XIV

Algebraic NT I: Rings of Integers

Part XIV: Contents

53 Algebraic integers

Here's a first taste of algebraic number theory.

This is really close to the border between olympiads and higher math. You've always known that $a+\sqrt{2}b$ had a "norm" a^2-2b^2, and that somehow this norm was multiplicative. You've also always known that roots come in conjugate pairs. You might have heard of minimal polynomials but not know much about them.

This chapter and the next one will make all these vague notions precise. It's drawn largely from the first chapter of [Og10].

§53.1 Motivation from high school algebra

This is adapted from my blog, *Power Overwhelming*[1].

In high school precalculus, you'll often be asked to find the roots of some polynomial with integer coefficients. For instance,

$$x^3 - x^2 - x - 15 = (x-3)(x^2+2x+5)$$

has roots 3, $-1+2i$, $-1-2i$. Or as another example,

$$x^3 - 3x^2 - 2x + 2 = (x+1)(x^2-4x+2)$$

has roots -1, $2+\sqrt{2}$, $2-\sqrt{2}$. You'll notice that the irrational roots, like $-1 \pm 2i$ and $2 \pm \sqrt{2}$, are coming up in pairs. In fact, I think precalculus explicitly tells you that the complex roots come in conjugate pairs. More generally, it seems like all the roots of the form $a + b\sqrt{c}$ come in "conjugate pairs". And you can see why.

But a polynomial like

$$x^3 - 8x + 4$$

has no rational roots. (The roots of this are approximately -3.0514, 0.51730, 2.5341.) Or even simpler,

$$x^3 - 2$$

has only one real root, $\sqrt[3]{2}$. These roots, even though they are irrational, have no "conjugate" pairs. Or do they?

Let's try and figure out exactly what's happening. Let α be any complex number. We define a **minimal polynomial** of α over $\mathbb{Q}$ to be a polynomial such that

- $P(x)$ has rational coefficients, and leading coefficient 1,
- $P(\alpha) = 0$.
- The degree of P is as small as possible. We call $\deg P$ the **degree** of α.

Example 53.1.1 (Examples of minimal polynomials)

(a) $\sqrt{2}$ has minimal polynomial $x^2 - 2$.

(b) The imaginary unit $i = \sqrt{-1}$ has minimal polynomial $x^2 + 1$.

[1] URL: https://blog.evanchen.cc/2014/10/19/why-do-roots-come-in-conjugate-pairs/

(c) A primitive pth root of unity, $\zeta_p = e^{\frac{2\pi i}{p}}$, has minimal polynomial $x^{p-1} + x^{p-2} + \dots + 1$, where p is a prime.

Note that $100x^2 - 200$ is also a polynomial of the same degree which has $\sqrt{2}$ as a root; that's why we want to require the polynomial to be monic. That's also why we choose to work in the rational numbers; that way, we can divide by leading coefficients without worrying if we get non-integers.

Why do we care? The point is as follows: suppose we have another polynomial $A(x)$ such that $A(\alpha) = 0$. Then we claim that $P(x)$ actually divides $A(x)$! That means that all the other roots of P will also be roots of A.

The proof is by contradiction: if not, by polynomial long division we can find a quotient and remainder $Q(x)$, $R(x)$ such that

$$A(x) = Q(x)P(x) + R(x)$$

and $R(x) \not\equiv 0$. Notice that by plugging in $x = \alpha$, we find that $R(\alpha) = 0$. But $\deg R < \deg P$, and $P(x)$ was supposed to be the minimal polynomial. That's impossible!

It follows from this and the monotonicity of the minimal polynomial that it is unique (when it exists), so actually it is better to refer to *the* minimal polynomial.

Exercise 53.1.2. Can you find an element in $\mathbb{C}$ that has no minimal polynomial?

Let's look at a more concrete example. Consider $A(x) = x^3 - 3x^2 - 2x + 2$ from the beginning. The minimal polynomial of $2 + \sqrt{2}$ is $P(x) = x^2 - 4x + 2$ (why?). Now we know that if $2 + \sqrt{2}$ is a root, then $A(x)$ is divisible by $P(x)$. And that's how we know that if $2 + \sqrt{2}$ is a root of A, then $2 - \sqrt{2}$ must be a root too.

As another example, the minimal polynomial of $\sqrt[3]{2}$ is $x^3 - 2$. So $\sqrt[3]{2}$ actually has **two** conjugates, namely, $\alpha = \sqrt[3]{2}\,(\cos 120^\circ + i \sin 120^\circ)$ and $\beta = \sqrt[3]{2}\,(\cos 240^\circ + i \sin 240^\circ)$. Thus any polynomial which vanishes at $\sqrt[3]{2}$ also has α and β as roots!

Question 53.1.3 (Important but tautological: irreducible $\iff$ minimal)**.** Let α be a root of the polynomial $P(x)$. Show that $P(x)$ is the minimal polynomial if and only if it is irreducible.

§53.2 Algebraic numbers and algebraic integers

Prototypical example for this section: $\sqrt{2}$ is an algebraic integer (root of $x^2 - 2$), $\frac{1}{2}$ is an algebraic number but not an algebraic integer (root of $x - \frac{1}{2}$).

Let's now work in much vaster generality. First, let's give names to the new numbers we've discussed above.

Definition 53.2.1. An **algebraic number** is any $\alpha \in \mathbb{C}$ which is the root of *some* polynomial with coefficients in $\mathbb{Q}$. The set of algebraic numbers is denoted $\overline{\mathbb{Q}}$.

Remark 53.2.2 — One can equally well say algebraic numbers are those that are roots of some polynomial with coefficients in $\mathbb{Z}$ (rather than $\mathbb{Q}$), since any polynomial in $\mathbb{Q}[x]$ can be scaled to one in $\mathbb{Z}[x]$.

Definition 53.2.3. Consider an algebraic number α and its minimal polynomial P (which is monic and has rational coefficients). If it turns out the coefficients of P are integers, then we say α is an **algebraic integer**.

The set of algebraic integers is denoted $\overline{\mathbb{Z}}$.

Remark 53.2.4 — One can show, using *Gauss's Lemma*, that if α is the root of *any* monic polynomial with integer coefficients, then α is an algebraic integer. So in practice, if I want to prove that $\sqrt{2}+\sqrt{3}$ is an algebraic integer, then I only have to say "the polynomial $(x^2-5)^2-24$ works" without checking that it's minimal.

Sometimes for clarity, we refer to elements of $\mathbb{Z}$ as **rational integers**.

Example 53.2.5 (Examples of algebraic integers)

The numbers

$$4,\ i=\sqrt{-1},\ \sqrt[3]{2},\ \sqrt{2}+\sqrt{3}$$

are all algebraic integers, since they are the roots of the monic polynomials $x-4$, x^2+1, x^3-2 and $(x^2-5)^2-24$.

The number $\frac{1}{2}$ has minimal polynomial $x-\frac{1}{2}$, so it's an algebraic number but not an algebraic integer. (In fact, the rational root theorem also directly implies that any monic integer polynomial does not have $\frac{1}{2}$ as a root!)

There are two properties I want to give for these off the bat, because they'll be used extensively in the tricky (but nice) problems at the end of the section. The first we prove now, since it's very easy:

Proposition 53.2.6 (Rational algebraic integers are rational integers)

An algebraic integer is rational if and only if it is a rational integer. In symbols,

$$\overline{\mathbb{Z}}\cap\mathbb{Q}=\mathbb{Z}.$$

Proof. Let α be a rational number. If α is an integer, it is the root of $x-\alpha$, hence an algebraic integer too.

Conversely, if P is a monic polynomial with integer coefficients such that $P(\alpha)=0$ then (by the rational root theorem, say) it follows α must be an integer. □

The other is that:

Proposition 53.2.7 ($\overline{\mathbb{Z}}$ is a ring and $\overline{\mathbb{Q}}$ is a field)

The algebraic integers $\overline{\mathbb{Z}}$ form a ring. The algebraic numbers $\overline{\mathbb{Q}}$ form a field.

We could prove this now if we wanted to, but the results in the next chapter will more or less do it for us, and so we take this on faith temporarily.

§53.3 Number fields

Prototypical example for this section: $\mathbb{Q}(\sqrt{2})$ *is a typical number field.*

Given any algebraic number α, we're able to consider fields of the form $\mathbb{Q}(\alpha)$. Let us write down the more full version.

Definition 53.3.1. A **number field** K is a field containing $\mathbb{Q}$ as a subfield which is a *finite-dimensional* $\mathbb{Q}$-vector space. The **degree** of K is its dimension.

Example 53.3.2 (Prototypical example)
Consider the field
$$K = \mathbb{Q}(\sqrt{2}) = \left\{ a + b\sqrt{2} \mid a, b \in \mathbb{Q} \right\}.$$
This is a field extension of $\mathbb{Q}$, and has degree 2 (the basis being 1 and $\sqrt{2}$).

You might be confused that I wrote $\mathbb{Q}(\sqrt{2})$ (which should permit denominators) instead of $\mathbb{Q}[\sqrt{2}]$, say. But if you read through Example 5.5.4, you should see that the denominators don't really matter: $\frac{1}{3-\sqrt{2}} = \frac{1}{7}(3+\sqrt{2})$ anyways, for example. You can either check this now in general, or just ignore the distinction and pretend I wrote square brackets everywhere.

Exercise 53.3.3 (Unimportant)**.** Show that if α is an algebraic number, then $\mathbb{Q}(\alpha) \cong \mathbb{Q}[\alpha]$.

Example 53.3.4 (Adjoining an algebraic number)
Let α be the root of some irreducible polynomial $P(x) \in \mathbb{Q}[x]$. The field $\mathbb{Q}(\alpha)$ is a field extension as well, and the basis is $1, \alpha, \alpha^2, \dots, \alpha^{m-1}$, where $m = \deg P$. In particular, the degree of $\mathbb{Q}(\alpha)$ is the degree of P.

Example 53.3.5 (Non-examples of number fields)
$\mathbb{R}$ and $\mathbb{C}$ are not number fields since there is no *finite* $\mathbb{Q}$-basis of them.

§53.4 Primitive element theorem, and monogenic extensions

Prototypical example for this section: $\mathbb{Q}(\sqrt{3}, \sqrt{5}) \cong \mathbb{Q}(\sqrt{3}+\sqrt{5})$. *Can you see why?*

I'm only putting this theorem here because I was upset that no one told me it was true (it's a very natural conjecture), and I hope to not do the same to the reader. However, I'm not going to use it in anything that follows.

Theorem 53.4.1 (Artin's primitive element theorem)
Every number field K is isomorphic to $\mathbb{Q}(\alpha)$ for some algebraic number α.

The proof is left as Problem 59F, since to prove it I need to talk about field extensions first.

The prototypical example
$$\mathbb{Q}(\sqrt{3}, \sqrt{5}) \cong \mathbb{Q}(\sqrt{3}+\sqrt{5})$$
makes it clear why this theorem should not be too surprising.

§53.5 A few harder problems to think about

Problem 53A. Find a polynomial with integer coefficients which has $\sqrt{2} + \sqrt[3]{3}$ as a root.

🌶 **Problem 53B** (Brazil 2006)**.** Let p be an irreducible polynomial in $\mathbb{Q}[x]$ and degree larger than 1. Prove that if p has two roots r and s whose product is 1 then the degree of p is even.

Problem 53C*. Consider n roots of unity $\varepsilon_1, \dots, \varepsilon_n$. Assume the average $\frac{1}{n}(\varepsilon_1 + \dots + \varepsilon_n)$ is an algebraic integer. Prove that either the average is zero or $\varepsilon_1 = \dots = \varepsilon_n$. (Used in Lemma 22.2.2.)

🌶 **Problem 53D†.** Which rational numbers q satisfy $\cos(q\pi) \in \mathbb{Q}$?

Problem 53E (MOP 2010)**.** There are $n > 2$ lamps arranged in a circle; initially one is on and the others are off. We may select any regular polygon whose vertices are among the lamps and toggle the states of all the lamps simultaneously. Show it is impossible to turn all lamps off.

🌶🌶 **Problem 53F** (Kronecker's theorem)**.** Let α be an algebraic integer. Suppose all its Galois conjugates have absolute value one. Prove that $\alpha^N = 1$ for some positive integer N.

🌶🌶 **Problem 53G.** Is there an algebraic integer with absolute value one which is not a root of unity?

Problem 53H. Is the ring of algebraic integers Noetherian?

54 The ring of integers

§54.1 Norms and traces

Prototypical example for this section: $a + b\sqrt{2}$ *as an element of* $\mathbb{Q}(\sqrt{2})$ *has norm* $a^2 - 2b^2$ *and trace* $2a$.

Remember when you did olympiads and we had like $a^2 + b^2$ was the "norm" of $a + bi$? Cool, let me tell you what's actually happening.

First, let me make precise the notion of a conjugate.

Definition 54.1.1. Let α be an algebraic number, and let $P(x)$ be its minimal polynomial, of degree m. Then the m roots of P are the (Galois) **conjugates** of α.

It's worth showing at the moment that there are no repeated conjugates.

Lemma 54.1.2 (Irreducible polynomials have distinct roots)

An irreducible polynomial in $\mathbb{Q}[x]$ cannot have a complex double root.

Proof. Let $f(x) \in \mathbb{Q}[x]$ be the irreducible polynomial and assume it has a double root α. **Take the derivative $f'(x)$.** This derivative has three interesting properties.

- The degree of f' is one less than the degree of f.
- The polynomials f and f' are not relatively prime because they share a factor $x - \alpha$.
- The coefficients of f' are also in $\mathbb{Q}$.

Consider $g = \gcd(f, f')$. We must have $g \in \mathbb{Q}[x]$ by Euclidean algorithm. But the first two facts about f' ensure that g is nonconstant and $\deg g < \deg f$. Yet g divides f, contradiction to the fact that f should be a minimal polynomial. □

Hence α has exactly as many conjugates as the degree of α.

Now, we would *like* to define the *norm* of an element $\mathrm{N}(\alpha)$ as the product of its conjugates. For example, we want $2 + i$ to have norm $(2 + i)(2 - i) = 5$, and in general for $a + bi$ to have norm $a^2 + b^2$. It would be *really cool* if the norm was multiplicative; we already know this is true for complex numbers!

Unfortunately, this doesn't quite work: consider

$$\mathrm{N}(2 + i) = 5 \text{ and } \mathrm{N}(2 - i) = 5.$$

But $(2 + i)(2 - i) = 5$, which doesn't have norm 25 like we want, since 5 is degree 1 and has no conjugates at all. The reason this "bad" thing is happening is that we're trying to define the norm of an *element*, when we really ought to be defining the norm of an element *with respect to a particular* K.

What I'm driving at is that the norm should have different meanings depending on which field you're in. If we think of 5 as an element of $\mathbb{Q}$, then its norm is 5. But thought of as an element of $\mathbb{Q}(i)$, its norm really ought to be 25. Let's make this happen: for K a number field, we will now define $\mathrm{N}_{K/\mathbb{Q}}(\alpha)$ to be the norm of α *with respect to* K as follows.

Definition 54.1.3. Let $\alpha \in K$ have degree n, so $\mathbb{Q}(\alpha) \subseteq K$, and set $k = (\deg K)/n$. The **norm** of α is defined as

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha) := \left(\prod \text{Galois conj of } \alpha\right)^k .$$

The **trace** is defined as

$$\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) := k \cdot \left(\sum \text{Galois conj of } \alpha\right) .$$

The exponent of k is a "correction factor" that makes the norm of 5 into $5^2 = 25$ when we view 5 as an element of $\mathbb{Q}(i)$ rather than an element of $\mathbb{Q}$. For a "generic" element of K, we expect $k = 1$.

Exercise 54.1.4. Use what you know about nested vector spaces to convince yourself that k is actually an integer.

Example 54.1.5 (Norm of $a + b\sqrt{2}$)

Let $\alpha = a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2}) = K$. If $b \neq 0$, then α and K have the degree 2. Thus the only conjugates of α are $a \pm b\sqrt{2}$, which gives the norm

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

The trace is $(a - b\sqrt{2}) + (a + b\sqrt{2}) = 2a$.

Nicely, the formula $a^2 - 2b^2$ and $2a$ also works when $b = 0$.

Example 54.1.6 (Norm of $a + b\sqrt[3]{2} + c\sqrt[3]{4}$)

Let $\alpha = a + b\sqrt[3]{2} + c\sqrt[3]{4} \in \mathbb{Q}(\sqrt[3]{2}) = K$. As above, if $b \neq 0$ or $c \neq 0$, then α and K have the same degree 3. The conjugates of α are $a + b\sqrt[3]{2}\omega + c\sqrt[3]{4}\omega^2$ and $a + b\sqrt[3]{2}\omega^2 + c\sqrt[3]{4}\omega$, and we can compute $\mathrm{N}_{K/\mathbb{Q}}(\alpha) = a^3 + 2b^3 + 4c^3 - 6abc$ and $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) = 3a$.

Note that in this case the conjugates of α does not lie in the field K!

Of importance is:

Proposition 54.1.7 (Norms and traces are rational integers)

If α is an algebraic integer, its norm and trace are rational integers.

Question 54.1.8. Prove it. (Vieta formula.)

That's great, but it leaves a question unanswered: why is the norm multiplicative? To do this, I have to give a new definition of norm and trace.

Remark 54.1.9 — Another way to automatically add the "corrective factor" is to use the embeddings of K into $\mathbb{C}$.

As we will see later, in Theorem 59.3.1, there are exactly $d = \deg K$ embeddings of K into $\mathbb{C}$, say $\sigma_1, \dots, \sigma_d$. Then, $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) = \sum_{i=1}^d \sigma_i(\alpha)$ and $\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \prod_{i=1}^d \sigma_i(\alpha)$.

Theorem 54.1.10 (Morally correct definition of norm and trace)
Let K be a number field of degree n, and let $\alpha \in K$. Let $\mu_\alpha \colon K \to K$ denote the map

$$x \mapsto \alpha x$$

viewed as a linear map of $\mathbb{Q}$-vector spaces. Then,

- the norm of α equals the determinant $\det \mu_\alpha$, and
- the trace of α equals the trace $\operatorname{Tr} \mu_\alpha$.

The definition of the determinant has an obvious geometrical interpretation: viewing $K \cong \mathbb{Q}^n$ as a vector space, the determinant measures how much $\mathbb{Q}^n$ is stretched when multiplied by α. That is, given a parallelepiped with volume v in $\mathbb{Q}^n$, it will be transformed to one with volume $|\operatorname{N}(\alpha)|v$ under the transformation μ_α.

Since the trace and determinant don't depend on the choice of basis, you can pick whatever basis you want and use whatever definition you got in high school. Fantastic, right?

Example 54.1.11 (Explicit computation of matrices for $a + b\sqrt{2}$)
Let $K = \mathbb{Q}(\sqrt{2})$, and let 1, $\sqrt{2}$ be the basis of K. Let

$$\alpha = a + b\sqrt{2}$$

(possibly even $b = 0$), and notice that

$$\left(a + b\sqrt{2}\right)\left(x + y\sqrt{2}\right) = (ax + 2yb) + (bx + ay)\sqrt{2}.$$

We can rewrite this in matrix form as

$$\begin{bmatrix} a & 2b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + 2yb \\ bx + ay \end{bmatrix}.$$

Consequently, we can interpret μ_α as the matrix

$$\mu_\alpha = \begin{bmatrix} a & 2b \\ b & a \end{bmatrix}.$$

Of course, the matrix will change if we pick a different basis, but the determinant and trace do not: they are always given by

$$\det \mu_\alpha = a^2 - 2b^2 \text{ and } \operatorname{Tr} \mu_\alpha = 2a.$$

This interpretation explains why the same formula should work for $a + b\sqrt{2}$ even in the case $b = 0$.

Proof. I'll prove the result for just the norm; the trace falls out similarly. Set

$$n = \deg \alpha, \qquad kn = \deg K.$$

The proof is split into two parts, depending on whether or not $k = 1$.

Proof if $k = 1$. Set $n = \deg \alpha = \deg K$. Thus the norm actually *is* the product of the Galois conjugates. Also,

$$\{1, \alpha, \dots, \alpha^{n-1}\}$$

is linearly independent in K, and hence a basis (as $\dim K = n$). Let's use this as the basis for μ_α.

Let

$$x^n + c_{n-1}x^{n-1} + \dots + c_0$$

be the minimal polynomial of α. Thus $\mu_\alpha(1) = \alpha$, $\mu_\alpha(\alpha) = \alpha^2$, and so on, but $\mu_\alpha(\alpha^{n-1}) = -c_{n-1}\alpha^{n-1} - \dots - c_0$. Therefore, μ_α is given by the matrix

$$M = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & 0 & \dots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & 0 & -c_{n-2} \\ 0 & 0 & 0 & \dots & 1 & -c_{n-1} \end{bmatrix}.$$

Thus

$$\det M = (-1)^n c_0$$

and we're done by Vieta's formulas. ■

Proof if $k > 1$. We have nested vector spaces

$$\mathbb{Q} \subseteq \mathbb{Q}(\alpha) \subseteq K.$$

Let e_1, ..., e_k be a $\mathbb{Q}(\alpha)$-basis for K (meaning: interpret K as a vector space over $\mathbb{Q}(\alpha)$, and pick that basis). Since $\{1, \alpha, \dots, \alpha^{n-1}\}$ is a $\mathbb{Q}$ basis for $\mathbb{Q}(\alpha)$, the elements

$$\begin{matrix} e_1, & e_1\alpha, & \dots, & e_1\alpha^{n-1} \\ e_2, & e_2\alpha, & \dots, & e_2\alpha^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ e_k, & e_k\alpha, & \dots, & e_k\alpha^{n-1} \end{matrix}$$

constitute a $\mathbb{Q}$-basis of K. Using *this* basis, the map μ_α looks like

$$\underbrace{\begin{bmatrix} M & & & \\ & M & & \\ & & \ddots & \\ & & & M \end{bmatrix}}_{k \text{ times}}$$

where M is the same matrix as above: we just end up with one copy of our old matrix for each e_i. Thus $\det \mu_\alpha = (\det M)^k$, as needed. ■

> **Question 54.1.12.** Verify the result for traces as well. □

From this it follows immediately that

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha\beta) = \mathrm{N}_{K/\mathbb{Q}}(\alpha)\,\mathrm{N}_{K/\mathbb{Q}}(\beta)$$

because by definition we have

$$\mu_{\alpha\beta} = \mu_\alpha \circ \mu_\beta,$$

and that the determinant is multiplicative. In the same way, the trace is additive.

§54.2 The ring of integers

Prototypical example for this section: If $K = \mathbb{Q}(\sqrt{2})$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{2}]$. But if $K = \mathbb{Q}(\sqrt{5})$, then $\mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$.

$\mathbb{Z}$ makes for better number theory than $\mathbb{Q}$. In the same way, focusing on the *algebraic integers* of K gives us some really nice structure, and we'll do that here.

Definition 54.2.1. Given a number field K, we define

$$\mathcal{O}_K := K \cap \overline{\mathbb{Z}}$$

to be the **ring of integers** of K; in other words $\mathcal{O}_K$ consists of the algebraic integers of K.

We do the classical example of a quadratic field now. Before proceeding, I need to write a silly number theory fact.

Exercise 54.2.2 (Annoying but straightforward)**.** Let a and b be rational numbers, and d a squarefree positive integer.

- If $d \equiv 2, 3 \pmod 4$, prove that $2a, a^2 - db^2 \in \mathbb{Z}$ if and only if $a, b \in \mathbb{Z}$.
- For $d \equiv 1 \pmod 4$, prove that $2a, a^2 - db^2 \in \mathbb{Z}$ if and only if $a, b \in \mathbb{Z}$ OR if $a - \frac{1}{2}, b - \frac{1}{2} \in \mathbb{Z}$.

You'll need to take mod 4.

Example 54.2.3 (Ring of integers of $K = \mathbb{Q}(\sqrt{3})$)

Let K be as above. We claim that

$$\mathcal{O}_K = \mathbb{Z}[\sqrt{3}] = \left\{ m + n\sqrt{3} \mid m, n \in \mathbb{Z} \right\}.$$

We set $\alpha = a + b\sqrt{3}$. Then $\alpha \in \mathcal{O}_K$ when the minimal polynomial has integer coefficients.

If $b = 0$, then the minimal polynomial is $x - \alpha = x - a$, and thus α works if and only if it's an integer. If $b \neq 0$, then the minimal polynomial is

$$(x-a)^2 - 3b^2 = x^2 - 2a \cdot x + (a^2 - 3b^2).$$

From the exercise, this occurs exactly for $a, b \in \mathbb{Z}$.

Example 54.2.4 (Ring of integers of $K = \mathbb{Q}(\sqrt{5})$)

We claim that in this case

$$\mathcal{O}_K = \mathbb{Z}\left[\frac{1+\sqrt{5}}{2}\right] = \left\{ m + n \cdot \frac{1+\sqrt{5}}{2} \mid m, n \in \mathbb{Z} \right\}.$$

The proof is exactly the same, except the exercise tells us instead that for $b \neq 0$, we have both the possibility that $a, b \in \mathbb{Z}$ or that $a, b \in \mathbb{Z} - \frac{1}{2}$. This reflects the fact that $\frac{1+\sqrt{5}}{2}$ is the root of $x^2 - x - 1 = 0$; no such thing is possible with $\sqrt{3}$.

In general, the ring of integers of $K = \mathbb{Q}(\sqrt{d})$ is

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{d}] & d \equiv 2, 3 \pmod 4 \\ \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] & d \equiv 1 \pmod 4. \end{cases}$$

What we're going to show is that $\mathcal{O}_K$ behaves in K a lot like the integers do in $\mathbb{Q}$. First we show K consists of quotients of numbers in $\mathcal{O}_K$. In fact, we can do better:

Example 54.2.5 (Rationalizing the denominator)

For example, consider $K = \mathbb{Q}(\sqrt{3})$. The number $x = \frac{1}{4+\sqrt{3}}$ is an element of K, but by "rationalizing the denominator" we can write

$$\frac{1}{4+\sqrt{3}} = \frac{4-\sqrt{3}}{13}.$$

So we see that in fact, x is $\frac{1}{13}$ of an integer in $\mathcal{O}_K$.

The theorem holds true more generally.

Theorem 54.2.6 ($K = \mathbb{Q} \cdot \mathcal{O}_K$)

Let K be a number field, and let $x \in K$ be any element. Then there exists an integer n such that $nx \in \mathcal{O}_K$; in other words,

$$x = \frac{1}{n}\alpha$$

for some $\alpha \in \mathcal{O}_K$.

> **Exercise 54.2.7.** Prove this yourself. (Start by using the fact that x has a minimal polynomial with rational coefficients. Alternatively, take the norm.)

Now we are going to show $\mathcal{O}_K$ is a ring; we'll check it is closed under addition and multiplication. To do so, the easiest route is:

Lemma 54.2.8 ($\alpha \in \overline{\mathbb{Z}} \iff \mathbb{Z}[\alpha]$ finitely generated)

Let $\alpha \in \overline{\mathbb{Q}}$. Then α is an algebraic integer if and only if the abelian group $\mathbb{Z}[\alpha]$ is finitely generated.

Proof. Note that α is an algebraic integer if and only if it's the root of some nonzero, monic polynomial with integer coefficients. Suppose first that

$$\alpha^N = c_{N-1}\alpha^{N-1} + c_{N-2}\alpha^{N-2} + \cdots + c_0.$$

Then the set $1, \alpha, \dots, \alpha^{N-1}$ generates $\mathbb{Z}[\alpha]$, since we can repeatedly replace α^N until all powers of α are less than N.

Conversely, suppose that $\mathbb{Z}[\alpha]$ is finitely generated by some $b_1, \dots, b_m$. Viewing the b_i as polynomials in α, we can select a large integer N (say $N = \deg b_1 + \cdots + \deg b_m + 2015$) and express α^N in the b_i's to get

$$\alpha^N = c_1 b_1(\alpha) + \cdots + c_m b_m(\alpha).$$

The above gives us a monic polynomial in α, and the choice of N guarantees it is not zero. So α is an algebraic integer. $\square$

Example 54.2.9 ($\frac{1}{2}$ isn't an algebraic integer)
We already know $\frac{1}{2}$ isn't an algebraic integer. So we expect

$$\mathbb{Z}\left[\frac{1}{2}\right] = \left\{\frac{a}{2^m} \mid a, m \in \mathbb{Z} \text{ and } m \geq 0\right\}$$

to not be finitely generated, and this is the case.

Question 54.2.10. To make the last example concrete: name all the elements of $\mathbb{Z}[\frac{1}{2}]$ that cannot be written as an integer combination of

$$\left\{\frac{1}{2}, \frac{7}{8}, \frac{13}{64}, \frac{2015}{4096}, \frac{1}{1048576}\right\}$$

Now we can state the theorem.

Theorem 54.2.11 (Algebraic integers are closed under $+$ and $\times$)
The set $\overline{\mathbb{Z}}$ is closed under addition and multiplication; i.e. it is a ring. In particular, $\mathcal{O}_K$ is also a ring for any number field K.

Proof. Let $\alpha, \beta \in \overline{\mathbb{Z}}$. Then $\mathbb{Z}[\alpha]$ and $\mathbb{Z}[\beta]$ are finitely generated. Hence so is $\mathbb{Z}[\alpha, \beta]$. (Details: if $\mathbb{Z}[\alpha]$ has $\mathbb{Z}$-basis $a_1, \dots, a_m$ and $\mathbb{Z}[\beta]$ has $\mathbb{Z}$-basis $b_1, \dots, b_n$, then take the mn elements $a_i b_j$.)

Now $\mathbb{Z}[\alpha \pm \beta]$ and $\mathbb{Z}[\alpha\beta]$ are subsets of $\mathbb{Z}[\alpha, \beta]$ and so they are also finitely generated. Hence $\alpha \pm \beta$ and $\alpha\beta$ are algebraic integers. $\square$

In fact, something even better is true. As you saw, for $\mathbb{Q}(\sqrt{3})$ we had $\mathcal{O}_K = \mathbb{Z}[\sqrt{3}]$; in other words, $\mathcal{O}_K$ was generated by 1 and $\sqrt{3}$. Something similar was true for $\mathbb{Q}(\sqrt{5})$. We claim that in fact, the general picture looks exactly like this.

Theorem 54.2.12 ($\mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank n)
Let K be a number field of degree n. Then $\mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank n, i.e. $\mathcal{O}_K \cong \mathbb{Z}^{\oplus n}$ as an abelian group. In other words, $\mathcal{O}_K$ has a $\mathbb{Z}$-basis of n elements as

$$\mathcal{O}_K = \{c_1\alpha_1 + \dots + c_{n-1}\alpha_{n-1} + c_n\alpha_n \mid c_i \in \mathbb{Z}\}$$

where α_i are algebraic integers in $\mathcal{O}_K$.

The proof will be postponed to a later chapter.

This last theorem shows that in many ways $\mathcal{O}_K$ is a "lattice" in K. That is, for a number field K we can find $\alpha_1, \dots, \alpha_n$ in $\mathcal{O}_K$ such that

$$\begin{aligned} \mathcal{O}_K &\cong \alpha_1\mathbb{Z} \oplus \alpha_2\mathbb{Z} \oplus \dots \oplus \alpha_n\mathbb{Z} \\ K &\cong \alpha_1\mathbb{Q} \oplus \alpha_2\mathbb{Q} \oplus \dots \oplus \alpha_n\mathbb{Q} \end{aligned}$$

as abelian groups.

§54.3 On monogenic extensions

Recall that it turned out number fields K could all be expressed as $\mathbb{Q}(\alpha)$ for some α. We might hope that something similar is true of the ring of integers: that we can write

$$\mathcal{O}_K = \mathbb{Z}[\theta]$$

in which case $\{1, \theta, \dots, \theta^{n-1}\}$ serves both as a basis of K and as the $\mathbb{Z}$-basis for $\mathcal{O}_K$ (here $n = [K : \mathbb{Q}]$). In other words, we hope that the basis of $\mathcal{O}_K$ is actually a "power basis".

This is true for the most common examples we use:

- the quadratic field, and
- the cyclotomic field in Problem 54E†.

Unfortunately, it is not true in general: the first counterexample is $K = \mathbb{Q}(\alpha)$ for α a root of $X^3 - X^2 - 2X - 8$.

We call an extension with this nice property **monogenic**. As we'll later see, monogenic extensions have a really nice factoring algorithm, Theorem 55.5.4.

Remark 54.3.1 (What went wrong with $\mathcal{O}_K$?) — As we have just mentioned above, as an abelian group, $\mathcal{O}_K \cong \mathbb{Z}^3$, so it's generated by finitely many elements.

In fact, $\{1, \alpha, \beta\}$ is a basis of $\mathcal{O}_K$, where $\beta = \frac{\alpha+\alpha^2}{2}$. The group generated by $\{1, \alpha, \alpha^2\}$ has index 2 in $\mathcal{O}_K$ – that is, $|\mathcal{O}_K / \langle 1, \alpha, \alpha^2 \rangle| = 2$, and we misses β.

If we try to pick $\{1, \beta, \beta^2\}$ as a basis instead, again we get $|\mathcal{O}_K / \langle 1, \beta, \beta^2 \rangle| = 2$, and we misses α. If you explicitly compute it out, you can get $\beta^2 = \frac{3\alpha^2+7\alpha}{2} + 6 = 3\beta + 2\alpha + 6$.

While this is not a proof that the extension is not monogenic, hopefully it gives you a feeling of the structure of $\mathcal{O}_K$.

§54.4 A few harder problems to think about

Problem 54A⋆. Show that α is a unit of $\mathcal{O}_K$ (meaning $\alpha^{-1} \in \mathcal{O}_K$) if and only if $\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \pm 1$.

Problem 54B⋆. Let K be a number field. What is the field of fractions of $\mathcal{O}_K$?

Problem 54C (Russian olympiad 1984)**.** Find all integers m and n such that

$$\left(5 + 3\sqrt{2}\right)^m = \left(3 + 5\sqrt{2}\right)^n.$$

Problem 54D (USA TST 2012)**.** Decide whether there exist $a, b, c > 2010$ satisfying

$$a^3 + 2b^3 + 4c^3 = 6abc + 1.$$

Problem 54E† (Cyclotomic Field)**.** Let p be an odd rational prime and ζ_p a primitive pth root of unity. Let $K = \mathbb{Q}(\zeta_p)$. Prove that $\mathcal{O}_K = \mathbb{Z}[\zeta_p]$. (In fact, the result is true even if p is not a prime.)

55 Unique factorization (finally!)

Took long enough.

§55.1 Motivation

Suppose we're interested in solutions to the Diophantine equation $n = x^2 + 5y^2$ for a given n. The idea is to try and "factor" n in $\mathbb{Z}[\sqrt{-5}]$, for example

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

Unfortunately, this is not so simple, because as I've said before we don't have unique factorization of elements:

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

One reason this doesn't work is that we don't have a notion of a *greatest common divisor*. We can write $(35, 77) = 7$, but what do we make of $(3, 1 + \sqrt{-5})$?

The trick is to use ideals as a "generalized GCD". Recall that by (a, b) I mean the ideal $\{ax + by \mid x, y \in \mathbb{Z}[\sqrt{-5}]\}$. You can see that $(35, 77) = (7)$, but $(3, 1 + \sqrt{-5})$ will be left "unsimplified" because it doesn't represent an actual value in the ring. Using these *sets* (ideals) as elements, it turns out that we can develop a full theory of prime factorization, and we do so in this chapter.

In other words, we use the ideal $(a_1, \dots, a_m)$ to interpret a "generalized GCD" of a_1, $\dots$, a_m. In particular, if we have a number x we want to represent, we encode it as just (x).

Going back to our example of 6,

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

Please take my word for it that in fact, the complete prime factorization of (6) into prime ideals is

$$(6) = (2, 1 - \sqrt{-5})^2(3, 1 + \sqrt{-5})(3, 1 - \sqrt{-5}) = \mathfrak{p}^2\mathfrak{q}_1\mathfrak{q}_2.$$

In fact, $(2) = \mathfrak{p}^2$, $(3) = \mathfrak{q}_1\mathfrak{q}_2$, $(1 + \sqrt{-5}) = \mathfrak{p}\mathfrak{q}_1$, $(1 - \sqrt{-5}) = \mathfrak{p}\mathfrak{q}_2$. So 6 indeed factorizes uniquely into ideals, even though it doesn't factor into elements.

As one can see above, ideal factorization is more refined than element factorization. Once you have the factorization into *ideals*, you can from there recover all the factorizations into *elements*. The upshot of this is that if we want to write n as $x^2 + 5y^2$, we just have to factor n into ideals, and from there we can recover all factorizations into elements, and finally all ways to write n as $x^2 + 5y^2$. Since we can already break n into rational prime factors (for example $6 = 2 \cdot 3$ above) we just have to figure out how each rational prime $p \mid n$ breaks down. There's a recipe for this, Theorem 55.5.4! In fact, I'll even tell you what is says in this special case:

- If $t^2 + 5$ factors as $(t + c)(t - c) \pmod p$, then $(p) = (p, c + \sqrt{-5})(p, c - \sqrt{-5})$.
- Otherwise, (p) is a prime ideal.

In this chapter we'll develop this theory of unique factorization in full generality.

Remark 55.1.1 — In this chapter, I'll be using the letters $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{p}$, $\mathfrak{q}$ for ideals of $\mathcal{O}_K$. When fractional ideals arise, I'll use I and J for them.

§55.2 Ideal arithmetic

Prototypical example for this section: $(x)(y) = (xy)$, *and* $(x) + (y) = (\gcd(x, y))$. *In any case, think in terms of generators.*

First, I have to tell you how to add and multiply two ideals $\mathfrak{a}$ and $\mathfrak{b}$.

Definition 55.2.1. Given two ideals $\mathfrak{a}$ and $\mathfrak{b}$ of a ring R, we define

$$\begin{aligned} \mathfrak{a} + \mathfrak{b} &:= \{a + b \mid a \in \mathfrak{a}, b \in \mathfrak{b}\} \\ \mathfrak{a} \cdot \mathfrak{b} &:= \{a_1 b_1 + \dots + a_n b_n \mid a_i \in \mathfrak{a}, b_i \in \mathfrak{b}\}. \end{aligned}$$

(Note that infinite sums don't make sense in general rings, which is why in $\mathfrak{a} \cdot \mathfrak{b}$ we cut off the sum after some finite number of terms.) You can readily check these are actually ideals. This definition is more natural if you think about it in terms of the generators of $\mathfrak{a}$ and $\mathfrak{b}$.

Proposition 55.2.2 (Ideal arithmetic via generators)

Suppose $\mathfrak{a} = (a_1, a_2, \dots, a_n)$ and $\mathfrak{b} = (b_1, \dots, b_m)$ are ideals in a ring R. Then

(a) $\mathfrak{a} + \mathfrak{b}$ is the ideal generated by $a_1, \dots, a_n, b_1, \dots, b_m$.

(b) $\mathfrak{a} \cdot \mathfrak{b}$ is the ideal generated by $a_i b_j$, for $1 \le i \le n$ and $1 \le j \le m$.

Proof. Pretty straightforward; just convince yourself that this result is correct. □

In other words, for sums you append the two sets of generators together, and for products you take products of the generators. Note that for principal ideals, this coincides with "normal" multiplication, for example

$$(3) \cdot (5) = (15)$$

in $\mathbb{Z}$.

Remark 55.2.3 — Note that for an ideal $\mathfrak{a}$ and an element c, the set

$$c\mathfrak{a} = \{ca \mid a \in \mathfrak{a}\}$$

is equal to $(c) \cdot \mathfrak{a}$. So "scaling" and "multiplying by principal ideals" are the same thing. This is important, since we'll be using the two notions interchangeably.

Remark 55.2.4 — The addition of two ideals does not correspond to the addition of elements — for example, $(4) + (6) = (4, 6) = (2)$, but $4 + 6 = 10$.

This is the best we can hope for — addition of elements does not make sense for ideals — for example, $1 + 1 = 2$ and $1 + (-1) = 0$, but as ideals, $(1) = (-1)$.

In fact, addition of ideal is the straightforward generalization of gcd of elements — as you can check in the example above, $\gcd(4, 6) = 2$.

Nevertheless, I hope you agree that $\mathfrak{a} + \mathfrak{b}$ is a natural notation, compared to something like $(\mathfrak{a}, \mathfrak{b})$.

> Because factorization involves *multiplying*, instead of adding, the ideals together, we will not need to use the notation $\mathfrak{a}+\mathfrak{b}$ any time soon.

Finally, since we want to do factorization we better have some notion of divisibility. So we define:

Definition 55.2.5. We say $\mathfrak{a}$ divides $\mathfrak{b}$ and write $\mathfrak{a} \mid \mathfrak{b}$ if $\mathfrak{a} \supseteq \mathfrak{b}$.

Note the reversal of inclusions! So (3) divides (15), because (15) is contained in (3); every multiple of 15 is a multiple of 3. And from the example in the previous section: In $\mathbb{Z}[\sqrt{-5}]$, $(3, 1-\sqrt{-5})$ divides (3) and $(1-\sqrt{-5})$.

Finally, the **prime ideals** are defined as in Definition 5.3.1: $\mathfrak{p}$ is prime if $xy \in \mathfrak{p}$ implies $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. This is compatible with the definition of divisibility:

> **Exercise 55.2.6.** A nonzero proper ideal $\mathfrak{p}$ is prime if and only if whenever $\mathfrak{p}$ divides $\mathfrak{a}\mathfrak{b}$, $\mathfrak{p}$ divides one of $\mathfrak{a}$ or $\mathfrak{b}$.

As mentioned in Remark 5.3.3, this also lets us ignore multiplication by units: $(-3) = (3)$.

§55.3 Dedekind domains

Prototypical example for this section: Any $\mathcal{O}_K$ is a Dedekind domain.

We now define a Dedekind domain as follows.

Definition 55.3.1. An integral domain $\mathcal{A}$ is a **Dedekind domain** if it is Noetherian, integrally closed, and *every nonzero prime ideal of $\mathcal{A}$ is in fact maximal.* (The last condition is the important one.)

> **Remark 55.3.2** — Note that $\mathcal{A}$ is a Dedekind domain if and only if $\mathcal{A} = \mathcal{O}_K$ for some field K, as we will prove below. We're just defining this term for historical reasons...

Here there's one new word I have to define for you, but we won't make much use of it.

Definition 55.3.3. Let R be an integral domain and let K be its field of fractions. We say R is **integrally closed** if the only elements $a \in K$ which are roots of *monic* polynomials in R are the elements of R (which are roots of the trivial $x - r$ polynomial).

The *interesting* condition in the definition of a Dedekind domain is the last one: prime ideals and maximal ideals are the same thing. The other conditions are just technicalities, but "primes are maximal" has real substance.

Example 55.3.4 ($\mathbb{Z}$ is a Dedekind domain)

The ring $\mathbb{Z}$ is a Dedekind domain. Note that

- $\mathbb{Z}$ is Noetherian (for obvious reasons).
- $\mathbb{Z}$ has field of fractions $\mathbb{Q}$. If $f(x) \in \mathbb{Z}[x]$ is monic, then by the rational root theorem any rational roots are integers (this is the same as the proof that $\overline{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}$). Hence $\mathbb{Z}$ is integrally closed.
- The nonzero prime ideals of $\mathbb{Z}$ are (p), which also happen to be maximal.

The case of interest is a ring $\mathcal{O}_K$ in which we wish to do factorizing. We're now going to show that for any number field K, the ring $\mathcal{O}_K$ is a Dedekind domain. First, the boring part.

Proposition 55.3.5 ($\mathcal{O}_K$ integrally closed and Noetherian)

For any number field K, the ring $\mathcal{O}_K$ is integrally closed and Noetherian.

Proof. Boring, but here it is anyways for completeness.

Since $\mathcal{O}_K \cong \mathbb{Z}^{\oplus n}$,[1] we get that it's Noetherian.

Now we show that $\mathcal{O}_K$ is integrally closed. Suppose that $\eta \in K$ is the root of some polynomial with coefficients in $\mathcal{O}_K$. Thus

$$\eta^n = \alpha_{n-1} \cdot \eta^{n-1} + \alpha_{n-2} \cdot \eta^{n-2} + \cdots + \alpha_0$$

where $\alpha_i \in \mathcal{O}_K$. We want to show that $\eta \in \mathcal{O}_K$ as well.

Well, from the above, $\mathcal{O}_K[\eta]$ is finitely generated... thus $\mathbb{Z}[\eta] \subseteq \mathcal{O}_K[\eta]$ is finitely generated. So $\eta \in \overline{\mathbb{Z}}$, and hence $\eta \in K \cap \overline{\mathbb{Z}} = \mathcal{O}_K$. □

Now let's do the fun part. We'll prove a stronger result, which will re-appear repeatedly.

Theorem 55.3.6 (Important: prime ideals divide rational primes)

Let $\mathcal{O}_K$ be a ring of integers and $\mathfrak{p}$ a nonzero prime ideal inside it. Then $\mathfrak{p}$ contains a rational prime p. Moreover, $\mathfrak{p}$ is maximal.

For a concrete example, consider $(2+i) \subseteq \mathbb{Z}[i]$. In this case, $p = 5$, and because $p \in (2+i)$, we get that the lattice is periodic in both dimensions with period 5, which implies finitely many points in $\{a + bi \mid 0 \le a, b \le 4, a, b \in \mathbb{Z}\}$ suffices to cover all cosets modulo the ideal.

The proof of the finiteness of the quotient here is closely related to the statement that the mesh of the lattice is finite, which will be covered in the next section.

Proof. Take any $\alpha \neq 0$ in $\mathfrak{p}$. Its Galois conjugates are algebraic integers so their product $\operatorname{N}(\alpha)/\alpha$ is in $\mathcal{O}_K$ (even though each individual conjugate need not be in K). Consequently, $\operatorname{N}(\alpha) \in \mathfrak{p}$, and we conclude $\mathfrak{p}$ contains some integer.

Then take the smallest positive integer in $\mathfrak{p}$, say p. We must have that p is a rational prime, since otherwise $\mathfrak{p} \ni p = xy$ implies one of $x, y \in \mathfrak{p}$. This shows the first part.

We now do something pretty tricky to show $\mathfrak{p}$ is maximal. Look at $\mathcal{O}_K/\mathfrak{p}$; since $\mathfrak{p}$ is prime it's supposed to be an integral domain... but we claim that it's actually finite! To do this, we forget that we can multiply on $\mathcal{O}_K$. Recalling that $\mathcal{O}_K \cong \mathbb{Z}^{\oplus n}$ as an abelian group, we obtain a map

$$\mathbb{F}_p{}^{\oplus n} \cong \mathcal{O}_K/(p) \twoheadrightarrow \mathcal{O}_K/\mathfrak{p}.$$

Hence $|\mathcal{O}_K/\mathfrak{p}| \le p^n$ is *finite*. Since finite integral domains are fields (Problem 5D*) we are done. □

Since every nonzero prime $\mathfrak{p}$ is maximal, we now know that $\mathcal{O}_K$ is a Dedekind domain. Note that this tricky proof is essentially inspired by the solution to Problem 5G†.

[1] By Theorem 54.2.12.

Remark 55.3.7 — An alternative proof for the first part is: because $\mathfrak{p}$ is an ideal, $\alpha \cdot \mathcal{O}_K \subseteq \mathfrak{p}$, but $\alpha \cdot \mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank n, so $\mathfrak{p}$ is squeezed between two free $\mathbb{Z}$-modules of rank n, by Theorem 18.1.5 we must have $\mathfrak{p}$ is also free of rank n. So the quotient is finite, then use Lemma 56.8.8.

§55.4 Unique factorization works

Okay, I'll just say it now!

Unique factorization works perfectly in Dedekind domains!

Remark 55.4.1 (Comparison between Dedekind domain and UFD) — If we temporarily forget about the Noetherian and integrally closed condition, we have:

- An integral domain admits unique factorization of elements if the prime elements and the irreducible elements are the same.
- An integral domain admits unique factorization of ideals if the prime ideals and the maximal ideals are the same.

Notice the similarity — in either case, the Noetherian condition is "merely" to ensure that, if you keep extracting prime factors, you will terminate in a finite time.

Example 55.4.2 (What went wrong if $\mathcal{A}$ is not integrally closed?)

Consider $\mathcal{A} = 2\mathbb{Z}$, which is an ideal of $\mathbb{Z}$. Clearly, every nonzero prime ideal is maximal.

Nevertheless, in $\mathcal{A}$, $(2 \cdot 3 \cdot 5) = (60)$ is not a prime ideal (so of course it isn't a maximal ideal), but we cannot break it down into, for example, $(2 \cdot 3) \cdot (5)$.

Theorem 55.4.3 (Prime factorization works)

Let $\mathfrak{a}$ be a nonzero proper ideal of a Dedekind domain $\mathcal{A}$. Then $\mathfrak{a}$ can be written as a finite product of nonzero prime ideals $\mathfrak{p}_i$, say

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \mathfrak{p}_2^{e_2} \dots \mathfrak{p}_g^{e_g}$$

and this factorization is unique up to the order of the $\mathfrak{p}_i$.

Moreover, $\mathfrak{a}$ divides $\mathfrak{b}$ if and only if for every prime ideal $\mathfrak{p}$, the exponent of $\mathfrak{p}$ in $\mathfrak{a}$ is less than or equal to the corresponding exponent in $\mathfrak{b}$.

I won't write out the proof, but I'll describe the basic method of attack. Section 3 of [Ul08] does a nice job of explaining it. When we proved the fundamental theorem of arithmetic, the basic plot was:

(1) Show that if p is a rational prime[2] then $p \mid bc$ means $p \mid b$ or $p \mid c$. (This is called

[2] Note that the kindergarten definition of a prime is that "p isn't the product of two smaller integers". This isn't the correct definition of a prime: the definition of a prime is that $p \mid bc$ means $p \mid b$ or $p \mid c$. The kindergarten definition is something called "irreducible". Fortunately, in $\mathbb{Z}$, primes and irreducibles are the same thing, so no one ever told you that your definition of "prime" was wrong.

Euclid's Lemma.)

(2) Use strong induction to show that every $N > 1$ can be written as the product of primes (easy).

(3) Show that if $p_1 \dots p_m = q_1 \dots q_n$ for some primes (not necessarily unique), then $p_1 = q_i$ for some i, say q_1.

(4) Divide both sides by p_1 and use induction.

What happens if we try to repeat the proof here? We get step 1 for free, because we're using a better definition of "prime". We can also do step 3, since it follows from step 1. But step 2 doesn't work, because for abstract Dedekind domains we don't really have a notion of size. And step 4 doesn't work because we don't yet have a notion of what the inverse of a prime ideal is.

Well, it turns out that we *can* define the inverse $\mathfrak{a}^{-1}$ of an ideal, and I'll do so by the end of this chapter. You then need to check that $\mathfrak{a} \cdot \mathfrak{a}^{-1} = (1) = \mathcal{A}$. In fact, even this isn't easy. You have to check it's true for prime ideals $\mathfrak{p}$, *then* prove prime factorization, and then prove that this is true. Moreover, $\mathfrak{a}^{-1}$ is not actually an ideal, so you need to work in the field of fractions K instead of $\mathcal{A}$.

So the main steps in the new situation are as follows:

(1) First, show that every ideal $\mathfrak{a}$ divides $\mathfrak{p}_1 \dots \mathfrak{p}_g$ for some finite collection of primes. (This is an application of Zorn's Lemma.)

(2) Define $\mathfrak{p}^{-1}$ and show that $\mathfrak{p}\mathfrak{p}^{-1} = (1)$.

(3) Show that a factorization exists (again using Zorn's Lemma).

(4) Show that it's unique, using the new inverse we've defined.

Finally, let me comment on how nice this is if $\mathcal{A}$ is a PID (like $\mathbb{Z}$). Thus every element $a \in \mathcal{A}$ is in direct correspondence with an ideal (a). Now suppose (a) factors as a product of ideals $\mathfrak{p}_i = (p_i)$, say,

$$(a) = (p_1)^{e_1}(p_2)^{e_2} \dots (p_n)^{e_n}.$$

This verbatim reads

$$a = up_1^{e_1} p_2^{e_2} \dots p_n^{e_n}$$

where u is some unit (recall Definition 4.4.1). Hence, Dedekind domains which are PID's satisfy unique factorization for *elements*, just like in $\mathbb{Z}$. (In fact, the converse of this is true.)

§55.5 The factoring algorithm

Let's look at some examples from quadratic fields. Recall that if $K = \mathbb{Q}(\sqrt{d})$, then

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{d}] & d \equiv 2, 3 \pmod 4 \\ \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] & d \equiv 1 \pmod 4. \end{cases}$$

Also, recall that the norm of $a + b\sqrt{-d}$ is given by $a^2 + db^2$.

Example 55.5.1 (Factoring 6 in the integers of $\mathbb{Q}(\sqrt{-5})$)
Let $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ arise from $K = \mathbb{Q}(\sqrt{-5})$. We've already seen that

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

and you can't get any further with these principal ideals. But let

$$\mathfrak{p} = \left(1+\sqrt{-5}, 2\right) = \left(1-\sqrt{-5}, 2\right) \quad \text{and} \quad \mathfrak{q}_1 = (1+\sqrt{-5}, 3),\ \mathfrak{q}_2 = (1-\sqrt{-5}, 3).$$

Then it turns out $(6) = \mathfrak{p}^2\mathfrak{q}_1\mathfrak{q}_2$. More specifically, $(2) = \mathfrak{p}^2$, $(3) = \mathfrak{q}_1\mathfrak{q}_2$, and $(1+\sqrt{-5}) = \mathfrak{p}\mathfrak{q}_1$ and $(1-\sqrt{-5}) = \mathfrak{p}\mathfrak{q}_2$. (Proof in just a moment.)

I want to stress that all our ideals are computed relative to $\mathcal{O}_K$. So for example,

$$(2) = \{2x \mid x \in \mathcal{O}_K\}.$$

How do we know in this example that $\mathfrak{p}$ is prime/maximal? (Again, these are the same since we're in a Dedekind domain.) Answer: look at $\mathcal{O}_K/\mathfrak{p}$ and see if it's a field. There is a trick to this: we can express

$$\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}] \cong \mathbb{Z}[x]/(x^2+5).$$

So when we take *that* mod $\mathfrak{p}$, we get that

$$\mathcal{O}_K/\mathfrak{p} = \mathbb{Z}[x]/(x^2+5, 2, 1+x) \cong \mathbb{F}_2[x]/(x^2+5, x+1)$$

as rings.

> **Question 55.5.2.** Conclude that $\mathcal{O}_K/\mathfrak{p} \cong \mathbb{F}_2$, and satisfy yourself that $\mathfrak{q}_1$ and $\mathfrak{q}_2$ are also maximal.

I should give an explicit example of an ideal multiplication: let's compute

$$\begin{aligned}\mathfrak{q}_1\mathfrak{q}_2 &= \left((1+\sqrt{-5})(1-\sqrt{-5}), 3(1+\sqrt{-5}), 3(1-\sqrt{-5}), 9\right)\\ &= \left(6, 3+3\sqrt{-5}, 3-3\sqrt{-5}, 9\right)\\ &= \left(6, 3+3\sqrt{-5}, 3-3\sqrt{-5}, 3\right)\\ &= (3)\end{aligned}$$

where we first did $9-6=3$ (think Euclidean algorithm!), then noted that all the other generators don't contribute anything we don't already have with the 3 (again these are ideals computed in $\mathcal{O}_K$). You can do the computation for $\mathfrak{p}^2$, $\mathfrak{p}\mathfrak{q}_1$, $\mathfrak{p}\mathfrak{q}_2$ in the same way.

Finally, it's worth pointing out that we should quickly verify that $\mathfrak{p} \neq (x)$ for some x; in other words, that $\mathfrak{p}$ is not principal. Assume for contradiction that it is. Then x divides both $1+\sqrt{-5}$ and 2, in the sense that $1+\sqrt{-5} = \alpha_1 x$ and $2 = \alpha_2 x$ for some $\alpha_1, \alpha_2 \in \mathcal{O}_K$. (Principal ideals are exactly the "multiples" of x, so $(x) = x\mathcal{O}_K$.) Taking the norms, we find that $\mathrm{N}_{K/\mathbb{Q}}(x)$ divides both

$$\mathrm{N}_{K/\mathbb{Q}}(1+\sqrt{-5}) = 6 \quad \text{and} \quad \mathrm{N}_{K/\mathbb{Q}}(2) = 4.$$

Since $\mathfrak{p} \neq (1)$, x cannot be a unit, so its norm must be 2. But there are no elements of norm $2 = a^2 + 5b^2$ in $\mathcal{O}_K$.

Example 55.5.3 (Factoring 3 in the integers of $\mathbb{Q}(\sqrt{-17})$)
Let $\mathcal{O}_K = \mathbb{Z}[\sqrt{-17}]$ arise from $K = \mathbb{Q}(\sqrt{-17})$. We know $\mathcal{O}_K \cong \mathbb{Z}[x]/(x^2+17)$. Now

$$\mathcal{O}_K/3\mathcal{O}_K \cong \mathbb{Z}[x]/(3, x^2+17) \cong \mathbb{F}_3[x]/(x^2-1).$$

This already shows that (3) cannot be a prime (i.e. maximal) ideal, since otherwise our result should be a field. Anyways, we have a projection

$$\mathcal{O}_K \twoheadrightarrow \mathbb{F}_3[x]/\left((x-1)(x+1)\right).$$

Let $\mathfrak{q}_1$ be the pre-image of $(x-1)$ in the image, that is,

$$\mathfrak{q}_1 = (3, \sqrt{-17}-1).$$

Similarly,

$$\mathfrak{q}_2 = (3, \sqrt{-17}+1).$$

We have $\mathcal{O}_K/\mathfrak{q}_1 \cong \mathbb{F}_3$, so $\mathfrak{q}_1$ is maximal (prime). Similarly $\mathfrak{q}_2$ is prime. Magically, you can check explicitly that

$$\mathfrak{q}_1\mathfrak{q}_2 = (3).$$

Hence this is the factorization of (3) into prime ideals.

The fact that $\mathfrak{q}_1\mathfrak{q}_2 = (3)$ looks magical, but it's really true:

$$\begin{aligned}
\mathfrak{q}_1\mathfrak{q}_2 &= (3, \sqrt{-17}-1)(3, \sqrt{-17}+1) \\
&= (9, 3\sqrt{-17}+3, 3\sqrt{-17}-3, 18) \\
&= (9, 3\sqrt{-17}+3, 6) \\
&= (3, 3\sqrt{-17}+3, 6) \\
&= (3).
\end{aligned}$$

In fact, it turns out this always works in general: given a rational prime p, there is an algorithm to factor p in any $\mathcal{O}_K$ of the form $\mathbb{Z}[\theta]$.

Theorem 55.5.4 (Factoring algorithm / Dedekind-Kummer theorem)
Let K be a number field. Let $\theta \in \mathcal{O}_K$ with $|\mathcal{O}_K/\mathbb{Z}[\theta]| = j < \infty$, and let p be a prime not dividing j. Then $(p) = p\mathcal{O}_K$ is factored as follows:

> Let f be the minimal polynomial of θ and factor $\overline{f}$ mod p as
>
> $$\overline{f} \equiv \prod_{i=1}^{g} (\overline{f}_i)^{e_i} \pmod{p}.$$
>
> Then $\mathfrak{p}_i = (f_i(\theta), p)$ is prime for each i and the factorization of (p) is
>
> $$\mathcal{O}_K \supseteq (p) = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}.$$

In particular, if K is monogenic with $\mathcal{O}_K = \mathbb{Z}[\theta]$ then $j = 1$ and the theorem applies for all primes p.

In almost all our applications in this book, K will be monogenic; i.e. $j = 1$. Here $\overline{\psi}$ denotes the image in $\mathbb{F}_p[x]$ of a polynomial $\psi \in \mathbb{Z}[x]$.

> **Question 55.5.5.** There are many possible pre-images f_i we could have chosen (for example if $\overline{f_i} = x^2 + 1 \pmod 3$, we could pick $f_i = x^2 + 3x + 7$.) Why does this not affect the value of $\mathfrak{p}_i$?

Note that earlier, we could check the factorization worked for any particular case. The proof that this works is much the same, but we need one extra tool, the ideal norm. After that we leave the proposition as Problem 55E.

This algorithm gives us a concrete way to compute prime factorizations of (p) in any monogenic number field with $\mathcal{O}_K = \mathbb{Z}[\theta]$. To summarize the recipe:

1. Find the minimal polynomial of θ, say $f \in \mathbb{Z}[x]$.
2. Factor f mod p into irreducible polynomials $\overline{f_1}^{e_1} \overline{f_2}^{e_2} \dots \overline{f_g}^{e_g}$.
3. Compute $\mathfrak{p}_i = (f_i(\theta), p)$ for each i.

Then your $(p) = \mathfrak{p}_1^{e_1} \dots \mathfrak{p}_g^{e_g}$.

Or even shorter:

In order to factorize p in $\mathbb{Z}[x]/(f(x))$, we can factorize $f(x)$ in $\mathbb{Z}[x]/(p)$ instead.

Both are equivalent to factorizing 0 in $\mathbb{Z}[x]/(f(x), p)$ — in other words, writing $\mathbb{Z}[x]/(f(x), p)$ as a direct sum of $\mathbb{Z}[x]$-modules.

> **Exercise 55.5.6.** Factor (29) in $\mathbb{Z}[i]$ using the above algorithm.

§55.6 Fractional ideals

Prototypical example for this section: Analog to $\mathbb{Q}$ for $\mathbb{Z}$, allowing us to take inverses of ideals. Prime factorization works in the nicest way possible.

We now have a neat theory of factoring ideals of $\mathcal{A}$, just like factoring the integers. Now note that our factorization of $\mathbb{Z}$ naturally gives a way to factor elements of $\mathbb{Q}$; just factor the numerator and denominator separately.

Let's make the analogy clearer. The analogue of a rational number is as follows.

Definition 55.6.1. Let $\mathcal{A}$ be a Dedekind domain with field of fractions K. A **fractional ideal** J of K is a set of the form

$$J = \frac{1}{x} \cdot \mathfrak{a} \quad \text{where } x \in \mathcal{A}, \text{ and } \mathfrak{a} \text{ is an integral ideal.}$$

For emphasis, ideals of $\mathcal{A}$ will be sometimes referred to as **integral ideals**.

You might be a little surprised by this definition: one would expect that a fractional ideal should be of the form $\frac{\mathfrak{a}}{\mathfrak{b}}$ for some integral ideals $\mathfrak{a}$, $\mathfrak{b}$. But in fact, it suffices to just take $x \in \mathcal{A}$ in the denominator. The analogy is that when we looked at $\mathcal{O}_K$, we found that we only needed integer denominators: $\frac{1}{4-\sqrt{3}} = \frac{1}{13}(4+\sqrt{3})$. Similarly here, it will turn out that we only need to look at $\frac{1}{x} \cdot \mathfrak{a}$ rather than $\frac{\mathfrak{a}}{\mathfrak{b}}$, and so we define it this way from the beginning. See Problem 55D[†] for a different equivalent definition.

Example 55.6.2 ($\frac{5}{2}\mathbb{Z}$ is a fractional ideal)

The set

$$\frac{5}{2}\mathbb{Z} = \left\{ \frac{5}{2}n \mid n \in \mathbb{Z} \right\} = \frac{1}{2}(5)$$

is a fractional ideal of $\mathbb{Z}$.

Now, as we prescribed, the fractional ideals form a multiplicative group:

Theorem 55.6.3 (Fractional ideals form a group)

Let $\mathcal{A}$ be a Dedekind domain and K its field of fractions. For any integral ideal $\mathfrak{a}$, the set

$$\mathfrak{a}^{-1} = \{x \in K \mid x\mathfrak{a} \subseteq (1) = \mathcal{A}\}$$

is a fractional ideal with $\mathfrak{a}\mathfrak{a}^{-1} = (1)$.

(This result is nontrivial. To prove $\mathfrak{a}\mathfrak{a}^{-1} = (1)$, one approach is to prove it first for prime $\mathfrak{a}$, then consider the factorization of $\mathfrak{a}$ into prime ideals.)

Definition 55.6.4. Thus nonzero fractional ideals of K form a group under multiplication with identity $(1) = \mathcal{A}$. This **ideal group** is denoted J_K.

Example 55.6.5 ($(3)^{-1}$ in $\mathbb{Z}$)

Please check that in $\mathbb{Z}$ we have

$$(3)^{-1} = \left\{ \frac{1}{3}n \mid n \in \mathbb{Z} \right\} = \frac{1}{3}\mathbb{Z}.$$

It follows that every fractional ideal J can be uniquely written as

$$J = \prod_i \mathfrak{p}_i^{n_i} \cdot \prod_i \mathfrak{q}_i^{-m_i}$$

where n_i and m_i are positive integers. In fact, $\mathfrak{a}$ is an integral ideal if and only if all its exponents are nonnegative, just like the case with integers. So, a perhaps better way to think about fractional ideals is as products of prime ideals, possibly with negative exponents.

§55.7 The ideal norm

One last tool is the ideal norm, which gives us a notion of the "size" of an ideal.

Definition 55.7.1. The **ideal norm** (or absolute norm) of a nonzero ideal $\mathfrak{a} \subseteq \mathcal{O}_K$ is defined as $|\mathcal{O}_K/\mathfrak{a}|$ and denoted $\mathrm{N}(\mathfrak{a})$.

Example 55.7.2 (Ideal norm of (5) in the Gaussian integers)

Let $K = \mathbb{Q}(i)$, $\mathcal{O}_K = \mathbb{Z}[i]$. Consider the ideal (5) in $\mathcal{O}_K$. We have that

$$\mathcal{O}_K/(5) \cong \{a + bi \mid a, b \in \mathbb{Z}/5\mathbb{Z}\}$$

so (5) has ideal norm 25, corresponding to the fact that $\mathcal{O}_K/(5)$ has $5^2 = 25$ elements.

Example 55.7.3 (Ideal norm of $(2+i)$ in the Gaussian integers)
You'll notice that

$$\mathcal{O}_K/(2+i) \cong \mathbb{F}_5$$

since mod $2+i$ we have both $5 \equiv 0$ and $i \equiv -2$. (Indeed, since $(2+i)$ is prime we had better get a field!) Thus $\mathrm{N}\left((2+i)\right) = 5$; similarly $\mathrm{N}\left((2-i)\right) = 5$.

Thus the ideal norm measures how "roomy" the ideal is: that is, (5) is a lot more spaced out in $\mathbb{Z}[i]$ than it is in $\mathbb{Z}$. (This intuition will be important when we will actually view $\mathcal{O}_K$ as a lattice.)

Question 55.7.4. What are the ideals with ideal norm one?

Our example with (5) suggests several properties of the ideal norm which turn out to be true:

Lemma 55.7.5 (Properties of the absolute norm)
Let $\mathfrak{a}$ be a nonzero ideal of $\mathcal{O}_K$.

(a) $\mathrm{N}(\mathfrak{a})$ is finite.

(b) For any other nonzero ideal $\mathfrak{b}$, $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$.

(c) If $\mathfrak{a} = (a)$ is principal, then $\mathrm{N}(\mathfrak{a}) = \left|\mathrm{N}_{K/\mathbb{Q}}(a)\right|$.

I unfortunately won't prove these properties, though we already did (a) in our proof that $\mathcal{O}_K$ was a Dedekind domain.

As with the case of the norm of an element, the ideal norm also has a geometrical interpretation: Recall that if $\mathfrak{a} = (a)$, let μ_a be the multiplication-by-a map, then $\mathrm{N}_{K/\mathbb{Q}}(a) = \left|\det \mu_a\right|$ measures how much K is stretched under μ_a when viewed as a $\mathbb{Q}$-vector space.

Exercise 55.7.6. Convince yourself that if $\mu_a(\mathcal{O}_K) \subseteq \mathcal{O}_K$, then $|\mathcal{O}_K/a\mathcal{O}_K|$ is exactly equal to $\left|\det \mu_a\right|$.

This explains why $\mathrm{N}(\mathfrak{a}) = \left|\mathrm{N}_{K/\mathbb{Q}}(a)\right|$, although note that $\mathfrak{a} = (a) = (-a)$, so there need not be an unique multiplication-by-$\mathfrak{a}$ map.

The fact that N is completely multiplicative lets us also consider the norm of a fractional ideal J by the natural extension

$$J = \prod_i \mathfrak{p}_i^{n_i} \cdot \prod_i \mathfrak{q}_i^{-m_i} \qquad \Longrightarrow \qquad \mathrm{N}(J) := \frac{\prod_i \mathrm{N}(\mathfrak{p}_i)^{n_i}}{\prod_i \mathrm{N}(\mathfrak{q}_i)^{m_i}}.$$

Thus N is a natural group homomorphism $J_K \to \mathbb{Q}^\times$.

§55.8 A few harder problems to think about

Problem 55A. Show that there are three different factorizations of 77 in $\mathcal{O}_K$, where $K = \mathbb{Q}(\sqrt{-13})$.

Problem 55B. Let $K = \mathbb{Q}(\sqrt[3]{2})$; take for granted that $\mathcal{O}_K = \mathbb{Z}[\sqrt[3]{2}]$. Find the factorization of (5) in $\mathcal{O}_K$.

Problem 55C (Fermat's little theorem)**.** Let $\mathfrak{p}$ be a prime ideal in some ring of integers $\mathcal{O}_K$. Show that for $\alpha \in \mathcal{O}_K$,

$$\alpha^{\mathrm{N}(\mathfrak{p})} \equiv \alpha \pmod{\mathfrak{p}}.$$

Problem 55D†. Let $\mathcal{A}$ be a Dedekind domain with field of fractions K, and pick $J \subseteq K$. Show that J is a fractional ideal if and only if

(i) J is closed under addition and multiplication by elements of $\mathcal{A}$, and

(ii) J is finitely generated as an abelian group.

More succinctly: J is a fractional ideal $\iff$ J is a finitely generated $\mathcal{A}$-module.

Problem 55E. In the notation of Theorem 55.5.4, let $I = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$. Assume for simplicity that K is monogenic, hence $\mathcal{O}_K = \mathbb{Z}[\theta]$.

(a) Prove that each $\mathfrak{p}_i$ is prime.

(b) Show that (p) divides I.

(c) Use the norm to show that $(p) = I$.

56 Minkowski bound and class groups

We now have a neat theory of unique factorization of ideals. In the case of a PID, this in fact gives us a UFD. Sweet.

We'll define, in a moment, something called the *class group* which measures how far $\mathcal{O}_K$ is from being a PID; the bigger the class group, the farther $\mathcal{O}_K$ is from being a PID. In particular, $\mathcal{O}_K$ is a PID if it has trivial class group.

Then we will provide some inequalities which let us put restrictions on the class group; for instance, this will let us show in some cases that the class group must be trivial. Astonishingly, the proof will use Minkowski's theorem, a result from geometry.

§56.1 The class group

Prototypical example for this section: PID's have trivial class group.

Let K be a number field, and let J_K denote the multiplicative group of fractional ideals of $\mathcal{O}_K$. Let P_K denote the multiplicative group of **principal fractional ideals**: those of the form $(x) = x\mathcal{O}_K$ for some $x \in K$.

> **Question 56.1.1.** Check that P_K is also a multiplicative group. (This is really easy: name $x\mathcal{O}_K \cdot y\mathcal{O}_K$ and $(x\mathcal{O}_K)^{-1}$.)

As J_K is abelian, we can now define the **class group** (or **ideal class group**) to be the quotient

$$\mathrm{Cl}_K := J_K/P_K.$$

The elements of Cl_K are called **classes**.

Equivalently,

The class group Cl_K is the set of nonzero fractional ideals modulo scaling by a constant in K.

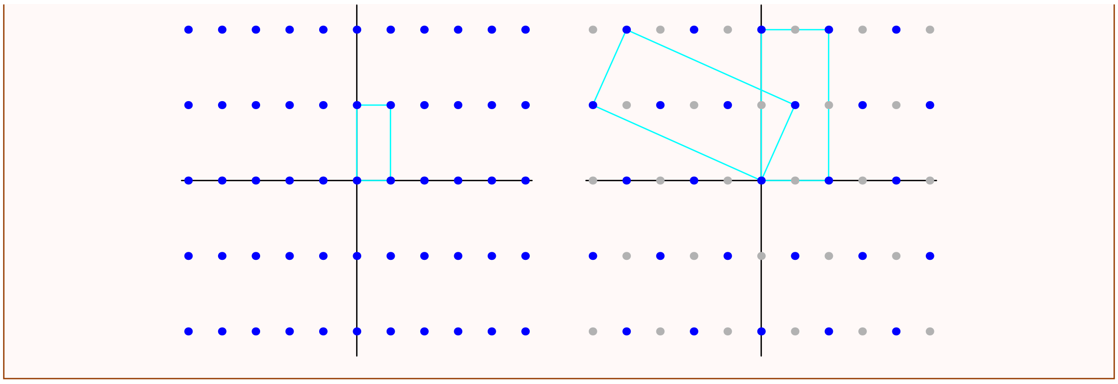
You can also think of the classes as the "shapes" of the ideals, as two ideals belong to the same class if and only if they're isomorphic as $\mathcal{O}_K$-modules.

Example 56.1.2 (Ideal classes in $\mathbb{Q}(\sqrt{-5})$)

If the field is an imaginary quadratic field, visualizing the class of an ideal is really easy: because multiplication by a complex number corresponds to a combination of a scaling and a rotation (i.e. it preserves angles), two ideals belong to the same class if they are similar, that is, you can overlap one onto the another using rotation and scaling.

When the field is $K = \mathbb{Q}(\sqrt{-5})$, the ring of integers is $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$.

The first picture below depicts the ideal $(1) \subseteq \mathcal{O}_K$. The second picture depicts $(2, 1+\sqrt{-5}) \subseteq \mathcal{O}_K$, which is not a principal ideal.



In particular, Cl_K is trivial if all ideals are principal, since the nonzero principal ideals are the same up to scaling.

The size of the class group is called the **class number**. It's a beautiful theorem that the class number is always finite, and the bulk of this chapter will build up to this result. It requires several ingredients.

§56.2 The discriminant of a number field

Prototypical example for this section: Quadratic fields.

Let's say I have $K = \mathbb{Q}(\sqrt{2})$. As we've seen before, this means $\mathcal{O}_K = \mathbb{Z}[\sqrt{2}]$, meaning

$$\mathcal{O}_K = \left\{ a + b\sqrt{2} \mid a, b \in \mathbb{Z} \right\}.$$

The key insight now is that you might think of this as a *lattice*: geometrically, we want to think about this the same way we think about $\mathbb{Z}^2$.

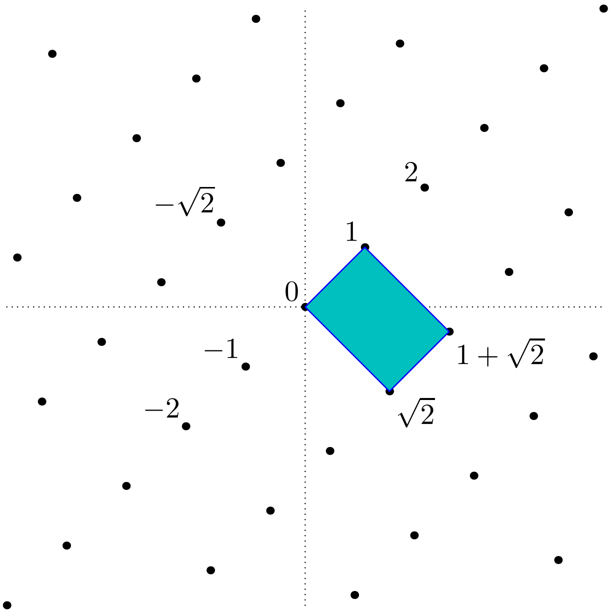
Perversely, we might try to embed this into $\mathbb{Q}^2$ by sending $a + b\sqrt{2}$ to (a, b). But this is a little stupid, since we're rudely making K, which somehow lives inside $\mathbb{R}$ and is "one-dimensional" in that sense, into a two-dimensional space. It also depends on a choice of basis, which we don't like. A better way is to think about the fact that there are two embeddings $\sigma_1 \colon K \to \mathbb{C}$ and $\sigma_2 \colon K \to \mathbb{C}$, namely the identity, and conjugation:

$$\begin{aligned} \sigma_1(a + b\sqrt{2}) &= a + b\sqrt{2} \\ \sigma_2(a + b\sqrt{2}) &= a - b\sqrt{2}. \end{aligned}$$

Fortunately for us, these embeddings both have real image. This leads us to consider the set of points

$$(\sigma_1(\alpha), \sigma_2(\alpha)) \in \mathbb{R}^2 \quad \text{as} \quad \alpha \in K.$$

This lets us visualize what $\mathcal{O}_K$ looks like in $\mathbb{R}^2$. The points of K are dense in $\mathbb{R}^2$, but the points of $\mathcal{O}_K$ cut out a lattice.



To see how big the lattice is, we look at how $\{1, \sqrt{2}\}$, the generators of $\mathcal{O}_K$, behave. The point corresponding to $a + b\sqrt{2}$ in the lattice is

$$a \cdot (1,1) + b \cdot (\sqrt{2}, -\sqrt{2}).$$

The **mesh** of the lattice[1] is defined as the hypervolume of the "fundamental parallelepiped" I've colored blue above. For this particular case, it ought to be equal to the area of that parallelogram, which is

$$\det \begin{bmatrix} 1 & -\sqrt{2} \\ 1 & \sqrt{2} \end{bmatrix} = 2\sqrt{2}.$$

The definition of the discriminant is precisely this, except with an extra square factor (since permutation of rows could lead to changes in sign in the matrix above). Problem 57B* shows that the squaring makes Δ_K an integer.

To make the next definition, we invoke:

Theorem 56.2.1 (The n embeddings of a number field)
Let K be a number field of degree n. Then there are exactly n field homomorphisms $K \hookrightarrow \mathbb{C}$, say $\sigma_1, \dots, \sigma_n$, which fix $\mathbb{Q}$.

Proof. Deferred to Theorem 59.3.1, once we have the tools of Galois theory. □

In fact, in Theorem 59.3.4 we see that for $\alpha \in K$, we have that $\sigma_i(\alpha)$ runs over the conjugates of α as $i = 1, \dots, n$. It follows that

$$\operatorname{Tr}_{K/\mathbb{Q}}(\alpha) = \sum_{i=1}^{n} \sigma_i(\alpha) \quad \text{and} \quad \operatorname{N}_{K/\mathbb{Q}}(\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha).$$

This allows us to define:

[1] Most authors call this the volume, but I think this is not the right word to use – lattices have "volume" zero since they are just a bunch of points! In contrast, the English word "mesh" really does refer to the width of a "gap".

Definition 56.2.2. Suppose $\alpha_1, \dots, \alpha_n$ is a $\mathbb{Z}$-basis of $\mathcal{O}_K$. The **discriminant** of the number field K is defined by

$$\Delta_K := \det \begin{bmatrix} \sigma_1(\alpha_1) & \dots & \sigma_n(\alpha_1) \\ \vdots & \ddots & \vdots \\ \sigma_1(\alpha_n) & \dots & \sigma_n(\alpha_n) \end{bmatrix}^2 .$$

This does not depend on the choice of the $\{\alpha_i\}$; we will not prove this here.

Example 56.2.3 (Discriminant of $K = \mathbb{Q}(\sqrt{2})$)

We have $\mathcal{O}_K = \mathbb{Z}[\sqrt{2}]$ and as discussed above the discriminant is

$$\Delta_K = (-2\sqrt{2})^2 = 8.$$

Example 56.2.4 (Discriminant of $\mathbb{Q}(i)$)

Let $K = \mathbb{Q}(i)$. We have $\mathcal{O}_K = \mathbb{Z}[i] = \mathbb{Z} \oplus i\mathbb{Z}$. The embeddings are the identity and complex conjugation which take 1 to $(1,1)$ and i to $(i, -i)$. So

$$\Delta_K = \det \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^2 = (-2i)^2 = -4.$$

This example illustrates that the discriminant need not be positive for number fields which wander into the complex plane (the lattice picture is a less perfect analogy). But again, as we'll prove in the problems the discriminant is always an integer.

Example 56.2.5 (Discriminant of $\mathbb{Q}(\sqrt{5})$)

Let $K = \mathbb{Q}(\sqrt{5})$. This time, $\mathcal{O}_K = \mathbb{Z} \oplus \frac{1+\sqrt{5}}{2}\mathbb{Z}$, and so the discriminant is going to look a little bit different. The embeddings are still $a + b\sqrt{5} \mapsto a + b\sqrt{5}, a - b\sqrt{5}$.

Applying this to the $\mathbb{Z}$-basis $\left\{1, \frac{1+\sqrt{5}}{2}\right\}$, we get

$$\Delta_K = \det \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}^2 = (-\sqrt{5})^2 = 5.$$

Exercise 56.2.6. Extend all this to show that if $K = \mathbb{Q}(\sqrt{d})$ for $d \neq 1$ squarefree, we have

$$\Delta_K = \begin{cases} d & \text{if } d \equiv 1 \pmod 4 \\ 4d & \text{if } d \equiv 2, 3 \pmod 4. \end{cases}$$

Actually, let me point out something curious: recall that the polynomial discriminant of $Ax^2 + Bx + C$ is $B^2 - 4AC$. Then:

- In the $d \equiv 1 \pmod 4$ case, Δ_K is the discriminant of $x^2 - x - \frac{d-1}{4}$, which is the minimal polynomial of $\frac{1}{2}(1 + \sqrt{d})$. Of course, $\mathcal{O}_K = \mathbb{Z}[\frac{1}{2}(1 + \sqrt{d})]$.
- In the $d \equiv 2, 3 \pmod 4$ case, Δ_K is the discriminant of $x^2 - d$ which is the minimal polynomial of $\sqrt{d}$. Once again, $\mathcal{O}_K = \mathbb{Z}[\sqrt{d}]$.

This is not a coincidence! Problem 57C* asserts that this is true in general; hence the name "discriminant".

§56.3 The signature of a number field

Prototypical example for this section: $\mathbb{Q}(\sqrt[100]{2})$ *has signature* $(2, 49)$.

In the example of $K = \mathbb{Q}(i)$, we more or less embedded K into the space $\mathbb{C}$. However, K is a degree two extension, so what we'd really like to do is embed it into $\mathbb{R}^2$. To do so, we're going to take advantage of complex conjugation.

Let K be a number field and $\sigma_1, \dots, \sigma_n$ be its embeddings. We distinguish between the **real embeddings** (which map all of K into $\mathbb{R}$) and the **complex embeddings** (which map some part of K outside $\mathbb{R}$). Notice that if σ is a complex embedding, then so is the conjugate $\overline{\sigma} \neq \sigma$; hence complex embeddings come in pairs.

Definition 56.3.1. Let K be a number field of degree n, and set

$$\begin{aligned} r_1 &= \text{number of real embeddings} \\ r_2 &= \text{number of pairs of complex embeddings.} \end{aligned}$$

The **signature** of K is the pair (r_1, r_2). Observe that $r_1 + 2r_2 = n$.

Example 56.3.2 (Basic examples of signatures)

(a) $\mathbb{Q}$ has signature $(1, 0)$.

(b) $\mathbb{Q}(\sqrt{2})$ has signature $(2, 0)$.

(c) $\mathbb{Q}(i)$ has signature $(0, 1)$.

(d) Let $K = \mathbb{Q}(\sqrt[3]{2})$, and let ω be a cube root of unity. The elements of K are

$$K = \left\{ a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in \mathbb{Q} \right\}.$$

Then the signature is $(1, 1)$, because the three embeddings are

$$\sigma_1 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}, \quad \sigma_2 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}\omega, \quad \sigma_3 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}\omega^2.$$

The first of these is real and the latter two are conjugate pairs.

Example 56.3.3 (Even more signatures)

In the same vein $\mathbb{Q}(\sqrt[99]{2})$ and $\mathbb{Q}(\sqrt[100]{2})$ have signatures $(1, 49)$ and $(2, 49)$.

> **Question 56.3.4.** Verify the signatures of the above two number fields.

From now on, we will number the embeddings of K in such a way that

$$\sigma_1, \sigma_2, \dots, \sigma_{r_1}$$

are the real embeddings, while

$$\sigma_{r_1+1} = \overline{\sigma_{r_1+r_2+1}}, \quad \sigma_{r_1+2} = \overline{\sigma_{r_1+r_2+2}}, \quad \dots, \quad \sigma_{r_1+r_2} = \overline{\sigma_{r_1+2r_2}}.$$

are the r_2 pairs of complex embeddings. We define the **canonical embedding** of K as

$$K \overset{\iota}{\hookrightarrow} \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \quad \text{by} \quad \alpha \mapsto \left(\sigma_1(\alpha), \dots, \sigma_{r_1}(\alpha), \sigma_{r_1+1}(\alpha), \dots, \sigma_{r_1+r_2}(\alpha)\right).$$

All we've done is omit, for the complex case, the second of the embeddings in each conjugate pair. This is no big deal, since they are just conjugates; the above tuple is all the information we need.

For reasons that will become obvious in a moment, I'll let τ denote the isomorphism

$$\tau \colon \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \xrightarrow{\sim} \mathbb{R}^{r_1+2r_2} = \mathbb{R}^n$$

by breaking each complex number into its real and imaginary part, as

$$\begin{aligned} \alpha \mapsto \big(&\sigma_1(\alpha), \dots, \sigma_{r_1}(\alpha), \\ &\operatorname{Re}\sigma_{r_1+1}(\alpha), \ \operatorname{Im}\sigma_{r_1+1}(\alpha), \\ &\operatorname{Re}\sigma_{r_1+2}(\alpha), \ \operatorname{Im}\sigma_{r_1+2}(\alpha), \\ &\dots, \\ &\operatorname{Re}\sigma_{r_1+r_2}(\alpha), \ \operatorname{Im}\sigma_{r_1+r_2}(\alpha)\big). \end{aligned}$$

Example 56.3.5 (Example of canonical embedding)

As before let $K = \mathbb{Q}(\sqrt[3]{2})$ and set

$$\sigma_1 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}, \quad \sigma_2 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}\omega, \quad \sigma_3 \colon \sqrt[3]{2} \mapsto \sqrt[3]{2}\omega^2$$

where $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, noting that we've already arranged indices so $\sigma_1 = \mathrm{id}$ is real while σ_2 and σ_3 are a conjugate pair. So the embeddings $K \overset{\iota}{\hookrightarrow} \mathbb{R} \times \mathbb{C} \xrightarrow{\sim} \mathbb{R}^3$ are given by

$$\alpha \overset{\iota}{\longmapsto} (\sigma_1(\alpha), \sigma_2(\alpha)) \overset{\tau}{\longmapsto} \left(\sigma_1(\alpha), \ \operatorname{Re}\sigma_2(\alpha), \ \operatorname{Im}\sigma_2(\alpha)\right).$$

For concreteness, taking $\alpha = 9 + \sqrt[3]{2}$ gives

$$\begin{aligned} 9 + \sqrt[3]{2} &\overset{\iota}{\mapsto} \left(9 + \sqrt[3]{2}, \ 9 + \sqrt[3]{2}\omega\right) \\ &= \left(9 + \sqrt[3]{2}, \ 9 - \frac{1}{2}\sqrt[3]{2} + \frac{\sqrt[6]{108}}{2} i\right) \in \mathbb{R} \times \mathbb{C} \\ &\overset{\tau}{\mapsto} \left(9 + \sqrt[3]{2}, \ 9 - \frac{1}{2}\sqrt[3]{2}, \ \frac{\sqrt[6]{108}}{2}\right) \in \mathbb{R}^3. \end{aligned}$$

Now, the whole point of this is that we want to consider the resulting lattice when we take $\mathcal{O}_K$. In fact, we have:

Lemma 56.3.6

Consider the composition of the embeddings $K \hookrightarrow \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \xrightarrow{\sim} \mathbb{R}^n$. Then as before, $\mathcal{O}_K$ becomes a lattice L in $\mathbb{R}^n$, with mesh equal to

$$\frac{1}{2^{r_2}}\sqrt{|\Delta_K|}.$$

Proof. Fun linear algebra problem (you just need to manipulate determinants). Left as Problem 56D. □

From this we can deduce:

Lemma 56.3.7

Consider the composition of the embeddings $K \hookrightarrow \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \xrightarrow{\sim} \mathbb{R}^n$. Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$. Then the image of $\mathfrak{a}$ is a lattice $L_{\mathfrak{a}}$ in $\mathbb{R}^n$ with mesh equal to

$$\frac{\mathrm{N}(\mathfrak{a})}{2^{r_2}}\sqrt{|\Delta_K|}.$$

Sketch of Proof. Let

$$d = \mathrm{N}(\mathfrak{a}) := |\mathcal{O}_K/\mathfrak{a}|.$$

Then in the lattice $L_{\mathfrak{a}}$, we somehow only take $\frac{1}{d}$th of the points which appear in the lattice L, which is why the area increases by a factor of $\mathrm{N}(\mathfrak{a})$. To make this all precise I would need to do a lot more with lattices and geometry than I have space for in this chapter, so I will omit the details. But I hope you can see why this is intuitively true. □

§56.4 Minkowski's theorem

Now I can tell you why I insisted we move from $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ to $\mathbb{R}^n$. In geometry, there's a really cool theorem of Minkowski's that goes as follows.

Theorem 56.4.1 (Minkowski)

Let $S \subseteq \mathbb{R}^n$ be a convex set containing 0 which is centrally symmetric (meaning that $x \in S \iff -x \in S$). Let L be a lattice with mesh d. If either

(a) The volume of S exceeds $2^n d$, or

(b) The volume of S equals $2^n d$ and S is compact,

then S contains a nonzero lattice point of L.

> **Question 56.4.2.** Show that the condition $0 \in S$ is actually extraneous in the sense that any nonempty, convex, centrally symmetric set contains the origin.

Sketch of Proof. Part (a) is surprisingly simple and has a very olympiad-esque solution: it's basically Pigeonhole on areas. We'll prove part (a) in the special case $n = 2$, $L = \mathbb{Z}^2$ for simplicity as the proof can easily be generalized to any lattice and any n. Thus we want to show that any such convex set S with area more than 4 contains a lattice point.

Dissect the plane into 2×2 squares

$$[2a-1, 2a+1] \times [2b-1, 2b+1]$$

and overlay all these squares on top of each other. By the Pigeonhole Principle, we find there exist two points $p \neq q \in S$ which is mapped to the same point. Since S is symmetric, $-q \in S$. Then $\frac{1}{2}(p-q) \in S$ (convexity) and is a nonzero lattice point.

I'll briefly sketch part (b): the idea is to consider $(1+\varepsilon)S$ for $\varepsilon > 0$ (this is "S magnified by a small factor $1+\varepsilon$"). This satisfies condition (a). So for each $\varepsilon > 0$ the set of nonzero lattice points in $(1+\varepsilon)S$, say S_ε, is a *finite nonempty set* of (discrete) points (the "finite" part follows from the fact that $(1+\varepsilon)S$ is bounded). So there has to be some point that's in S_ε for every $\varepsilon > 0$ (why?), which implies it's in S. □

§56.5 The trap box

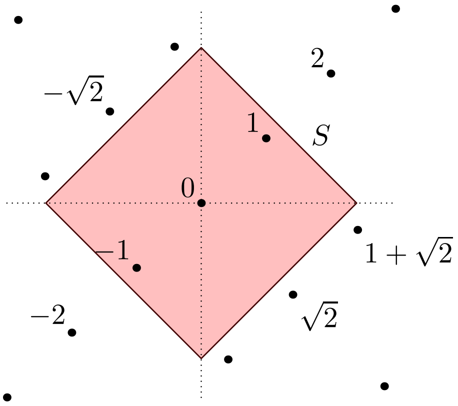
The last ingredient we need is a set to apply Minkowski's theorem to. I propose:

Definition 56.5.1. Let M be a positive real. In $\mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$, define the box S to be the set of points $(x_1, \dots, x_{r_1}, z_1, \dots, z_{r_2})$ such that

$$\sum_{i=1}^{r_1} |x_i| + 2\sum_{j=1}^{r_2} |z_j| \le M.$$

Note that this depends on the value of M.

Think of this box as a *mousetrap*: anything that falls in it is going to have a small norm, and our goal is to use Minkowski to lure some nonzero element into it.



That is, suppose $\alpha \in \mathfrak{a}$ falls into the box I've defined above, which means

$$M \ge \sum_{i=1}^{r_1} |\sigma_i(\alpha)| + 2 \sum_{i=r_1+1}^{r_1+r_2} |\sigma_i(\alpha)| = \sum_{i=1}^{n} |\sigma_i(\alpha)|,$$

where we are remembering that the last few σ's come in conjugate pairs. This looks like the trace, but the absolute values are in the way. So instead, we apply AM-GM to obtain:

Lemma 56.5.2 (Effect of the mousetrap)
Let $\alpha \in \mathcal{O}_K$, and suppose $\iota(\alpha)$ is in S (where $\iota \colon K \hookrightarrow \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$ as usual). Then

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \prod_{i=1}^{n} |\sigma_i(\alpha)| \le \left(\frac{M}{n}\right)^n.$$

The last step we need to do is compute the volume of the box. This is again some geometry I won't do, but take my word for it:

Lemma 56.5.3 (Size of the mousetrap)

Let $\tau \colon \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \xrightarrow{\sim} \mathbb{R}^n$ as before. Then the image of S under τ is a convex, compact, centrally symmetric set with volume

$$2^{r_1} \cdot \left(\frac{\pi}{2}\right)^{r_2} \cdot \frac{M^n}{n!}.$$

> **Question 56.5.4.** (Sanity check) Verify that the above is correct for the signatures $(r_1, r_2) = (2, 0)$ and $(r_1, r_2) = (0, 1)$, which are the possible signatures when $n = 2$.

§56.6 The Minkowski bound

We can now put everything we have together to obtain the great Minkowski bound.

Theorem 56.6.1 (Minkowski bound)

Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be any nonzero ideal. Then there exists $0 \neq \alpha \in \mathfrak{a}$ such that

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha) \le \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n} \sqrt{|\Delta_K|} \cdot \mathrm{N}(\mathfrak{a}).$$

Proof. This is a matter of putting all our ingredients together. Let's see what things we've defined already:

$$\begin{tikzcd}
K \arrow[r, hook, "\iota"] & \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \arrow[r, "\tau"] & \mathbb{R}^n & \\
& \text{box } S \arrow[r, mapsto] & \tau^{\text{img}}(S) & \text{with volume } 2^{r_1} \left(\frac{\pi}{2}\right)^{r_2} \frac{M^n}{n!} \\
\mathcal{O}_K \arrow[rr, mapsto] & & \text{Lattice } L & \text{with mesh } 2^{-r_2} \sqrt{|\Delta_K|} \\
\mathfrak{a} \arrow[rr, mapsto] & & \text{Lattice } L_{\mathfrak{a}} & \text{with mesh } 2^{-r_2} \sqrt{|\Delta_K|}\, \mathrm{N}(\mathfrak{a})
\end{tikzcd}$$

Pick a value of M such that the mesh of $L_{\mathfrak{a}}$ equals 2^{-n} of the volume of the box. Then Minkowski's theorem gives that some $0 \neq \alpha \in \mathfrak{a}$ lands inside the box — the mousetrap is configured to force $\mathrm{N}_{K/\mathbb{Q}}(\alpha) \le \frac{1}{n^n} M^n$. The correct choice of M is

$$M^n = M^n \cdot 2^n \cdot \frac{\text{mesh}}{\text{vol box}} = 2^n \cdot \frac{n!}{2^{r_1} \cdot \left(\frac{\pi}{2}\right)^{r_2}} \cdot 2^{-r_2} \sqrt{|\Delta_K|}\, \mathrm{N}(\mathfrak{a})$$

which gives the bound after some arithmetic. □

§56.7 The class group is finite

Definition 56.7.1. Let $M_K = \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n} \sqrt{|\Delta_K|}$ for brevity. Note that it is a constant depending on K.

So that's cool and all, but what we really wanted was to show that the class group is finite. How can the Minkowski bound help? The key idea is the following:

The class of $\mathfrak{a}$ is entirely determined by $(\alpha) \cdot \mathfrak{a}^{-1}$.

Question 56.7.2. Verify this. (That is, if $\mathfrak{a}$ and $\mathfrak{b}$ are such that $(\alpha) \cdot \mathfrak{a}^{-1} = (\beta) \cdot \mathfrak{b}^{-1}$ for some $\alpha \in \mathfrak{a}$, $\beta \in \mathfrak{b}$, then prove that $\mathfrak{a} = (\gamma)\mathfrak{b}$ for some $\gamma \in K$.)

Example 56.7.3

Recall this example:

$$(6) = (2, 1 - \sqrt{-5})^2 (3, 1 + \sqrt{-5})(3, 1 - \sqrt{-5}) = \mathfrak{p}^2 \mathfrak{q}_1 \mathfrak{q}_2.$$

Consider $\mathfrak{a} = (2)$ being principal, pick $\alpha = 2$, then $(\alpha) \cdot \mathfrak{a}^{-1} = (1)$. If $(\alpha) \cdot \mathfrak{a}^{-1} = (1)$ (or (2), or anything principal), we know $\mathfrak{a}$ must be principal.

On the other hand, $\mathfrak{q}_1 = (3, 1 + \sqrt{-5})$ is not principal, pick $\alpha = 3$. We know $(3) = \mathfrak{q}_1 \mathfrak{q}_2$, so $(\alpha) \cdot \mathfrak{q}_1^{-1} = \mathfrak{q}_2 \neq (1)$.

In both examples above, $\mathfrak{a} \mid (\alpha)$, so their quotient should be an "integer". Indeed:

Question 56.7.4. Show that $(\alpha) \cdot \mathfrak{a}^{-1}$ is an integral ideal. (Unwind definitions.)

You might notice that we can rewrite the Minkowski bound to say

$$\mathrm{N}\left((\alpha) \cdot \mathfrak{a}^{-1}\right) \le M_K$$

where M_K is some constant depending on K.

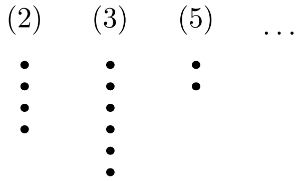
This statement is helpful, because in fact, there are only finitely many integral ideals with norm $\le M_K$.

Corollary 56.7.5 (Finiteness of class group)

Class groups are always finite.

Proof. We just have to show there are finitely many integral ideals as above; this will mean there are finitely many classes.

Suppose we want to build such an ideal $\mathfrak{a} = \mathfrak{p}_1^{e_1} \dots \mathfrak{p}_m^{e_m}$. Recall that a prime ideal $\mathfrak{p}_i$ must have some rational prime p inside it, meaning $\mathfrak{p}_i$ divides (p) and p divides $\mathrm{N}(\mathfrak{p}_i)$. So let's group all the $\mathfrak{p}_i$ we want to build $\mathfrak{a}$ with based on which (p) they came from.



To be more dramatic: imagine you have a *cherry tree*; each branch corresponds to a prime (p) and contains as cherries (prime ideals) the factors of (p) (finitely many). Your bucket (the ideal $\mathfrak{a}$ you're building) can only hold a total weight (norm) of M_K. So you can't even touch the branches higher than M_K. You can repeat cherries (oops), but the weight of a cherry on branch (p) is definitely $\ge p$; all this means that the number of ways to build $\mathfrak{a}$ is finite. □

§56.8 Computation of class numbers

Definition 56.8.1. The order of Cl_K is called the **class number** of K.

> **Remark 56.8.2** — If $\mathrm{Cl}_K = 1$, then $\mathcal{O}_K$ is a PID, hence a UFD.

By computing the actual value of M_K, we can quite literally build the entire "cherry tree" mentioned in the previous proof. Let's give an example how!

Proposition 56.8.3
The field $\mathbb{Q}(\sqrt{-67})$ has class number 1.

Proof. Since $K = \mathbb{Q}(\sqrt{-67})$ has signature $(0,1)$ and discriminant $\Delta_K = -67$ (since $-67 \equiv 1 \pmod 4$) we can compute

$$M_K = \left(\frac{4}{\pi}\right)^1 \cdot \frac{2!}{2^2}\sqrt{67} \approx 5.2.$$

That means we can cut off the cherry tree after (2), (3), (5), since any cherries on these branches will necessarily have norm $\ge M_K$. We now want to factor each of these in $\mathcal{O}_K = \mathbb{Z}[\theta]$, where $\theta = \frac{1+\sqrt{-67}}{2}$ has minimal polynomial $x^2 - x + 17$. But something miraculous happens:

- When we try to reduce $x^2 - x + 17 \pmod 2$, we get an irreducible polynomial $x^2 - x + 1$. By the factoring algorithm (Theorem 55.5.4) this means (2) is prime.
- Similarly, reducing mod 3 gives $x^2 - x + 2$, which is irreducible. This means (3) is prime.
- Finally, for the same reason, (5) is prime.

It's our lucky day; all of the ideals (2), (3), (5) are prime (already principal). To put it another way, each of the three branches has only one (large) cherry on it. That means any time we put together an integral ideal with norm $\le M_K$, it is actually principal. In fact, these guys have norm 4, 9, 25 respectively... so we can't even touch (3) and (5), and the only ideals we can get are (1) and (2) (with norms 1 and 4).

Now we claim that's all. Suppose $\mathfrak{b}$ is an integral ideal such that $\mathrm{N}(\mathfrak{b}) \le M_K$. By the above, either $\mathfrak{b} = (1)$ or $\mathfrak{b} = (2)$, both of which are principal, and hence trivial in Cl_K. So J is trivial in Cl_K too, as needed. □

Let's do a couple more.

Theorem 56.8.4 (Gaussian integers $\mathbb{Z}[i]$ form a UFD)
The field $\mathbb{Q}(i)$ has class number 1.

Proof. This is $\mathcal{O}_K$ where $K = \mathbb{Q}(i)$, so we just want Cl_K to be trivial. We have $M_K = \frac{2}{\pi}\sqrt{4} < 2$. So every class has an integral ideal of norm $\mathfrak{b}$ satisfying

$$\mathrm{N}(\mathfrak{b}) \le \left(\frac{4}{\pi}\right)^1 \cdot \frac{2!}{2^2} \cdot \sqrt{4} = \frac{4}{\pi} < 2.$$

Well, that's silly: we don't have any branches to pick from at all. In other words, we can only have $\mathfrak{b} = (1)$. □

Here's another example of something that still turns out to be unique factorization, but this time our cherry tree will actually have cherries that can be picked.

Proposition 56.8.5 ($\mathbb{Z}[\sqrt{7}]$ is a UFD)
The field $\mathbb{Q}(\sqrt{7})$ has class number 1.

Proof. First we compute the Minkowski bound.

Question 56.8.6. Check that $M_K \approx 2.646$.

So this time, the only branch is (2). Let's factor (2) as usual: the polynomial x^2+7 reduces as $(x-1)(x+1) \pmod 2$, and hence

$$(2) = \left(2, \sqrt{7}-1\right)\left(2, \sqrt{7}+1\right).$$

Oops! We now have two cherries, and they both seem reasonable. But actually, I claim that

$$\left(2, \sqrt{7}-1\right) = \left(3-\sqrt{7}\right).$$

Question 56.8.7. Prove this.

So both the cherries are principal ideals, and as before we conclude that Cl_K is trivial. But note that this time, the prime ideal (2) actually splits; we got lucky that the two cherries were principal but this won't always work. □

How about some nontrivial class groups? First, we use a lemma that will help us with narrowing down the work in our cherry tree.

Lemma 56.8.8 (Ideals divide their norms)
Let $\mathfrak{b}$ be an integral ideal with $\mathrm{N}(\mathfrak{b}) = n$. Then $\mathfrak{b}$ divides the ideal (n).

Proof. By definition, $n = |\mathcal{O}_K/\mathfrak{b}|$. Treating $\mathcal{O}_K/\mathfrak{b}$ as an (additive) abelian group and using Lagrange's theorem, we find

$$0 \equiv \underbrace{\alpha + \dots + \alpha}_{n \text{ times}} = n\alpha \pmod{\mathfrak{b}} \qquad \text{for all } \alpha \in \mathcal{O}_K.$$

Thus $(n) \subseteq \mathfrak{b}$, done. □

Alternatively, if you have read Chapter 59: If the extension $K/\mathbb{Q}$ is Galois, we can actually prove that, analogous to Remark 54.1.9, $\prod_{\sigma \in \mathrm{Gal}(K/\mathbb{Q})} \sigma(\mathfrak{b}) = (n)$, implying the result $id(\mathfrak{b}) \mid (n)$.

Now we can give such an example.

Proposition 56.8.9 (Class group of $\mathbb{Q}(\sqrt{-17})$)
The number field $K = \mathbb{Q}(\sqrt{-17})$ has class group $\mathbb{Z}/4\mathbb{Z}$.

You are not obliged to read the entire proof in detail, as it is somewhat gory. The idea is just that there are some cherries which are not trivial in the class group.

Proof. Since $\Delta_K = -68$, we compute the Minkowski bound

$$M_K = \frac{4}{\pi}\sqrt{17} < 6.$$

Now, it suffices to factor with (2), (3), (5). The minimal polynomial of $\sqrt{-17}$ is x^2+17, so as usual

$$\begin{aligned}(2) &= (2, \sqrt{-17}+1)^2 \\ (3) &= (3, \sqrt{-17}-1)(3, \sqrt{-17}+1) \\ (5) &= (5)\end{aligned}$$

corresponding to the factorizations of x^2+17 modulo each of 2, 3, 5. Set $\mathfrak{p} = (2, \sqrt{-17}+1)$ and $\mathfrak{q}_1 = (3, \sqrt{-17}-1)$, $\mathfrak{q}_2 = (3, \sqrt{-17}+1)$. We can compute

$$\mathrm{N}(\mathfrak{p}) = 2 \quad \text{and} \quad \mathrm{N}(\mathfrak{q}_1) = \mathrm{N}(\mathfrak{q}_2) = 3.$$

In particular, they are not principal. The ideal (5) is out the window; it has norm 25. Hence, the three cherries are $\mathfrak{p}$, $\mathfrak{q}_1$, $\mathfrak{q}_2$.

The possible ways to arrange these cherries into ideals with norm ≤ 5 are

$$\left\{(1), \mathfrak{p}, \mathfrak{q}_1, \mathfrak{q}_2, \mathfrak{p}^2\right\}.$$

However, you can compute

$$\mathfrak{p}^2 = (2)$$

so $\mathfrak{p}^2$ and (1) are in the same class group; that is, they are trivial. In particular, the class group has order at most 4.

From now on, let $[\mathfrak{a}]$ denote the class (member of the class group) that $\mathfrak{a}$ is in. Since $\mathfrak{p}$ isn't principal (so $[\mathfrak{p}] \neq [(1)]$), it follows that $\mathfrak{p}$ has order two. So Lagrange's theorem says that Cl_K has order either 2 or 4.

Now we claim $[\mathfrak{q}_1]^2 \neq [(1)]$, which implies that $\mathfrak{q}_1$ has order greater than 2. If not, $\mathfrak{q}_1^2$ is principal. We know $\mathrm{N}(\mathfrak{q}_1) = 3$, so this can only occur if $\mathfrak{q}_1^2 = (3)$; this would force $\mathfrak{q}_1 = \mathfrak{q}_2$. This is impossible since $\mathfrak{q}_1 + \mathfrak{q}_2 = (1)$.

Thus, $\mathfrak{q}_1$ has even order greater than 2. So it has to have order 4. From this we deduce

$$\mathrm{Cl}_K \cong \mathbb{Z}/4\mathbb{Z}. \qquad \square$$

Remark 56.8.10 — When we did this at Harvard during Math 129, there was a five-minute interruption in which students (jokingly) complained about the difficulty of evaluating $\frac{4}{\pi}\sqrt{17}$. Excerpt:

> "Will we be allowed to bring a small calculator on the exam?" – Student 1
> "What does the size have to do with anything? You could have an Apple Watch" – Professor
> "Just use the fact that $\pi \ge 3$" – me
> "Even [other professor] doesn't know that, how are we supposed to?" – Student 2
> "You have to do this yourself!" – Professor
> "This is an outrage." – Student 1

§56.9 Optional: Proof that $\mathcal{O}_K$ is a free $\mathbb{Z}$-module

We have the suitable tools to prove Theorem 54.2.12 now.

We know $\mathcal{O}_K$ is a ring, so obviously it must be a $\mathbb{Z}$-module. Suppose it is not a free $\mathbb{Z}$-module of degree $n = |K : \mathbb{Q}|$. What could go wrong?

- First, it may happen that it is dense like $\mathbb{Q}$ or the ring extension $\mathbb{Z}[\frac{1}{2}]$ (which makes it not finitely generated and not free).
- Even without that, it may happen that its rank is less than n.

The second possibility is much easier to discard. Let $\alpha_1, \dots, \alpha_n$ be a basis of K. Using Theorem 54.2.6, we have positive integers $d_1, \dots, d_n$ such that $d_1\alpha_1, \dots, d_n\alpha_n \in \mathcal{O}_K$. Since $\alpha_1, \dots, \alpha_n$ are linearly independent, this implies $\operatorname{rank} \mathcal{O}_K \ge n$.

The other direction is harder. We wish to prove $\mathcal{O}_K$ is "discrete" in some sense.

From now on, replace α_i with $d_i\alpha_i$, so they are still a basis of the $\mathbb{Q}$-vector space K, and furthermore they are now in $\mathcal{O}_K$.

Three distinct proofs will be provided.

§56.9.i First proof

We will show that $\mathcal{O}_K$ is contained in some free $\mathbb{Z}$-module of rank n.

Specifically, we will show that $\mathcal{O}_K \subseteq \langle \frac{1}{d}\alpha_1, \dots, \frac{1}{d}\alpha_n \rangle$ for some integer $d \ne 0$.

Let us pretend that we already know $\mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank n. How would we compute d?

Exercise 56.9.1. These are a few naive attempts to compute d; unfortunately, they wouldn't work. Verify that on $\langle 1, 1+2i \rangle \subseteq \mathbb{Z}[i]$.

- Take d to be the first positive integer that belongs to $\langle \alpha_1, \dots, \alpha_n \rangle$. (Attempt inspired by Theorem 55.3.6.)
- Take d to be the product of the norm of $\alpha_1, \dots, \alpha_n$.

Instead, we compute d by using an idea inspired by how we compute the mesh of the lattice. Let $A = \langle \alpha_1, \dots, \alpha_n \rangle$, then it is a lattice and a free $\mathbb{Z}$-module of rank n.

Exercise 56.9.2. Assume you already know $\mathcal{O}_K$ is free of rank n. Show that $|\mathcal{O}_K/A|$ is finite. Conclude that for every $x \in \mathcal{O}_K$, then $|\mathcal{O}_K/A| \cdot x \in A$.

Using the same argument as Problem 57B*, you can prove that the "discriminant" (squared mesh) of the lattice spanned by $\alpha_1, \dots, \alpha_n$ is an integer. Formally, let

$$d \coloneqq \det \begin{bmatrix} \sigma_1(\alpha_1) & \dots & \sigma_n(\alpha_1) \\ \vdots & \ddots & \vdots \\ \sigma_1(\alpha_n) & \dots & \sigma_n(\alpha_n) \end{bmatrix}^2 .$$

Exercise 56.9.3. Convince yourself (at least when all embeddings are real) that the ratio of the meshes of $\mathcal{O}_K$ and A is exactly the size of the quotient abelian group $\mathcal{O}_K/A$, that is $\left|\frac{d}{\Delta_K}\right| = |\mathcal{O}_K/A|^2$. Conclude that for every $x \in \mathcal{O}_K$, then $d \cdot x \in A$.

Which implies $\mathcal{O}_K \subseteq \langle \frac{1}{d}\alpha_1, \dots, \frac{1}{d}\alpha_n \rangle$, which is another free $\mathbb{Z}$-module of rank n.

However, the argument above is circular since it assumes Δ_K exists (it can only serve as a motivation for where the value d comes from). The actual proof is the following.

Similar to Problem 57B*, we define

$$d := \det[\operatorname{Tr}_{K/\mathbb{Q}}(\alpha_i \alpha_j)]_{i,j}.$$

Then, because $\{\alpha_i\}_i$ spans K as a $\mathbb{Q}$-vector space, there is some $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{Q}^{\oplus n}$ such that

$$\begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x \end{pmatrix}.$$

Which means

$$\begin{bmatrix} \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1\alpha_1) & \cdots & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1\alpha_n) \\ \vdots & \ddots & \vdots \\ \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n\alpha_1) & \cdots & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n\alpha_n) \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1 x) \\ \vdots \\ \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n x) \end{pmatrix}.$$

Or, in short,

$$(\text{some matrix} \in \mathrm{GL}_n(\mathbb{Z}) \text{ with determinant } d) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (\text{some vector} \in \mathbb{Z}^{\oplus n}).$$

Question 56.9.4. Finish the proof that $d \cdot x \in A$. (Cramer's rule. Or take the adjoint.)

Finally, we have that $\mathcal{O}_K$ is squeezed between two free $\mathbb{Z}$-modules of rank n, so it is also free of rank n.[2]

Question 56.9.5. Finish this. (Theorem 18.1.5 can be used here.)

§56.9.ii Second proof

This time around, instead of dividing by d, we instead take the *dual lattice*.

What is the dual lattice, and why would we think about using it? One way to motivate this proof is to look at the inverse of an ideal: if $(1) \mid \mathfrak{a}$ (i.e. $\mathfrak{a}$ is an integral ideal), then $\mathfrak{a}^{-1} \mid (1)$.

Here, $\mathfrak{a}^{-1} := \{x \in K \mid xa \in \mathcal{O}_K \text{ for all } a \in \mathfrak{a}\}$.

We can think of doing something similar, considering

$$\{x \in K \mid xa \in \mathcal{O}_K \text{ for all } a \in \langle \alpha_1, \dots, \alpha_n \rangle\}.$$

This set is actually a lattice of rank n, but this won't work to prove the argument! We're trying to prove $\mathcal{O}_K$ is "discrete" in the first place, if $\mathcal{O}_K = K$ then the set above would equal K as well.

Instead, we must rely on what we already know — the discreteness of $\mathbb{Z}$. Define

$$S := \{x \in K \mid \operatorname{Tr}_{K/\mathbb{Q}}(xa) \in \mathbb{Z} \text{ for all } a \in \langle \alpha_1, \dots, \alpha_n \rangle\}.$$

This set is a bit larger than the previous set.

[2] We did something similar in Remark 55.3.7.

Question 56.9.6. Verify that this set is a superset of the previous one. Then conclude that $\mathcal{O}_K \subseteq S$.

Exercise 56.9.7. Consider $\langle 1, 2i \rangle \subseteq \mathbb{Z}[i]$. What would the set S be? (Recall that the trace in $\mathbb{C}$ is just twice the real part.)

S is still a lattice of rank n — and this time, we can actually prove it! Since we already know $\mathbb{Z}$ is discrete.

Now, this is a purely algebraic problem, we will only need to use knowledge of vector space here. Each element $x \in K$ can be written as a vector

$$F(x) := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{Q}^{\oplus n}$$

if we use the basis $\{\alpha_1, \dots, \alpha_n\}$.

Exercise 56.9.8. Show that $(x, y) \mapsto \operatorname{Tr}_{K/\mathbb{Q}}(x \cdot y)$ can then be written as an *invertible* matrix in $\operatorname{GL}_n(\mathbb{Q})$ — specifically, there exists $M \in \operatorname{GL}_n(\mathbb{Q})$ such that $\operatorname{Tr}_{K/\mathbb{Q}}(x \cdot y) = F(x)^\top M F(y)$, where $F(x)^\top$ is the transpose of $F(x)$.

Which makes $(x, y) \mapsto \operatorname{Tr}_{K/\mathbb{Q}}(x \cdot y)$ *almost* an inner product (see Chapter 13), except that it is not positive definite (for example, in $\mathbb{C}$, we have $\operatorname{Tr}((1+i)^2) = 0$). But having the matrix invertible suffices to do the following:

Exercise 56.9.9. Finish the proof. (Hint: Consider the matrix $\begin{bmatrix} F(\alpha_1) & \cdots & F(\alpha_n) \end{bmatrix} \in \operatorname{GL}_n(\mathbb{Q})$. What is the condition on $F(x)$ such that $x \in S$?)

§56.9.iii Third proof

Recall that we wish to prove $\mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank n. To do that, we suppose it cannot be generated by n elements, then show that $\mathcal{O}_K$ is not discrete, and this causes trouble because we know the norm is continuous.

Exercise 56.9.10. Consider $K \cong \mathbb{Q}^{\oplus n}$, this gives a topology on K. Verify that $x \mapsto |\operatorname{N}(x)|$ is indeed continuous.

We have that for every $x \in \mathcal{O}_K$, then $|\operatorname{N}(x)| \in \mathbb{Z}$.

Exercise 56.9.11. Conclude that there is a ball $B(0, r)$ in the topology above that contains no element of $\mathcal{O}_K$, except 0.

Now, what goes wrong if $\mathcal{O}_K$ cannot be generated by n elements? Things go wrong quickly:

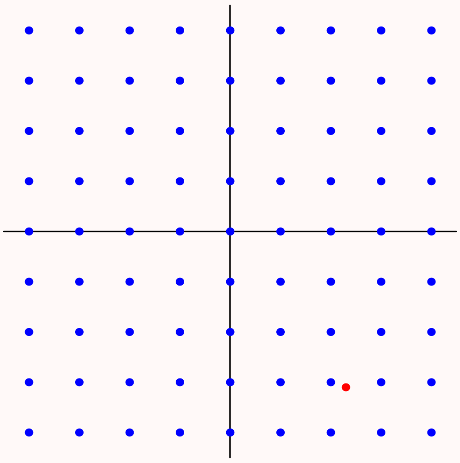
Exercise 56.9.12. Suppose $\langle \alpha_1, \dots, \alpha_n \rangle$ generates K as a $\mathbb{Q}$-vector space. Let $x \in \mathcal{O}_K$ such that x is not a $\mathbb{Z}$-linear combination of $\{\alpha_1, \dots, \alpha_n\}$. Show that there is $z_1, \dots, z_n \in [-\frac{1}{2}, \frac{1}{2}]$ not all zero, and z a $\mathbb{Z}$-linear combination of $\{\alpha_1, \dots, \alpha_n\}$ such that $x + z = \sum_{i=1}^n z_i \alpha_i$.

Exercise 56.9.13. With notation as above, suppose $z_1 \neq 0$. Show that if we replace α_1 with z_1, then the new set $\{z_1, \alpha_2, \dots, \alpha_n\}$ still spans K as a $\mathbb{Q}$-vector space; furthermore, the mesh of the lattice gets decreased by *at least a half.*

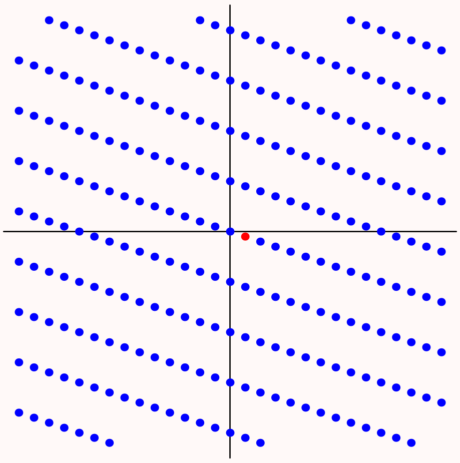
Example 56.9.14

Suppose $A \subseteq \mathbb{C}$ is a lattice. We know $1 \in A$ and $i \in A$, so $\mathbb{Z}[i] \subseteq A$.

Assume we know in addition that $2.3 - 3.1i \in A$.



Because A is closed under addition, we know that $(2.3 - 3.1i) + (-2 + 3i) = 0.3 - 0.1i \in A$. We replace 1 in the basis with $0.3 - 0.1i$. Then, the lattice $\langle i, 0.3 - 0.1i \rangle \subseteq A$ has smaller mesh than $\langle 1, i \rangle$ as expected.

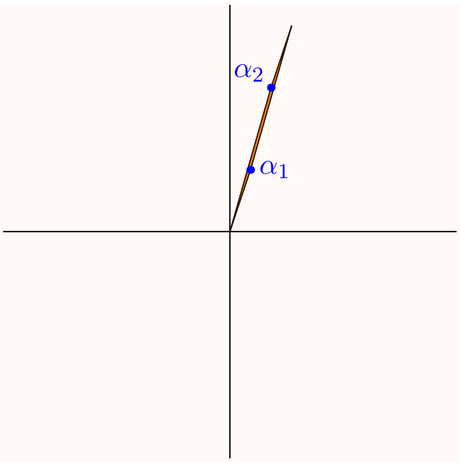


Since we can do this indefinitely ($\mathcal{O}_K$ never gets spanned), the mesh decreases to 0 rapidly.

So, are we done? Since the mesh goes to 0, maybe the smallest distance of some α_i also go to 0? Almost, but not quite:

Example 56.9.15

Consider $\alpha_1 = 1 + 3i$ and $\alpha_2 = 2 + 7i$.



The mesh of the lattice spanned by α_1 and α_2 is 1, however, neither α_1 nor α_2 are particularly close to the origin.

We need to apply Minkowski bound here again: suppose the mesh is d, then the cube centered at origin with volume $2^n d$ contains a nonzero lattice point.[3] This point's distance cannot be more than $\sqrt{n} \cdot \sqrt[n]{d}$ away from the origin.

Remark 56.9.16 — Speaking of which, the LLL lattice basis reduction algorithm can be used to find *in practice* a point on the lattice that is *close enough* to the origin.

For d small enough, $\sqrt{n} \cdot \sqrt[n]{d} < r$ where r is the ball of no lattice point we have shown above, which gives a contradiction. So we're done!

§56.10 A few harder problems to think about

Problem 56A. Show that $K = \mathbb{Q}(\sqrt{-163})$ has trivial class group, and hence $\mathcal{O}_K = \mathbb{Z}\left[\frac{1+\sqrt{-163}}{2}\right]$ is a UFD.[4]

Problem 56B. Determine the class group of $\mathbb{Q}(\sqrt{-31})$.

Problem 56C (China TST 1998)**.** Let n be a positive integer. A polygon in the plane (not necessarily convex) has area greater than n. Prove that one can translate it so that it contains at least $n+1$ lattice points.

Problem 56D (Lemma 56.3.6)**.** Consider the composition of the embeddings $K \hookrightarrow \mathbb{R}^{r_1} \times \mathbb{C}^{r_2} \xrightarrow{\sim} \mathbb{R}^n$. Show that the image of $\mathcal{O}_K \subseteq K$ has mesh equal to

$$\frac{1}{2^{r_2}}\sqrt{|\Delta_K|}.$$

Problem 56E. Let $p \equiv 1 \pmod 4$ be a prime. Show that there are unique integers $a > b > 0$ such that $a^2 + b^2 = p$.

Problem 56F (Korea national olympiad 2014)**.** Let p be an odd prime and k a positive integer such that $p \mid k^2 + 5$. Prove that there exist positive integers m, n such that $p^2 = m^2 + 5n^2$.

[3] Using a sphere would of course make for a better bound, but its volume is a bit harder to calculate.

[4] In fact, $n = 163$ is the largest number for which $\mathbb{Q}(\sqrt{-n})$ has trivial class group. The complete list is $1, 2, 3, 7, 11, 19, 43, 67, 163$, the **Heegner numbers**. You might notice Euler's prime-generating polynomial $t^2 + t + 41$ when doing the above problem. Not a coincidence!

57 More properties of the discriminant

I'll remind you that the discriminant of a number field K is given by

$$\Delta_K := \det \begin{bmatrix} \sigma_1(\alpha_1) & \dots & \sigma_n(\alpha_1) \\ \vdots & \ddots & \vdots \\ \sigma_1(\alpha_n) & \dots & \sigma_n(\alpha_n) \end{bmatrix}^2$$

where α_1, ..., α_n is a $\mathbb{Z}$-basis for K, and the σ_i are the n embeddings of K into $\mathbb{C}$.

Several examples, properties, and equivalent definitions follow.

§57.1 A few harder problems to think about

Problem 57A* (Discriminant of cyclotomic field)**.** Let p be an odd rational prime and ζ_p a primitive pth root of unity. Let $K = \mathbb{Q}(\zeta_p)$. Show that

$$\Delta_K = (-1)^{\frac{p-1}{2}} p^{p-2}.$$

🌶 **Problem 57B*** (Trace representation of Δ_K)**.** Let α_1, ..., α_n be a basis for $\mathcal{O}_K$. Prove that

$$\Delta_K = \det \begin{bmatrix} \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1^2) & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1\alpha_2) & \dots & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_1\alpha_n) \\ \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_2\alpha_1) & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_2^2) & \dots & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_2\alpha_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n\alpha_1) & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n\alpha_2) & \dots & \operatorname{Tr}_{K/\mathbb{Q}}(\alpha_n\alpha_n) \end{bmatrix}.$$

In particular, Δ_K is an integer.

Problem 57C* (Root representation of Δ_K)**.** The **discriminant** of a quadratic polynomial $Ax^2 + Bx + C$ is defined as $B^2 - 4AC$. More generally, the polynomial discriminant of a polynomial $f \in \mathbb{Z}[x]$ of degree n is

$$\Delta(f) := c^{2n-2} \prod_{1 \le i < j \le n} (z_i - z_j)^2$$

where $z_1, \dots, z_n$ are the roots of f, and c is the leading coefficient of f.

Suppose K is monogenic with $\mathcal{O}_K = \mathbb{Z}[\theta]$. Let f denote the minimal polynomial of θ (hence monic). Show that

$$\Delta_K = \Delta(f).$$

Problem 57D. Show that if $K \neq \mathbb{Q}$ is a number field then $|\Delta_K| > 1$.

Problem 57E (Brill's theorem)**.** For a number field K with signature (r_1, r_2), show that $\Delta_K > 0$ if and only if r_2 is even.

🌶🌶🌶 **Problem 57F** (Stickelberger theorem)**.** Let K be a number field. Prove that

$$\Delta_K \equiv 0 \text{ or } 1 \pmod 4.$$

58 Bonus: Let's solve Pell's equation!

This is an optional aside, and can be safely ignored. (On the other hand, it's pretty short.)

§58.1 Units

Prototypical example for this section: ± 1, *roots of unity*, $3-2\sqrt{2}$ *and its powers.*

Recall according to Problem 54A* that $\alpha \in \mathcal{O}_K$ is invertible if and only if

$$\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \pm 1.$$

We let $\mathcal{O}_K^\times$ denote the set of units of $\mathcal{O}_K$.

> **Question 58.1.1.** Show that $\mathcal{O}_K^\times$ is a group under multiplication. Hence we name it the **unit group** of $\mathcal{O}_K$.

What are some examples of units?

Example 58.1.2 (Examples of units in a number field)

1. ± 1 are certainly units, present in any number field.
2. If $\mathcal{O}_K$ contains a root of unity ω (i.e. $\omega^n = 1$), then ω is a unit. (In fact, ± 1 are special cases of this.)
3. Of course, not all units of $\mathcal{O}_K$ are roots of unity. For example, if $\mathcal{O}_K = \mathbb{Z}[\sqrt{3}]$ (from $K = \mathbb{Q}(\sqrt{3})$) then the number $2+\sqrt{3}$ is a unit, as its norm is

   $$\mathrm{N}_{K/\mathbb{Q}}(2+\sqrt{3}) = 2^2 - 3 \cdot 1^2 = 1.$$

   Alternatively, just note that the inverse $2-\sqrt{3} \in \mathcal{O}_K$ as well:

   $$\left(2-\sqrt{3}\right)\left(2+\sqrt{3}\right) = 1.$$

   Either way, $2-\sqrt{3}$ is a unit.
4. Given any unit $u \in \mathcal{O}_K^\times$, all its powers are also units. So for example, $(3-2\sqrt{2})^n$ is always a unit of $\mathbb{Z}[\sqrt{2}]$, for any n. If u is not a root of unity, then this generates infinitely many new units in $\mathcal{O}_K^\times$.

> **Question 58.1.3.** Verify the claims above that
>
> (a) Roots of unity are units, and
>
> (b) Powers of units are units.
>
> One can either proceed from the definition or use the characterization $\mathrm{N}_{K/\mathbb{Q}}(\alpha) = \pm 1$. If one definition seems more natural to you, use the other.

§58.2 Dirichlet's unit theorem

Prototypical example for this section: The units of $\mathbb{Z}[\sqrt{3}]$ *are* $\pm(2+\sqrt{3})^n$.

Definition 58.2.1. Let $\mu(\mathcal{O}_K)$ denote the set of roots of unity contained in a number field K (equivalently, in $\mathcal{O}_K$).

Example 58.2.2 (Examples of $\mu(\mathcal{O}_K)$)

(a) If $K = \mathbb{Q}(i)$, then $\mathcal{O}_K = \mathbb{Z}[i]$. So

$$\mu(\mathcal{O}_K) = \{\pm 1, \pm i\} \quad \text{where } K = \mathbb{Q}(i).$$

(b) If $K = \mathbb{Q}(\sqrt{3})$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{3}]$. So

$$\mu(\mathcal{O}_K) = \{\pm 1\} \quad \text{where } K = \mathbb{Q}(\sqrt{3}).$$

(c) If $K = \mathbb{Q}(\sqrt{-3})$, then $\mathcal{O}_K = \mathbb{Z}[\frac{1}{2}(1+\sqrt{-3})]$. So

$$\mu(\mathcal{O}_K) = \left\{\pm 1, \frac{\pm 1 \pm \sqrt{-3}}{2}\right\} \quad \text{where } K = \mathbb{Q}(\sqrt{-3})$$

where the $\pm$'s in the second term need not depend on each other; in other words $\mu(\mathcal{O}_K) = \{z \mid z^6 = 1\}$.

Exercise 58.2.3. Show that we always have that $\mu(\mathcal{O}_K)$ comprises the roots to $x^n - 1$ for some integer n. (First, show it is a finite group under multiplication.)

We now quote, without proof, the so-called Dirichlet's unit theorem, which gives us a much more complete picture of what the units in $\mathcal{O}_K$ are. Legend says that Dirichlet found the proof of this theorem during an Easter concert in the Sistine Chapel.

Theorem 58.2.4 (Dirichlet's unit theorem)

Let K be a number field with signature (r_1, r_2) and set

$$s = r_1 + r_2 - 1.$$

Then there exist units $u_1, \dots, u_s$ such that every unit $\alpha \in \mathcal{O}_K^\times$ can be written *uniquely* in the form

$$\alpha = \omega \cdot u_1^{n_1} \dots u_s^{n_s}$$

for $\omega \in \mu(\mathcal{O}_K)$ is a root of unity, and $n_1, \dots, n_s \in \mathbb{Z}$.

More succinctly:

We have $\mathcal{O}_K^\times \cong \mathbb{Z}^{r_1+r_2-1} \times \mu(\mathcal{O}_K)$.

A choice of $u_1, \dots, u_s$ is called a choice of **fundamental units**.

Here are some example applications.

Example 58.2.5 (Some unit groups)

(a) Let $K = \mathbb{Q}(i)$ with signature $(0,1)$. Then we obtain $s = 0$, so Dirichlet's Unit theorem says that there are no units other than the roots of unity. Thus

$$\mathcal{O}_K^\times = \{\pm 1, \pm i\} \quad \text{where } K = \mathbb{Q}(i).$$

This is not surprising, since $a + bi \in \mathbb{Z}[i]$ is a unit if and only if $a^2 + b^2 = 1$.

(b) Let $K = \mathbb{Q}(\sqrt{3})$, which has signature $(2,0)$. Then $s = 1$, so we expect exactly one fundamental unit. A fundamental unit is $2 + \sqrt{3}$ (or $2 - \sqrt{3}$, its inverse) with norm 1, and so we find

$$\mathcal{O}_K^\times = \left\{ \pm (2 + \sqrt{3})^n \mid n \in \mathbb{Z} \right\}.$$

(c) Let $K = \mathbb{Q}(\sqrt[3]{2})$ with signature $(1,1)$. Then $s = 1$, so we expect exactly one fundamental unit. The choice $1 + \sqrt[3]{2} + \sqrt[3]{4}$. So

$$\mathcal{O}_K^\times = \left\{ \pm \left(1 + \sqrt[3]{2} + \sqrt[3]{4}\right)^n \mid n \in \mathbb{Z} \right\}.$$

I haven't actually shown you that these are fundamental units, and indeed computing fundamental units is in general hard.

§58.3 Finding fundamental units

Here is a table with some fundamental units.

d	Unit
$d = 2$	$1 + \sqrt{2}$
$d = 3$	$2 + \sqrt{3}$
$d = 5$	$\frac{1}{2}(1 + \sqrt{5})$
$d = 6$	$5 + 2\sqrt{6}$
$d = 7$	$8 + 3\sqrt{7}$
$d = 10$	$3 + \sqrt{10}$
$d = 11$	$10 + 3\sqrt{11}$

In general, determining fundamental units is computationally hard.

However, once I tell you what the fundamental unit is, it's not too bad (at least in the case $s = 1$) to verify it. For example, suppose we want to show that $10 + 3\sqrt{11}$ is a fundamental unit of $K = \mathbb{Q}(\sqrt{11})$, which has ring of integers $\mathbb{Z}[\sqrt{11}]$. If not, then for some $n > 1$, we would have to have

$$10 + 3\sqrt{11} = \pm \left(x + y\sqrt{11}\right)^n.$$

For this to happen, at the very least we would need $|y| < 3$. We would also have $x^2 - 11y^2 = \pm 1$. So one can just verify (using $y = 1, 2$) that this fails.

The point is that: Since $(10, 3)$ is the *smallest* (in the sense of $|y|$) integer solution to $x^2 - 11y^2 = \pm 1$, it must be the fundamental unit. This holds more generally, although in the case that $d \equiv 1 \pmod 4$ a modification must be made as x, y might be half-integers (like $\frac{1}{2}(1 + \sqrt{5})$).

Theorem 58.3.1 (Fundamental units of Pell's equations)
Assume d is a squarefree integer.

(a) If $d \equiv 2, 3 \pmod 4$, and (x, y) is a minimal integer solution to $x^2 - dy^2 = \pm 1$, then $x + y\sqrt{d}$ is a fundamental unit.

(b) If $d \equiv 1 \pmod 4$, and (x, y) is a minimal *half-integer* solution to $x^2 - dy^2 = \pm 1$, then $x + y\sqrt{d}$ is a fundamental unit. (Equivalently, the minimal integer solution to $a^2 - db^2 = \pm 4$ gives $\frac{1}{2}(a + b\sqrt{d})$.)

(Any reasonable definition of "minimal" will work, such as sorting by $|y|$.)

§58.4 Pell's equation

This class of results completely eradicates Pell's Equation. After all, solving

$$a^2 - d \cdot b^2 = \pm 1$$

amounts to finding elements of $\mathbb{Z}[\sqrt{d}]$ with norm ± 1. It's a bit weirder in the $d \equiv 1 \pmod 4$ case, since in that case $K = \mathbb{Q}(\sqrt{d})$ gives $\mathcal{O}_K = \mathbb{Z}[\frac{1}{2}(1 + \sqrt{d})]$, and so the fundamental unit may not actually be a solution. (For example, when $d = 5$, we get the solution $(\frac{1}{2}, \frac{1}{2})$.) Nonetheless, all *integer* solutions are eventually generated.

To make this all concrete, here's a simple example.

Example 58.4.1 ($x^2 - 5y^2 = \pm 1$)
Set $K = \mathbb{Q}(\sqrt{5})$, so $\mathcal{O}_K = \mathbb{Z}[\frac{1}{2}(1+\sqrt{5})]$. By Dirichlet's unit theorem, $\mathcal{O}_K^\times$ is generated by a single element u. The choice

$$u = \frac{1}{2} + \frac{1}{2}\sqrt{5}$$

serves as a fundamental unit, as there are no smaller integer solutions to $a^2 - 5b^2 = \pm 4$.

The first several powers of u are

n	u^n	Norm
-2	$\frac{1}{2}(3 - \sqrt{5})$	1
-1	$\frac{1}{2}(1 - \sqrt{5})$	-1
0	1	1
1	$\frac{1}{2}(1 + \sqrt{5})$	-1
2	$\frac{1}{2}(3 + \sqrt{5})$	1
3	$2 + \sqrt{5}$	-1
4	$\frac{1}{2}(7 + 3\sqrt{5})$	1
5	$\frac{1}{2}(11 + 5\sqrt{5})$	-1
6	$9 + 4\sqrt{5}$	1

One can see that the first integer solution is $(2, 1)$, which gives -1. The first solution with $+1$ is $(9, 4)$. Continuing the pattern, we find that every third power of u gives an integer solution (see also Problem 58B), with the odd ones giving a solution to

$x^2 - 5y^2 = -1$ and the even ones a solution to $x^2 - 5y^2 = +1$. All solutions are generated this way, up to ± signs (by considering $\pm u^{\pm n}$).

§58.5 A few harder problems to think about

Problem 58A (Fictitious account of the battle of Hastings)**.** Determine the number of soldiers in the following battle:

> The men of Harold stood well together, as their wont was, and formed thirteen squares, with a like number of men in every square thereof, and woe to the hardy Norman who ventured to enter their redoubts; for a single blow of Saxon war-hatched would break his lance and cut through his coat of mail . . . when Harold threw himself into the fray the Saxons were one might square of men, shouting the battle-cries, "Ut!", "Olicrosse!", "Godemite!"

Problem 58B. Let $d > 0$ be a squarefree integer, and let u denote the fundamental unit of $\mathbb{Q}(\sqrt{d})$. Show that either $u \in \mathbb{Z}[\sqrt{d}]$, or $u^n \in \mathbb{Z}[\sqrt{d}] \iff 3 \mid n$.

Problem 58C. Show that there are no integer solutions to

$$x^2 - 34y^2 = -1$$

despite the fact that -1 is a quadratic residue mod 34.