiv \pm 1 (\operatorname{mod} \lambda^3)$. However, μ is one of the six units $\{\pm 1, \pm\omega, \pm\omega^2\}$, and of these only ± 1 are congruent to $\pm 1 (\operatorname{mod} \lambda^3)$, an easy check. Thus $\mu \in \{\pm 1\}$, and so finally we may rewrite (13.6) as

$$(x')^3 + (\mu y')^3 = u' \lambda^{3(n-1)} (z')^3.$$

By the assumption $P(n-1)$, such a solution cannot exist. $\square$

14. Unsolved problems

There are very many quite basic unsolved problems about number fields, easily stated with the language we have developed in this course.

For instance

- It is not known if there are infinitely many real quadratic fields $\mathbf{Q}(\sqrt{d})$ whose rings of integers are UFDs, although it is conjectured (and supported by numerical evidence) that as d ranges over primes, more than 75% of them are.
- It *is* known that there are only nine imaginary quadratic fields $\mathbf{Q}(\sqrt{d})$, $d < 0$, whose rings of integers are UFDs, but this was only proven in the 1960s. The largest of them is $\mathbf{Q}(\sqrt{-163})$. (Note that we did show that the ring of integers of this field is a UFD, but we certainly did not show it is the biggest such field.) It is also known that the class number of $\mathbf{Q}(\sqrt{d})$ tends to infinity as $d \to -\infty$, but the question of exactly how quickly is related to notorious questions in analytic number theory, connected with the generalised Riemann Hypothesis.
- Even less about unique factorisation is known for fields of degree $\geqslant 3$.
- As we saw in the notes, the classification of quadratic fields is quite straightforward. Cubic fields already present significant computational challenges. It turns out that even roughly counting how many fields there are of a given degree is an unsolved problem in general. It is conjectured that the number of number fields with degree n and discriminant at most X grows like a linear function $c_n X$. This is easily checked for $n = 2$. The case $n = 3$ was established by

Davenport and Heilbronn in the 1970s, and the cases $n = 4$ and 5 only in the last fifteen years or so, by Bhargava. All cases with $n \geqslant 6$ are open.

15. *Quadratic forms and the class group

Throughout this chapter, let K be *imaginary* quadratic field with ring of integers $\mathcal{O}_K$ and discriminant Δ. Our aim is to describe a beautiful connection between the ideal class group of such fields and binary quadratic forms. One application of this is an algorithm for computing class numbers h_K.

From ideal classes to $\Gamma \setminus \mathbf{H}$. *Upper half-plane.* The upper half plane $\mathbf{H}$ is defined to be $\{z \in \mathbf{C} : \operatorname{Im} z > 0\}$. The group

$$\mathrm{SL}_2(\mathbf{R}) = \{g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a, b, c, d \in \mathbf{R}, \det g = 1\}$$

acts on $\mathbf{H}$ via Möbius transformations, thus

$$gz := \frac{az + b}{cz + d}.$$

(This is a simple exercise, if you have not seen it before.)

Modular group. Inside $\mathrm{SL}_2(\mathbf{R})$ sits the *modular group*

$$\Gamma := \mathrm{SL}_2(\mathbf{Z}) = \{\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a, b, c, d \in \mathbf{Z}, \det \gamma = 1\}.$$

Of course, this also acts on $\mathbf{H}$ via Möbius transformations.

By $\Gamma \setminus \mathbf{H}$ we mean the set of orbits for this action, that is to say the set of all $\Gamma z = \{\gamma z : \gamma \in \Gamma\}$, as z ranges over $\mathbf{H}$.

There is a famous picture, Figure 15, of this action. The shaded region depicts a fundamental domain $\mathcal{F}$, that is to say a region containing precisely one point of each orbit. We will define $\mathcal{F}$ carefully in Section 15 below. Thus $\Gamma \setminus \mathbf{H}$ may be identified with $\mathcal{F}$.

In Lemma 15.1 below, we are going to associate a point in $\Gamma \setminus \mathbf{H}$ to each ideal class in $\mathcal{O}_K$. However the discussion is cleaner if, instead of ideals, we work with the group $\mathrm{Div}(\mathcal{O}_K)$ of *fractional* ideals. These were (briefly) introduced in Chapter 10. The reader should recall the discussion there. The reader should additionally check that

- the norm function on ideals extends uniquely to a multiplicative function $N : \mathrm{Div}(\mathcal{O}_K) \to \mathbf{Q}$;
- every fractional ideal $\mathfrak{a}$ has an integral basis, that is to say is of the form $\mathbf{Z}e_1 \oplus \mathbf{Z}e_2$ for some $e_1, e_2 \in \mathfrak{a}$.

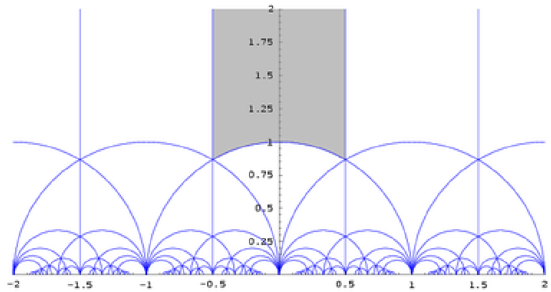

FIGURE 1. Fundamental domain for the action of Γ on $\mathbf{H}$

By an ideal class we mean an element of $\mathrm{Div}(\mathcal{O}_K)/K^*$ (the fractional ideals modulo the principal fractional ideals) which, as remarked in Chapter 10, is isomorphic to the class group $\mathrm{Cl}(K)$. In fact, many texts take this as the definition of the class group.

Lemma 15.1. *We have the following.*

(i) *Every ideal class contains a fractional ideal of the form* $\mathbf{Z} \oplus \mathbf{Z}\tau$ *with* $\tau \in \mathbf{H}$;

(ii) *Let* $\tau' \in \mathbf{H}$. *Then* $\mathbf{Z} \oplus \mathbf{Z}\tau'$ *is a fractional ideal in the same class as* $\mathbf{Z} \oplus \mathbf{Z}\tau$ *if and only if* $\Gamma\tau' = \Gamma\tau$.

Proof. (i) Suppose that $\mathfrak{a} = \mathbf{Z}e_1 \oplus \mathbf{Z}e_2$ is some fractional ideal in the class. Since K is imaginary, $\mathbf{R} \cap K = \mathbf{Q}$ and so we cannot have $e_1/e_2 \in \mathbf{R}$, since this would entail $e_1/e_2 \in \mathbf{Q}$ and so e_1, e_2 would not generate a free abelian group. By swapping e_1, e_2 if necessary, we may assume that $\tau := e_2/e_1 \in \mathbf{H}$. Then $\frac{1}{e_1}\mathfrak{a} = \mathbf{Z} \oplus \mathbf{Z}\tau$ is in the same (fractional) ideal class as $\mathfrak{a}$.

(ii) Suppose that $\tau' = \gamma\tau$, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Then, since γ is unimodular it follows from Proposition 2.13 that

$$\mathbf{Z} \oplus \mathbf{Z}\tau = \mathbf{Z}(c\tau + d) \oplus \mathbf{Z}(a\tau + b) = (c\tau + d)(\mathbf{Z} \oplus \mathbf{Z}\tau').$$

It follows that $\mathbf{Z} \oplus \mathbf{Z}\tau'$ is a fractional ideal, in the same class as $\mathbf{Z} \oplus \mathbf{Z}\tau$.

Conversely, suppose that $\mathbf{Z} \oplus \mathbf{Z}\tau' = (\alpha)(\mathbf{Z} \oplus \mathbf{Z}\tau) = \mathbf{Z}\alpha \oplus \mathbf{Z}\alpha\tau$, for some $\alpha \in K$. It follows from Proposition 2.13 that $1, \tau'$ and $\alpha, \alpha\tau$ are related by a unimodular transformation, thus

$$1 = \alpha(c\tau + d),$$

$$\tau' = \alpha(a\tau + b)$$

for some unimodular $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Thus $\tau' = \frac{a\tau+b}{c\tau+d}$. We must in fact have $ad - bc = 1$ (rather than -1) or else τ' would lie in the *lower* half plane. $\square$

Definition 15.2. Write $\mathbf{H}(K)$ for the set of all $\tau \in \mathbf{H}$ for which $\mathbf{Z} \oplus \mathbf{Z}\tau$ is a fractional ideal in K. These are called the Heegner points for K.

In this language, Lemma 15.1 shows that $\mathbf{H}(K)$ is a union of Γ-orbits, and the number of such orbits is precisely the class number h_K. That is,

$$|\Gamma \setminus \mathbf{H}(K)| = h_K. \tag{15.1}$$

Quadratic forms from points of H. By a *positive definite binary quadratic form* over $\mathbf{R}$ we mean $q(\mathbf{x}) = ax_1^2 + bx_1x_2 + cx_2^2$, with $a, b, c \in \mathbf{R}$, $a > 0$ and the *discriminant* $\mathrm{Disc}(q) := b^2 - 4ac$ negative. (We observe that this is the third distinct way in which we have used the word discriminant, but it will be linked to the other ones shortly.)

There is a very natural correspondence between points $\tau \in \mathbf{H}$ and positive definite binary quadratic forms over $\mathbf{R}$ of a fixed discriminant $D < 0$.

To a point $\tau \in \mathbf{H}$, we associate

$$q_{\tau;D}(\mathbf{x}) := \frac{\sqrt{-D}}{2\operatorname{Im}\tau}(x_1 - \tau x_2)(x_1 - \overline{\tau} x_2). \tag{15.2}$$

One may easily check that $\mathrm{Disc}(q_\tau) = D$.

Conversely, given q of discriminant D, we may recover τ as the unique element of $\mathbf{H}$ such that $q(\tau, 1) = 0$, i.e.

$$\tau = \frac{-b + \sqrt{D}}{2a},$$

where the square-root is a positive multiple of i. We refer to τ as the *root* of q.

As we have seen, the group $\mathrm{SL}_2(\mathbf{R})$ acts on $\mathbf{H}$ by Möbius transformations. It also acts on $\mathbf{C}^2$ in the usual linear way, that is to say if $g = \left(\begin{smallmatrix} g_{11} & g_{12} \\ g_{21} & g_{22}\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{R})$, then $g\mathbf{x} = (g_{11}x_1 + g_{12}x_2, g_{21}x_1 + g_{22}x_2)$. These actions are related in the following way, where we write elements of $\mathbf{C}^2$ as row vectors:

$$g(\tau, 1) = (g_{11}\tau + g_{12}, g_{21}\tau + g_{22}) = (g_{21}\tau + g_{22})(g\tau, 1). \tag{15.3}$$

The action of $\mathrm{SL}_2(\mathbf{R})$ on $\mathbf{R}^2$ gives rise to a (right-) action of $\mathrm{SL}_2(\mathbf{R})$ on quadratic forms of any given discriminant D via $(gq)(\mathbf{x}) = q(g^{-1}\mathbf{x})$. To see that the discriminant is preserved, note that if $q(\mathbf{x}) = x^T M \mathbf{x}$ with M symmetric then $\mathrm{Disc}(q) = -4\det M$. We have $(gq)(\mathbf{x}) = q(g^{-1}x) = \mathbf{x}^T g^{-T} M g^{-1}\mathbf{x}$, and so since $\det g = 1$

$$\mathrm{Disc}(gq) = -4\det(g^{-T} M g^{-1}) = -4\det M = \mathrm{Disc}(q).$$

Lemma 15.3. *Let $\tau \in \mathbf{H}$ and let $D < 0$ be arbitrary. Then we have $gq_{\tau;D} = q_{g\tau;D}$. That is, the $\mathrm{SL}_2(\mathbf{R})$-actions on $\mathbf{H}$ and on quadratic forms of discriminant D are the same under the correspondence between these two sets.*

Proof. It suffices to check that $g\tau$ is the root of gq_τ. But, by (15.3),

$$(g_{21}\tau + g_{22})^2(gq_{\tau;D})(g\tau, 1) = gq_{\tau;D}\big(g(\tau,1)\big) = q_{\tau;D}(\tau, 1) = 0.$$

This completes the proof. □

Action of $\mathrm{SL}_2(\mathbf{Z})$ and reduction theory. We saw in the last section that for any $D < 0$ there is a natural correspondence

$$\mathbf{H} \longleftrightarrow \text{positive definite quadratic forms of discriminant } D,$$

and that moreover this intertwines two natural actions of $\mathrm{SL}_2(\mathbf{R})$, the left action on $\mathbf{H}$ given by Möbius transformations, and the right action on quadratic forms given by $(gq)(\mathbf{x}) = q(g^{-1}\mathbf{x})$.

In this section we specialise this to the action of the modular group Γ.

Define

$$\mathcal{F} := \{\tau \in \mathbf{H} : -\frac{1}{2} \leqslant \operatorname{Re}\tau < \frac{1}{2}, |\tau| > 1\} \cup \{\tau \in \mathbf{H} : -\frac{1}{2} \leqslant \operatorname{Re}\tau \leqslant 0, |\tau| = 1\}.$$

Thus $\mathcal{F}$ is the shaded area in Figure 15 (but we have been precise about what the boundary is).

Lemma 15.4. *$\mathcal{F}$ is a* fundamental domain *for the action of Γ on $\mathbf{H}$: every $z \in \mathbf{H}$ is in the Γ-orbit of precisely one point of $\mathcal{F}$. Thus we can identify $\mathcal{F}$ with $\Gamma \backslash \mathbf{H}$.*

Proof. First note that if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ then $\operatorname{Im}(\gamma\tau) = |c\tau + d|^{-2}\operatorname{Im}\tau$. As c, d range over integers, $|c\tau + d|$ attains its minimum value, and so in any Γ-orbit there is τ with $\operatorname{Im}\tau$ maximal. Consider the elements $S := \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and $T := \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ of Γ. These act on $\mathbf{H}$ by inversion and translation respectively, that is to say $Sz = -1/z$, $Tz = z + 1$. Thus, applying a suitable power of T, we may additionally assume not only that $\operatorname{Im}\tau$ is maximal but also that $-\frac{1}{2} \leqslant \operatorname{Re}\tau < \frac{1}{2}$. Since $\operatorname{Im}\tau$ *is* maximal, $\operatorname{Im}(S\tau) \leqslant \operatorname{Im}(\tau)$, and this immediately implies that $|\tau| \geqslant 1$, so τ lies in the set

$$\tilde{\mathcal{F}} := \{\tau \in \mathbf{H} : -\frac{1}{2} \leqslant \operatorname{Re}\tau < \frac{1}{2}, |\tau| \geqslant 1\}.$$

Moreover if $|\tau| = 1$ and $0 < \operatorname{Re}\tau < \frac{1}{2}$ then $|S\tau| = 1$ and $-\frac{1}{2} < \operatorname{Re}(S\tau) < 0$. It follows that every element of $\tilde{\mathcal{F}}$ is Γ-equivalent to a point of $\mathcal{F}$.

The proof that different points of $\mathcal{F}$ are inequivalent under Γ is straightforward but somewhat tedious; I will probably go over it quickly in lectures. Suppose as a hypothesis for contradiction that $\tau, \gamma\tau \in \mathcal{F}$ are distinct points, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Without loss of generality (replacing γ by γ^{-1} if necessary) we may assume that $\operatorname{Im}(\gamma\tau) \geqslant \operatorname{Im}\tau$, which means that

$$|c\tau + d| \leqslant 1. \tag{15.4}$$

Taking imaginary parts, we have $|c \operatorname{Im}\tau| \leqslant 1$ which, since $|\operatorname{Im}\tau| \geqslant \frac{1}{2}\sqrt{3}$, means that $c \in \{-1, 0, 1\}$. Taking real parts, we have $\operatorname{Re}(c\tau + d) \leqslant 1$ and so $|d| \leqslant 1 + \frac{1}{2}|c|$ and so $d \in \{-1, 0, 1\}$ as well.

Case $c = 0$. Then $d = \pm 1$. The two cases are similar, so we look at $d = 1$. Then $a = 1$ and $\gamma\tau = \tau + b$. Since $\tau, \gamma\tau \in \mathcal{F}$, taking real parts gives $b = 0$ and so γ is the identity, contrary to the assumption that $\tau, \gamma\tau$ are distinct.

Case $c = \pm 1$. The cases are similar, so suppose that $c = 1$. If $d = 1$ then (15.4) gives $|\tau + 1| \leqslant 1$. The only point of $\mathcal{F}$ with this property is $\tau = \omega = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$. Also $a - b = ad - bc = 1$ and so

$$\gamma\tau = \frac{a\tau + b}{\tau + 1} = -\tau(a\tau + b) = a + (a - b)\tau = a + \tau.$$

This only lies in $\mathcal{F}$ if $a = 0$, and so $\gamma\tau = \tau$, contrary to assumption. The case $d = -1$ is similar. Finally, if $d = 0$ then (15.4) gives $|\tau| \leqslant 1$, and therefore since $\tau \in \mathcal{F}$ we have $|\tau| = 1$. Also, $b = -1$ and $\gamma\tau = a - \frac{1}{\tau} = a - \overline{\tau}$. This only lies in $\mathcal{F}$ if $a = 0$, in which case $\gamma\tau = -\overline{\tau}$. Thus τ and $\gamma\tau$ both lie in $\mathcal{F}$, on $|z| = 1$, and their real parts have opposite signs. This is impossible, and the proof is complete. $\square$

Remark. The proof shows that any point of $\mathbf{H}$ may be moved into $\mathcal{F}$ using elements of $\langle S, T\rangle$. Take $\tau \in \mathcal{F}$ to be a point with trivial Γ-stabiliser (exercise: these exist, and in fact any interior point of $\mathcal{F}$ has this property). Then, for any $\gamma \in \Gamma$, we may find $\gamma' \in \langle S, T\rangle$ such that $\gamma'\gamma z = z$ which, since z has trivial stabiliser, implies that $\gamma \in \langle S, T\rangle$. Thus Γ is generated by S and T.

Definition 15.5. Let $q(x) = ax_1^2 + bx_1x_2 + cx_2^2$ be a positive definite form over $\mathbf{R}$. Then we say that q is *reduced* if $|b| \leqslant a \leqslant c$ and if either $|b| = a$ or $a = c$ then $b \geqslant 0$.

Lemma 15.6. *Let q be a positive definite form of discriminant D. Then its root τ lies in $\mathcal{F}$ if and only if q is reduced.*

Proof. If τ is the root $\frac{-b+\sqrt{D}}{2a}$ of q, then $\mathrm{Re}\,\tau = -b/2a$ and $|\tau|^2 = c/a$, and the lemma is then a quick check. $\square$

As a consequence of Lemmas 15.4 and 15.6 and the fact that the actions of Γ on $\mathbf{H}$ and on quadratic forms are equivalent, we have the following.

Corollary 15.7. *Every Γ-orbit of quadratic forms of discriminant D contains precisely one reduced form.*

We say that two quadratic forms q, q' are *equivalent* if they are in the same Γ-orbit. Thus q, q' are equivalent if and only if there is some $\gamma \in \Gamma$ such that $q'(\mathbf{x}) = q(\gamma\mathbf{x})$.

We can summarise the findings of this section as follows: for each fixed $D < 0$ there is a one-to-one correspondence

$$\begin{aligned} \mathcal{F} \cong \Gamma \setminus \mathbf{H} &\longleftrightarrow \text{equivalence classes of quadratic forms of discriminant } D \\ &\longleftrightarrow \text{reduced quadratic forms of discriminant } D. \end{aligned}$$

Integral binary quadratic forms and Heegner points. The material in the last two sections was purely geometric and contained no number theory. Let us now reintroduce the imaginary quadratic field K, with discriminant Δ. Recall the definition of the set $\mathbf{H}(K)$ of Heegner points (Definition 15.2).

A positive definite binary quadratic form over $\mathbf{R}$ is *integral* if its coefficients a, b, c all lie in $\mathbf{Z}$. It is easy to see that the action of Γ on quadratic forms preserves the property of being integral.

Proposition 15.8. *Let K be an imaginary quadratic field with discriminant Δ. Then correspondence $\tau \leftrightarrow q_{\tau;\Delta}$ of the previous section induces a correspondence between points in $\Gamma \setminus \mathbf{H}(K)$ and equivalence classes of integral quadratic forms of discriminant Δ. In particular, by (15.1) and Corollary 15.7, the class number h_K is precisely the number of reduced integral quadratic forms of discriminant Δ.*

Proof. Suppose first that $\tau \in \mathbf{H}(K)$, that is to say $\mathbf{Z} \oplus \mathbf{Z}\tau$ is a fractional ideal in K. We claim that the quadratic form $q_{\tau;\Delta}$ (as defined in (15.2)) is integral (and, as previously observed, it has discriminant Δ).

To prove this, we slightly rephrase the definition of $q_{\tau;\Delta}$, writing it in terms of objects in the ring of integers $\mathcal{O}_K$. To do this, pick $\alpha \in K$ such that $e_1 := \alpha$ and $e_2 := \alpha\tau$ are both in $\mathcal{O}_K$. Set $\mathfrak{a} := \mathbf{Z}e_1 \oplus \mathbf{Z}e_2$, and note that $\mathfrak{a}$ is an ideal in $\mathcal{O}_K$. We claim that

$$q_{\tau;\Delta}(\mathbf{x}) = \frac{N_{K/\mathbf{Q}}(x_1e_1 - x_2e_2)}{N(\mathfrak{a})}. \tag{15.5}$$

To prove this claim, write $\tilde{q}(\mathbf{x})$ for the right-hand side of (15.5). It is clear that $\tilde{q}(\tau, 1) = 0$, that is to say τ is the root of $\tilde{q}$, and therefore we need only check that $\mathrm{Disc}(\tilde{q}) = \Delta$. (It is easy to see that the root and the discriminant completely determine a binary quadratic form.) One may easily calculate that $\mathrm{Disc}(\tilde{q})$ equals

$$\frac{1}{N(\mathfrak{a})^2}(e_1\overline{e}_2 - \overline{e}_1e_2)^2 = \frac{1}{N(\mathfrak{a})^2}\begin{vmatrix} e_1 & \overline{e}_1 \\ e_2 & \overline{e}_2 \end{vmatrix}^2 = \frac{1}{N(\mathfrak{a})^2}\,\mathrm{disc}_{K/\mathbf{Q}}(e_1, e_2).$$

(Recall the notion of $\mathrm{disc}_{K/\mathbf{Q}}(e_1, e_2)$, as given in Definition 1.19). By Corollary 2.20,

$$\mathrm{disc}_{K/\mathbf{Q}}(e_1, e_2) = [\mathcal{O}_K : \mathfrak{a}]^2\Delta = N(\mathfrak{a})^2\Delta,$$

and so indeed $\mathrm{Disc}(\tilde{q}) = \Delta$. This proves the claim.

Now that we have (15.5), it is easy to check that $q_{\tau;\Delta}(\mathbf{x})$ is integral. We need only show that $N(\mathfrak{a})$ divides $e_1\overline{e}_1 = N_{K/\mathbf{Q}}(e_1)$, $e_2\overline{e}_2 = N_{K/\mathbf{Q}}(e_2)$ and $e_1\overline{e}_2 + \overline{e}_1e_2 = N_{K/\mathbf{Q}}(e_1 + e_2) - N_{K/\mathbf{Q}}(e_1) - N_{K/\mathbf{Q}}(e_2)$. However, for each of $\beta = e_1, e_2, e_1 + e_2$ we have $\beta \in \mathfrak{a}$, and so $\mathfrak{a}|(\beta)$, and therefore $N(\mathfrak{a})|N_{K/\mathbf{Q}}(\beta)$. The integrality of $q_{\tau;\Delta}$ follows.

Now we look at the opposite direction of the correspondence, supposing that $q(\mathbf{x}) = ax_1^2 + bx_1x_2 + cx_2^2$ is an integral binary quadratic form of discriminant Δ. Let $\tau = \frac{-b+\sqrt{\Delta}}{2a}$ be its root. We claim that $\tau \in \mathbf{H}(K)$, to which end we must check that $\alpha(\mathbf{Z} \oplus \mathbf{Z}\tau) \subseteq (\mathbf{Z} \oplus \mathbf{Z}\tau)$, where $\mathcal{O}_K = \mathbf{Z}[\alpha]$. There are two cases.

- Case $K = \mathbf{Q}(\sqrt{d})$, $d \equiv 2, 3 (\mathrm{mod}\, 4)$. Then $\Delta = 4d$ and we can take $\alpha = \sqrt{d}$. Now observe that
$$\alpha = \frac{b}{2} + a\tau, \quad \alpha\tau = -c - \frac{b}{2}\tau.$$
Moreover, $\Delta = b^2 - 4ac \equiv 0 (\mathrm{mod}\, 4)$, so b is even.
- Case $d \equiv 1 (\mathrm{mod}\, 4)$. Then $\Delta = d$ and we can take $\alpha = \frac{1+\sqrt{d}}{2}$. Now observe that
$$\alpha = \frac{1+b}{2} + a\tau, \quad \alpha\tau = -c + \frac{1-b}{2}\tau,$$
and b is odd so $\frac{1\pm b}{2}$ are both integers.

The claim is thus confirmed in all cases, and this completes the proof. □

Example: $\mathbf{Q}(\sqrt{-29})$. Proposition 15.8 gives an algorithmic and calculationally feasible way of calculating h_K when K is an imaginary quadratic field. Consider the particular case $K = \mathbf{Q}(\sqrt{-29})$. Then $\Delta = -116$, and so h_K is the number of reduced integral quadratic forms of discriminant -116.

Let us outline a general strategy for enumerating the reduced integral quadratic forms of discriminant $\Delta < 0$. It is convenient and standard to use the abbreviation (a, b, c) for the form $ax_1^2 + bx_1x_2 + cx_2^2$. We recall, in this notation, the notion of reduced form: (a, b, c) is reduced if we have

- $|b| \leqslant a \leqslant c$; if either $|b| = a$ or $a = c$, then $b \geqslant 0$.

In enumerating the reduced forms, the following simple inequality is very useful.

Lemma 15.9. *Suppose that (a, b, c) is reduced and has discriminant $\Delta = b^2 - 4ac < 0$. Then $a \leqslant \sqrt{|\Delta|/3}$.*

Proof. We have
$$|\Delta| = 4ac - b^2 \geqslant 4a^2 - a^2 = 3a^2,$$
so the result follows immediately. □

When $\Delta = -116$, we get $a \leqslant 6$. Now we simply enumerate:

- $a = 6$. Thus $b^2 = 24c - 116$, and $|b| \leqslant 6$. The only solution is $b = \pm 2$, but this leads to $c = 5$, which is not reduced since $c < a$.
- $a = 5$. Thus $b^2 = 20c - 116$, and $|b| \leqslant 5$. The only solution is $b = \pm 2$, which leads to $c = 6$ and the reduced forms $(5, \pm 2, 6)$.
- $a = 4$. Thus $b^2 = 16c - 116$, and $|b| \leqslant 4$. This has no solutions.
- $a = 3$. Thus $b^2 = 12c - 116$, and $|b| \leqslant 3$. This has the solutions $b = \pm 2$, giving reduced forms $(3, \pm 2, 10)$.
- $a = 2$. Thus $b^2 = 8c - 116$, and $|b| \leqslant 2$. This has the solutions $b = \pm 2$ and $c = 15$. Only $b = 2$ gives a reduced form, namely $(2, 2, 15)$.
- $a = 1$. Thus $b^2 = 4c - 116$, and $|b| \leqslant 1$. The only solution is $b = 0$, giving the reduced form $(1, 0, 29)$.

We have shown that there are six reduced forms of discriminant -116, and this confirms our earlier calculation that $h_{\mathbf{Q}(\sqrt{-29})} = 6$.

Further remarks. We have given a very bare-bones version of the correspondence between class groups and binary quadratic forms. In particular

- We focussed on the imaginary quadratic case, but there is also a theory for real quadratic fields;
- Our focus was on (imaginary quadratic) fields, and so we only considered binary quadratic forms whose discriminant Δ is the discriminant of one of these fields (that is, is either $4d$ for some squarefree $d \equiv 2, 3 \pmod 4$, or d for some squarefree $d \equiv 1 \pmod 4$). Such Δ are called *fundamental discriminants*.

  The discriminant of a binary quadratic form may take any value $D \equiv 0, 1 \pmod 4$, and so need not be a fundamental discriminant. There is a theory covering binary quadratic forms in this generality, requiring one to work with *orders* in quadratic fields rather than just with the rings of integers $\mathcal{O}_K$.

Appendix A. Free abelian groups and lattices

In this chapter we record some basic facts about free abelian groups and lattices.

A *free abelian group* G or rank n is a group of the form $G = \bigoplus_{i=1}^n \mathbf{Z}e_i$, for some $e_1, \dots, e_n$. All such groups are isomorphic, and they are all isomorphic to the "standard lattice" $\mathbf{Z}^n \subseteq \mathbf{R}^n$. The following is the key result about free abelian groups.

Proposition A.1. *Let $G = \bigoplus_{i=1}^n \mathbf{Z}e_i$ be a free abelian group of rank n. If $H \leqslant G$ is a finite index subgroup, H is also a free abelian group of rank n, that is to say $H = \bigoplus_{i=1}^n \mathbf{Z}e'_i$ with $e'_i \in G$. Suppose that $e'_i = \sum_j A_{ji}e_j$. Then $[G : H] = |\det A|$.*

Proof. This is non-examinable. I may write my own exposition of the proof here, but for now you may consult Stewart and Tall, Chapter 1. $\square$

Definition A.2 (Lattice)**.** A lattice $\Lambda \subset \mathbf{R}^n$ is a subgroup of the form $\Lambda = \bigoplus_{i=1}^n \mathbf{Z}e_i = \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n$, where $e_1, \dots, e_n \in \mathbf{R}^n$ are linearly independent vectors.

Remark. There are other, equivalent, ways to define what it means to be a lattice. For example, a lattice is the same things as a discrete, cocompact subgroup of $\mathbf{R}^n$. We will not prove the equivalence of these definitions here.

Definition A.3. The *determinant* of a lattice, $\det(\Lambda)$, is $|\det(e_1, \dots, e_n)|$.

Remark. We use the absolute values since otherwise $\det(\Lambda)$ is only defined up to sign, depending on the ordering of the e_i.

Lemma A.4. *The determinant* $\det(\Lambda)$ *depends only on* Λ*, and not on the particular choice of* e_i*.*

Proof. Suppose that $e_1', \dots, e_n'$ is another basis for the lattice, and suppose that $e_i' = \sum_j A_{ji} e_j$. Then $\bigoplus \mathbf{Z} e_i' = \bigoplus \mathbf{Z} e_i$. We saw in the main text that this is the case if and only if A is unimodular, that is to say $A \in \mathrm{Mat}_n(\mathbf{Z})$ and $\det A = \pm 1$.

However we have

$$\det(e_1', \dots, e_n') = \det A \det(e_1, \dots, e_n),$$

and so

$$|\det(e_1' \dots, e_n')| = |\det(e_1, \dots, e_n)|.$$

This completes the proof. □

Suppose that $\Lambda = \bigoplus_{i=1}^n \mathbf{Z} e_i$ is a lattice. Then the region

$$\mathcal{F} := \{x_1 e_1 + \cdots + x_n e_n : 0 \leqslant x_i < 1 \text{ for } i = 1, \dots, n\}$$

is called a *fundamental region* or *fundamental parallelepiped* for Λ. Note that translates of $\mathcal{F}$ by Λ tile $\mathbf{R}^n$ perfectly, that is to say $\mathcal{F} + \Lambda = \mathbf{R}^n$ with each point represented uniquely.

Note that $\mathcal{F}$ depends on the choice of basis $e_1, \dots, e_n$ for Λ; different choices will give different fundamental regions.

It is well-known that the volume of the parallelepiped $\mathcal{F}$ is $|\det(e_1, \dots, e_n)|$. (The reader may, however, wish to reflect on the fact that a proper and careful discussion of this leads to foundational issues in linear algebra and measure theory.) Let us record this as a lemma.

Lemma A.5. *Let* $\mathcal{F}$ *be a fundamental region for* Λ*. Then* $\mathrm{vol}(\mathcal{F}) = \det(\Lambda)$*.*

Note that $\mathcal{F}$ is a (particularly nice) set of representatives for $\mathbf{R}^n/\Lambda$, and so one sometimes sees the above written as

$$\det(\Lambda) = \mathrm{vol}(\mathbf{R}^n/\Lambda). \tag{A.1}$$

Consequently the determinant of Λ is sometimes referred to as the *covolume* of Λ.

Lemma A.6. *Suppose that Λ is a lattice in $\mathbf{R}^n$ and that Λ' is a finite index subgroup of Λ. Then Λ' is a lattice, and $[\Lambda : \Lambda'] = \det(\Lambda')/\det(\Lambda)$, where (as usual) $[\Lambda : \Lambda']$ denotes the index of Λ' as a subgroup of Λ.*

Proof. That Λ' is a lattice follows from Proposition A.1. Suppose that a basis for Λ is $e_1, \dots, e_n$, and that a basis for Λ' is $e'_1, \dots, e'_n$. Since $\Lambda' \subseteq \Lambda$, we have $e'_i = \sum_j A_{ji} e_j$ for some $A \in \mathrm{Mat}_n(\mathbf{Z})$. By Proposition A.1, $[\Lambda : \Lambda'] = |\det A|$. However we also have

$$\det(e'_1, \dots, e'_n) = \det A \det(e_1, \dots, e_n),$$

and so

$$\det(\Lambda') = |\det A| \det(\Lambda).$$

Combining these facts concludes the proof. □

Note that Lemma A.6 is very natural when interpreted in terms of covolumes. Indeed, if $\Lambda = \bigoplus_{i=1}^m (x_i + \Lambda')$, where $m = [\Lambda : \Lambda']$, and if $\mathcal{F}$ is a fundamental domain for Λ, then $\bigcup_{i=1}^m (x_i + \mathcal{F})$ is a set of representatives for $\mathbf{R}^n/\Lambda'$, so we see that $\mathrm{vol}(\mathbf{R}^n/\Lambda') = m \,\mathrm{vol}(\mathbf{R}^n/\Lambda)$. This does assume, however, that these quantities are well-defined, and it is not really natural to prove a basic algebraic result using (implicitly) the construction of Lebesgue measure.

Appendix B. Geometry of numbers

In this section we give the proof of Minkowski's first theorem, the key ingredient in the proof of the Minkowski bound. Let us begin by recalling the statement.

Theorem 10.6. *Suppose that $\Lambda \subseteq \mathbf{R}^n$ is a lattice, and that $B \subset \mathbf{R}^n$ is a centrally symmetric, compact, convex body. Suppose that $\mathrm{vol}(B) \geqslant 2^n \det(\Lambda)$. Then B contains a nonzero point of Λ.*

It is convenient to prove the following variant which has no compactness assumption and a slightly weaker conclusion. (One could also use this version directly in the main text.)

Theorem B.1 (Minkowski)**.** *Suppose that $\Lambda \subseteq \mathbf{R}^n$ is a lattice, and that $B \subset \mathbf{R}^n$ is a centrally symmetric convex body. Suppose that $\mathrm{vol}(B) > 2^n \det(\Lambda)$. Then B contains a nonzero point of Λ.*

Theorem 10.6 follows from Theorem B.1 by a compactness argument, which we quickly sketch. Let assumptions be as in Theorem 10.6. For any ε, $0 < \varepsilon < 1$, consider the dilate $(1+\varepsilon)B$. This is centrally symmetric and convex, and has volume $(1+\varepsilon)^n \operatorname{vol}(B) > \operatorname{vol}(B)$. By Theorem B.1, $(1+\varepsilon)B$ contains a nonzero point $\lambda_\varepsilon \in \Lambda$. All of these points lie in $2B$, which is a bounded subset of $\mathbf{R}^n$, and hence contains only finitely many points of Λ. Thus as ε varies there are only finitely many different points λ_ε. In particular, there is some sequence of $\varepsilon \to 0$ such that $\lambda_\varepsilon = \lambda$ does not depend on ε. Since B is closed and $\lambda \in (1+\varepsilon)B$ for arbitrarily small ε, $\lambda \in B$.

Theorem B.1 is an easy consequence of the following result called *Blichfeldt's lemma*. Note that in this lemma there are no assumptions such as convexity or central symmetry.

Lemma B.2 (Blichfeldt's lemma)**.** *Suppose that $K \subset \mathbf{R}^n$, and suppose that* $\operatorname{vol}(K) > \det(\Lambda)$*. Then there are two distinct points $x, y \in K$ with $x - y \in \Lambda$.*

Proof. For each $\lambda \in \Lambda$, define $K_\lambda := (K - \lambda) \cap \mathcal{F}$. Then the translates $K_\lambda + \lambda$ tile K and so

$$\sum_\lambda \operatorname{vol}(K_\lambda) = \operatorname{vol}(K). \tag{B.1}$$

Suppose that there do not exist distinct points $x, y \in K$ whose difference lies in Λ. Then the K_λ are all disjoint. Since they all lie in $\mathcal{F}$, we therefore have

$$\sum_\lambda \operatorname{vol}(K_\lambda) \leqslant \operatorname{vol}(\mathcal{F}) = \det \Lambda. \tag{B.2}$$

Comparing (B.1) and (B.2), the result follows. □

Proof. [Proof of Theorem B.1] Let B be as in the statement of Theorem B.1, that is to say B is convex, centrally symmetric and $\operatorname{vol}(B) > 2^n \det(\Lambda)$. Set $K := \frac{1}{2}B = \{\frac{1}{2}x : x \in \mathbf{R}^n\}$. Then $\operatorname{vol}(K) = 2^{-n}\operatorname{vol}(B)$, and so $\operatorname{vol}(K) > \det(\Lambda)$. By Blichfeldt's lemma, the set K contains two distinct points whose difference is in Λ; thus there are $x, y \in B$ with $\frac{1}{2}(x-y) \in \Lambda$. However, since B is convex and centrally symmetric we have $\frac{1}{2}(x-y) \in B$.
□

Appendix C. Gauss's Lemma

There are more general versions of Gauss's lemma than the one we are about to state, but this is all we need in the course.

Lemma C.1 (Gauss's lemma)**.** *Let $f(X) \in \mathbf{Z}[X]$ be monic. Suppose that f is reducible in $\mathbf{Q}[X]$. Then f factors into monic polynomials in $\mathbf{Z}[X]$.*

Proof. Take the factorisation of $f(X)$ in $\mathbf{Q}[X]$, and clear denominators. Then we find some positive integer d such that

$$df(X) = g(X)h(X),$$

where $g(X), h(X) \in \mathbf{Z}[X]$. Suppose

$$g(X) = a_0 + a_1 X + \cdots + a_m X^m,$$

$$h(X) = b_0 + b_1 X + \cdots + b_n X^n.$$

Since f is monic, $d = a_m b_n$ and therefore any common factor of the a_i would have to divide d. We may then divide through by such a common factor, and in this way we may suppose that the a_i are coprime, and similarly that the b_j are coprime.

Suppose that $d \neq 1$. Then some prime p divides d. Let i be maximal so that $p \nmid a_i$, and j be maximal so that $p \nmid b_j$. Then the coefficient of X^{i+j} in $g(X)h(X)$ is $a_i b_j + \dots$, where everything in $\dots$ is divisible by p. Thus the coefficient of X^{i+j} in $g(X)h(X)$ is not divisible by p, which is evidently a contradiction since *all* coefficients of $df(X)$ are divisible by p.

Therefore $d = 1$ and the result is proven. □

References

[1] I. N. Stewart and D. O. Tall, *Algebraic number theory.*

[2] *Fermat's proof for $x^3 = y^2 = 2$,* https://mathoverflow.net/questions/142220/fermats-proof-for-x3-y2-2