Algebraic Number Theory

Term 1, 2019–20

Martin Orr

1. Introduction

What is algebraic number theory?

(1) *Theory of algebraic numbers*, that is, roots of polynomials with rational coefficients. This is also covered in Galois theory from a different perspective – we focus on "number theoretic" aspects like factorisations and primes.
(2) *Number theory using algebra* – studying prime numbers, integer solutions of equations etc using concepts like rings, ideals, polynomials.

These both happen together. For example, the question: What are all the integer solutions of

$$y^2 = x^3 + 6?$$

A classic number theory question just involving integers. By the end of the course, we will have a method to answer this using ideals in a ring ((2) above). To do this, we will have to use the algebraic number $\sqrt[3]{6}$ ((1) above) even though the question is only about integers.

A much harder example (which we won't get to in this course!) is Fermat's last theorem: the same question for the equation

$$x^n + y^n = z^n.$$

Gaussian integers.

The simplest example of what we will study in Algebraic Number Theory is the Gaussian integers

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

If you did Introduction to Number Theory, you will have seen these before, but whether you have or not you will encounter them often in this course – where they will be just one case of a broader theory.

The key facts about the Gaussian integers:

(1) Every Gaussian integer can be uniquely factorised as a product of irreducible Gaussian integers.
(2) We can describe the irreducible Gaussian integers in terms of the ordinary prime numbers (depending on whether a prime is 1 or 3 mod 4).

Gaussian integers are an example of algebraic numbers. We will generalise them to other rings of algebraic numbers. These are not always as nice as the Gaussian integers: for example, in the ring

$$\mathbb{Z}[\sqrt{-6}] = \{a + b\sqrt{-6} : a, b \in \mathbb{Z}\},$$

factorisation into irreducibles is non-unique (you might have seen this example before).

We will define factorisation of ideals, instead of elements of the ring, and see that this restores the uniqueness of prime factorisations. We will also define the "class group" of an algebraic number ring, which measures how far it is away from having unique factorisation of elements.

Practical information about the course.

The course involves a lot of down-to-earth calculation with examples e.g. determining how a prime factorises in a number field or computing the class group of a number field. There is also a lot of theory underpinning these calculations. Lectures will focus on the theory; example sheets and support classes on the examples. The exam will have both theoretical, proof-based questions and example-based questions.

Assignments – four pieces, best 3 of 4 will count (15% of module mark)
Deadlines: Monday 12 noon in weeks 4, 6, 8, 10

Example sheets, lecture capture and notes will be available on Moodle. (Notes usually 1-2 days after each lecture.)

My email address: `martin.orr@warwick.ac.uk`

Office hours: will be announced soon

Support classes should be organised soon.

Related courses.

Algebra 2 – the most important prerequisite. Rings, fields, ideals and factorisation of polynomials will be used throughout this course. We will also need quotient rings and the First Isomorphism Theorem for rings. You may like to revise this (we will not use the groups part of Algebra 2).

Algebra 1 – A little of Algebra 1 will be used in one part of the course (Smith Normal Form).

Introduction to Number Theory – not strictly a prerequisite, but it will provide very helpful background.

Galois Theory – The most closely related course in Year 3 (both study algebraic numbers). There is some overlap in the first 2 weeks, with various definitions and lemmas related to field extensions.

In Algebraic Number Theory, I will state all these definitions and lemmas, maybe giving brief examples, but omit the proofs. Galois Theory will (I expect!) prove everything and look at more examples – so Galois Theory might take a bit longer to cover this material. These proofs will not be examinable in this course.

2. Field extensions

In primary school, you learned about $\mathbb{Z}$ (a ring) before you learned about $\mathbb{Q}$ (a field). However for algebraic number theory it is easier to go in the opposite order: first we study fields, specifically a class called "number fields", then we study "rings of integers" inside them.

Definition of field extensions.

Our main object of study in this module will be number fields. A number field is defined as a finite extension of the field $\mathbb{Q}$. Let's unpack this definition.

Definition. Let K and L be fields. If K is a subfield of L, we say that L/K is a **field extension** (often, we will just call it an **extension**).

L/K here is just a piece of notation representing a pair of fields K, L. It does not mean quotient!

e.g. $\mathbb{C}/\mathbb{R}$, $\mathbb{R}/\mathbb{Q}$, $\mathbb{Q}(i)/\mathbb{Q}$, $\mathbb{C}/\mathbb{Q}(i)$ but not $\mathbb{R}/\mathbb{Q}(i)$
where $\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$.

This notation $\mathbb{Q}(i)$ can be generalised: let L/K be a field extension and let $\alpha_1, \ldots, \alpha_n$ be elements of L. We write $K(\alpha_1, \ldots, \alpha_m)$ to mean the "smallest subfield of L containing K and all of $\alpha_1, \ldots, \alpha_m$." That is:

$$K(\alpha_1, \ldots, \alpha_m) = \left\{ \frac{f(\alpha_1, \ldots, \alpha_m)}{g(\alpha_1, \ldots, \alpha_m)} : f, g \in K[X_1, \ldots, X_m], g(\alpha_1, \ldots, \alpha_m) \neq 0 \right\}.$$

This is called the extension of K **generated by** $\alpha_1, \ldots, \alpha_m$ or the extension of K obtained by **adjoining** $\alpha_1, \ldots, \alpha_m$.

e.g. $\mathbb{C} = \mathbb{R}(i)$

If d is a non-square rational number, then $\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}$.

This is a field because

$$(a + b\sqrt{d})(c + e\sqrt{d}) = (ac + bed) + (ae + bc)\sqrt{d}$$

and

$$(a + b\sqrt{d})^{-1} = (a - b\sqrt{d})/(a^2 - db^2)$$

(the denominator is non-zero because d is not the square of a rational number).

But! Is $\mathbb{Q}(\sqrt[3]{d}) = \{a + b\sqrt[3]{d} : a, b \in \mathbb{Q}\}$?

No: this set is not closed under multiplication so it is not a field. We will soon show that

$$\mathbb{Q}(\sqrt[3]{d}) = \{a + b\sqrt[3]{d} + cd^{2/3} : a, b, c \in \mathbb{Q}\}.$$

Algebraic elements.

Definition. Let L/K be a field extension and $\alpha \in L$. We say that α is **algebraic over** K if there exists a non-zero polynomial $f(X) \in K[X]$ such that $f(\alpha) = 0$.

e.g. $i \in \mathbb{C}$, $\sqrt[4]{7} \in \mathbb{R}$ are algebraic over $\mathbb{Q}$
π is not algebraic over $\mathbb{Q}$
πi is algebraic over $\mathbb{R}$ but not algebraic over $\mathbb{Q}$

Lemma 1. *Let α be algebraic over K.*

*(i) There exists a unique monic polynomial $\mu_{K,\alpha}(X) \in K[X]$ of smallest degree such that $\mu_{K,\alpha}(\alpha) = 0$. (**monic** means that the leading coefficient is 1)*
(ii) $\mu_{K,\alpha}$ is irreducible in $K[X]$.
(iii) $\mu_{K,\alpha}$ is the unique monic irreducible polynomial in $K[X]$ which vanishes at α.
(iv) If $f(X) \in K[X]$ satisfies $f(\alpha) = 0$, then $\mu_{K,\alpha}$ divides f.

This will be proved in Galois Theory. The proof relies on the fact that $K[X]$ is a PID (principal ideal domain), so the ideal $\{f \in K[X] : f(\alpha) = 0\}$ has a generator.

Definition. The polynomial $\mu_{K,\alpha}$ from Lemma 1 is called the **minimal polynomial** of α over K. We will write μ_α instead of $\mu_{K,\alpha}$ if the base field K is clear from the context.

But! K matters for determining the minimal polynomial! e.g. $\alpha = i + \sqrt{2} \in \mathbb{C}$.

- Over $K = \mathbb{C}$: the minimal polynomial is $\mu_{\mathbb{C},\alpha}(X) = X - \alpha$.
- Over $K = \mathbb{R}$: $\alpha \notin \mathbb{R}$, so the minimal polynomial has degree > 1. To find the minimal polynomial, we try to arrange things so that both sides square to expressions with real coefficients:

  $$\alpha - \sqrt{2} = i \quad \text{so} \quad (\alpha - \sqrt{2})^2 = -1 \quad \text{so} \quad \alpha^2 - 2\sqrt{2}\alpha + 3 = 0.$$

  Since the minimal polynomial has degree > 1 and we have just found a polynomial of degree 2 in $\mathbb{R}[X]$ which vanishes at α, it must be the minimal polynomial. That is, $\mu_{\mathbb{R},\alpha}(X) = X^2 - 2\sqrt{2}X + 3$.
- Over $K = \mathbb{Q}$: Repeat the same process of rearranging so that both sides square to something with rational coefficients.

  $$\alpha^2 + 3 = 2\sqrt{2}\alpha \quad \text{so} \quad (\alpha^2 + 3)^2 = 8\alpha^2 \quad \text{so} \quad \alpha^4 - 2\alpha^2 + 9 = 0.$$

  One can check that $X^4 - 2X^2 + 9 \in \mathbb{Q}[X]$ is irreducible over $\mathbb{Q}$, so this is $\mu_{\mathbb{Q},\alpha}$ by Lemma 1(iii). Checking the irreducibility by hand is tedious; we will soon give a quicker proof.

Degree.

Definition. Let $\alpha \in L$ be algebraic over K. The **degree of α over K** is the degree of the polynomial $\mu_{K,\alpha}$.

e.g. $i + \sqrt{2}$ has degree 1 over $\mathbb{C}$, 2 over $\mathbb{R}$, 4 over $\mathbb{Q}$.

Definition. If L/K is a field extension, then L is a K-vector space. The **degree** of L/K, written $[L : K]$, is the dimension of L as a K-vector space.

e.g. $[\mathbb{Q}(\sqrt{d}) : \mathbb{Q}] = 2$, $[\mathbb{C} : \mathbb{R}] = 2$.
$\mathbb{Q}(\pi)/\mathbb{Q}$ has infinite degree, even though it is generated by the finite set $\{\pi\}$, because $1, \pi, \pi^2, \ldots$ are $\mathbb{Q}$-linearly independent.

Definition. L/K is a **finite extension** if its degree is finite.

3. Algebraic and finite extensions

Algebraic extensions.

Definition. An extension L/K is **algebraic** if every $\alpha \in L$ is algebraic over K.

e.g. $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$ is algebraic since $a + b\sqrt{d}$ is a root of $(X-a)^2 - b^2 d \in \mathbb{Q}[X]$. $\mathbb{R}/\mathbb{Q}$ is not algebraic since $\pi \in \mathbb{R}$.

Algebraic extensions are related to finite extensions, which we defined last time.

Lemma 2. *If L/K is a finite extension, then it is an algebraic extension.*

(Proved in Galois Theory)
The converse is false: the set

$$\{\alpha \in \mathbb{C} : \alpha \text{ is algebraic over } \mathbb{Q}\}$$

is an algebraic extension of $\mathbb{Q}$, but not a finite extension of $\mathbb{Q}$. (Using this example now is cheating: we haven't proved that this set is a field. We will do that later in the lecture.)

Constructing finite extensions.

Now we give a couple of ways of constructing finite extensions.

Theorem 3. *Let α be algebraic over K, with minimal polynomial $\mu_\alpha \in K[X]$. Let $n = \deg(\mu_\alpha)$. Then:*

(i) $K(\alpha)$ has K-basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Hence $K(\alpha)/K$ is a finite extension and $[K(\alpha) : K] = n$.

(ii) $K(\alpha)$ is isomorphic as a ring to $K[X]/\langle \mu_\alpha \rangle$. More precisely, the following is a well defined isomorphism $K[X]/\langle \mu_\alpha \rangle \to K(\alpha)$:

$$f(X) + \langle \mu_\alpha \rangle \mapsto f(\alpha).$$

$\langle \mu_\alpha \rangle$ is the notation we will use in this module for "the ideal generated by μ_α."

This will be proved in Galois theory. For (i), the hardest part is showing that K-vector space spanned by $1, \alpha, \ldots, \alpha^{n-1}$ is a field. For (ii), we use the First Isomorphism Theorem for rings.

We have seen that not every algebraic extension is finite, but this theorem at least shows that every algebraic element is contained in some finite extension.

Theorem 3 tells us about extensions generated by a single element. We can build up more complicated extensions such as $K(\alpha_1, \ldots, \alpha_n)$ by adjoining elements successively, one at a time: $L = K(\alpha_1), M = L(\alpha_2) = K(\alpha_1, \alpha_2), \ldots$ Thus we need a theorem to tell us what happens when we extend an extension. This theorem is called the Tower Law because we often think of the extensions M/L and L/K as stacked one on top of the other.

Theorem 4 (Tower Law)**.** *Let M/L and L/K be two finite field extensions. Then M/K is also a finite extension, and*

$$[M : K] = [M : L][L : K].$$

Proof. (More detail in Galois Theory.)

Let $r = [L : K]$ and $s = [M : L]$. Let $\{\ell_1, \ldots, \ell_r\}$ be a K-basis for L and let $\{m_1, \ldots, m_s\}$ be an L-basis for M.

One can check that $\{l_i m_j : 1 \leq i \leq r, 1 \leq j \leq s\}$ is a K-basis for M. □

If $\alpha_1, \ldots, \alpha_r$ are algebraic over K, let $L = K(\alpha_1),\ M = K(\alpha_1, \alpha_2) = L(\alpha_2)$.

By Theorem 3, M/L and L/K are both finite extensions. Therefore by the Tower Law, M/K is a finite extension. Repeating this argument, $K(\alpha_1, \ldots, \alpha_r)/K$ is finite.

e.g. We can apply these results to the example from the previous lecture, in order to find the minimal polynomial of $\alpha = i + \sqrt{2}$:

Let $M = \mathbb{Q}(\alpha) \quad L = \mathbb{Q}(\sqrt{2}), \quad K = \mathbb{Q}$.

We saw that

$$\sqrt{2} = \frac{\alpha^2 + 3}{2\alpha}$$

so $L \subseteq M$. Thus we have a tower extensions M/L and L/K to which we can apply the Tower Law.

As observed last time, $[L : K] = [\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ since L has a K-basis $\{1, \sqrt{2}\}$.

Now $M = L(\alpha)$. To prove this: L and α are both contained in M, so $L(\alpha) \subseteq M$. Furthermore, $\mathbb{Q}$ and α are both contained in $L(\alpha)$, so $M = \mathbb{Q}(\alpha) \subseteq L(\alpha)$.

In fact, $M = L(i)$ because $i = \alpha - \sqrt{2}$ and $\sqrt{2} \in L$, so we can argue as in the previous paragraph: $L \subseteq L(i), \alpha \in L(i)$ so $L(\alpha) \subseteq L(i)$ and $L \subseteq M, i \in M$ so $L(i) \subseteq M$.

Note that $i \notin L$ because $L \subseteq \mathbb{R}$. Similarly to the extension $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$, $L(i) = L(\sqrt{-1})$ has L-basis $\{1, \sqrt{-1}\}$ so $[L(i) : L] = 2$.

So Theorem 4 tells us that $[M : K] = 2 \times 2 = 4$.

Now we can work out the minimal polynomial of α over $\mathbb{Q}$. Last lecture, we saw that $g(X) = X^4 - 2X^2 + 9$ is a polynomial which vanishes at α. We saw last time that $g(\alpha) = 0$. But Theorem 3 tells us that $\deg(\mu_{\mathbb{Q},\alpha}) = [K(\alpha) : K] = 4$. Hence g is monic and has the smallest possible degree for any polynomial vanishing at α, so it must be $\mu_{\mathbb{Q},\alpha}$. Using Lemma 1, we can also deduce that g is irreducible over $\mathbb{Q}$ without doing any more calculations.

We can write down two $\mathbb{Q}$-bases for $\mathbb{Q}(\alpha)$:

Using Theorem 3, a $\mathbb{Q}$-basis for $\mathbb{Q}(\alpha)$ is given by

$$\{1, \alpha, \alpha^2, \alpha^3\} = \{1, \sqrt{2} + i, 1 + 2\sqrt{2}i, -\sqrt{2} + 5i\}.$$

Using the Tower Law: L has a $\mathbb{Q}$-basis $\{1, \sqrt{2}\}$ while M has an L-basis $\{1, i\}$. Thus the proof of the Tower Law gives us the following $\mathbb{Q}$-basis for $\mathbb{Q}(\alpha)$:

$$\{1, \sqrt{2}, i, i\sqrt{2}\}.$$

Algebraic numbers.

Definition. An **algebraic number** is an element of $\mathbb{C}$ which is algebraic over $\mathbb{Q}$.

We write $\overline{\mathbb{Q}}$ for the set of algebraic numbers.

Lemma 5. *Let $\alpha, \beta \in \mathbb{C}$ be algebraic numbers. Then $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$ and α/β (if $\beta \neq 0$) are also algebraic numbers.*

Proof. Let $L = \mathbb{Q}(\alpha)$. Now β is algebraic over $\mathbb{Q}$, so it has a minimal polynomial $\mu_{\mathbb{Q},\beta}(X)$. Now $\mu_{\mathbb{Q},\beta}(X) \in \mathbb{Q}[X] \subseteq L[X]$, so β is algebraic over K. over L. (Note that $\mu_{\mathbb{Q},\beta}$ is not necessarily the minimal polynomial of β over L.)

By Theorem 3, $L/\mathbb{Q}$ and $L(\beta)/L$ are finite extensions. Hence by the Tower Law, $L(\beta)/\mathbb{Q}$ is also a finite extension.

By Lemma 2, we deduce that every element of $\mathbb{Q}(\alpha, \beta) = L(\beta)$ is an algebraic number. But $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$ and α/β are all elements of $\mathbb{Q}(\alpha, \beta)$. □

This is quite incredible! In a simple example, we had to do some work before to show that $i + \sqrt{2}$ was algebraic (and more to find its minimal polynomial). For example, if α is a root of

$$X^{10000} + 5X^{73} + 2X^8 - 6X - 22$$

and β is a root of

$$X^{99999} + 777X^2 - 5$$

then there is a polynomial with rational coefficients which has $\alpha + \beta$ as a root. Finding this polynomial is a hard computational problem (can you guess what its degree might be?) but the theorem tells us that it exists.

Corollary 6. *$\overline{\mathbb{Q}}$ is a field.*

Proof. Immediate corollary of Lemma 5. □

There will be a couple more lemmas from time to time, but we have now mostly finished the overlap with Galois theory.

4. Number fields and embeddings

Number fields.

Definition. A **number field** is a finite extension of $\mathbb{Q}$.

If we are more careful, we should perhaps say "a number field is a field K such that $K/\mathbb{Q}$ is a finite extension" – an extension is really a pair of fields, while a number field is just the bigger field in that pair (the smaller one is fixed to be $\mathbb{Q}$).

e.g. $\mathbb{Q}$ is the only number field of degree 1 (why?)
$\mathbb{Q}(\sqrt{d})$ is a number field of degree 2 when d is a non-square rational number
For any finite list of algebraic numbers $\alpha_1, \ldots, \alpha_r$, the field $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$ is a number field (we showed this last time).
Every number field can be written as $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$ for some $\alpha_1, \ldots, \alpha_r$: just pick $\alpha_1, \ldots, \alpha_r$ to be a $\mathbb{Q}$-basis for the field.

Examples of number fields.
The examples we will use most frequently in this course are quadratic fields – they are small enough to calculate many things by hand.

Definition. A **quadratic field** is a number field of degree 2.

We have already seen that $\mathbb{Q}(\sqrt{d})$ is a quadratic field if d is a non-square rational. Lots of d give the same quadratic field e.g.

$$\mathbb{Q}(\sqrt{2}) = \mathbb{Q}(\sqrt{8}) = \mathbb{Q}(\sqrt{18}) = \mathbb{Q}(\sqrt{1/2}) = \mathbb{Q}(\sqrt{9/8}) = \cdots$$

We can eliminate this redundancy by requiring d to be a square-free integer.

Definition. $d \in \mathbb{Z}$ is **square-free** if it is not divisible by m^2 for any integer $m > 1$. (Note: 1 is square-free, 0 is not.)

In fact all quadratic fields have this form and once we insist that d is a square-free integer, the representation is unique.

Lemma 7. *Let K be a quadratic field. Then $K = \mathbb{Q}(\sqrt{d})$ for a unique square-free integer $d \neq 1$.*

The proof is on example sheet 1.
Note that Lemma 7 does not generalise to higher-degree fields. For example a cubic field (i.e. a field of degree 3) does not always have the form $\mathbb{Q}(\sqrt[3]{d})$. An example of this will be on example sheet 1.

Another important example is cyclotomic fields.

Definition. Let n be a positive integer and let $\zeta_n = \exp(2\pi i/n)$ (a primitive n-th root of unity). We call $\mathbb{Q}(\zeta_n)$ the **n-th cyclotomic field**.

Lemma 8. *If $n = p$ is prime, then the minimal polynomial of ζ_p is $X^{p-1} + X^{p-2} + \cdots + X + 1$ and hence $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = p - 1$.*

Proof. This is on the example sheet. You will need to use Eisenstein's criterion from Algebra 2. □

If n is not a prime, then $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ (the Euler φ-function). There is no general formula for the minimal polynomial of ζ_n when n is not prime, so this is harder to prove.

Aside: The Primitive Element Theorem.

(non-examinable)

We said that every number field can be written in the form $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$. It turns out that actually every number field can be written in the form $\mathbb{Q}(\alpha)$ for a single element α. This doesn't require any tools beyond what we have done so far in this course, but would take most of a lecture to prove. Using the idea of separable polynomials from Galois theory, you could shorten that to half a lecture.

This is called the Primitive Element Theorem, and will be stated formally in lecture 11. The statement is examinable, but not the proof (which we won't give). Just an idea of the proof: an example we worked with in the last two lectures showed that

$$\mathbb{Q}(i, \sqrt{2}) = \mathbb{Q}(i + \sqrt{2}).$$

In general, it turns out that if you have $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$, then there is some linear combination of $\alpha_1, \ldots, \alpha_r$ which generates the extension as a single element – but it's not always just adding them up. If you are interested, there is a proof in my lecture notes from last year.

Embeddings of number fields.

Definition. Let K be a number field. An **embedding** of K is a field homomorphism $\sigma\colon K \to \mathbb{C}$.

To understand embeddings, we will need the following basic facts about field homomorphisms (whose proofs we omit because they are easy algebra).

Lemma 9. *Every homomorphism of fields is injective.*

This justifies the name "embedding."

Lemma 10. *Let K be a number field and let $\sigma\colon K \to \mathbb{C}$ be an embedding. Then $\sigma(a) = a$ for all $a \in \mathbb{Q}$.*

Consequently there is exactly one embedding $\sigma\colon \mathbb{Q} \to \mathbb{C}$, namely the inclusion.

What are the embeddings of a quadratic field $\mathbb{Q}(\sqrt{d})$? Thanks to Lemma 10, every embedding $\sigma\colon \mathbb{Q}(\sqrt{d}) \to \mathbb{C}$ must satisfy

$$\sigma(a + b\sqrt{d}) = \sigma(a) + \sigma(b)\sigma(\sqrt{d}) = a + b\sigma(\sqrt{d}).$$

Hence the embedding is fully determined once we know $\sigma(\sqrt{d})$. This must satisfy

$$\sigma(\sqrt{d})^2 = \sigma(d) = d$$

so there are two choices: $\sigma(\sqrt{d}) = \sqrt{d}$ or $\sigma(\sqrt{d}) = -\sqrt{d}$. Thus we get two possible embeddings:

$$\sigma_1(a + b\sqrt{d}) = a + b\sqrt{d}, \quad \sigma_2(a + b\sqrt{d}) = a - b\sqrt{d}.$$

One can check by calculation that these are both field homomorphisms.

We have seen that $\mathbb{Q}$, the only number field of degree 1, has one embedding. Quadratic fields, that is, number fields of degree 2, have two embeddings. One might guess that a number field of degree n has n embeddings. We will prove that in the next lecture.

In the proof, we write a number field as $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$ and build up first embeddings of $\mathbb{Q}(\alpha_1)$, then embeddings of $\mathbb{Q}(\alpha_1, \alpha_2)$ etc. We will need the following definition.

Definition. Let L/K be an extension of number fields. Let $\sigma\colon K \to \mathbb{C}$ and $\tau\colon L \to \mathbb{C}$ be embeddings. We say that τ **extends** σ if $\tau_{|K} = \sigma$.

We will also need a piece of notation. If $\sigma\colon K \to \mathbb{C}$ is a field homomorphism, then it induces an injective ring homomorphism $K[X] \to \mathbb{C}[X]$ which we also call σ, defined by

$$\sigma(a_0 + a_1X + \cdots + a_nX^n) = \sigma(a_0) + \sigma(a_1)X + \cdots + \sigma(a_n)X^n.$$

(There is nothing special about $\mathbb{C}$ here, or even about fields. We could define the same homomorphism of polynomial rings from any homomorphism of rings.)

5. Embeddings of number fields

Extending embeddings of number fields.

Lemma 11. *Let L/K be an extension of number fields such that $L = K(\alpha)$. Let $\sigma\colon K \to \mathbb{C}$ be an embedding.*

Let μ_α be the minimal polynomial of α over K and let $\alpha_1, \dots, \alpha_n$ be the roots of $\sigma(\mu_\alpha)$ in $\mathbb{C}$.

(i) For each embedding $\tau\colon L \to \mathbb{C}$ extending σ, $\tau(\alpha)$ is one of the roots $\alpha_1, \dots, \alpha_n$.

(ii) For each $i = 1, \dots, n$, there is a unique embedding $\tau_i\colon L \to \mathbb{C}$ extending σ such that $\tau_i(\alpha) = \alpha_i$.

Proof. The proofs of (i) and the uniqueness part of (ii) are just algebraic manipulation, resembling how we found all the embeddings of a quadratic field.

(i) Let $\tau\colon L \to \mathbb{C}$ be an embedding extending σ. We have

$$\sigma(\mu_\alpha)(\tau(\alpha)) = \tau(\mu_\alpha)(\tau(\alpha)) = \tau(\mu_\alpha(\alpha)) = \tau(0) = 0.$$

The first equality uses the fact that τ extends σ, the second the fact that τ is a homomorphism of rings.

Thus $\tau(\alpha)$ is a root of $\sigma(\mu_\alpha)$. In other words, it is one of $\alpha_1, \dots, \alpha_n$.

(ii) First we show uniqueness – like (i), this is just algebraic manipulation.

Let $\tau\colon L \to \mathbb{C}$ be an embedding extending σ. Thank to Theorem 3, we can write any $\beta \in L$ in the form

$$\beta = b_0 + b_1\alpha + \cdots + b_r\alpha^r$$

for some $b_0, b_1, \dots, b_r \in K$ (where $r = [L : K] - 1$. Then

$$\begin{aligned}\tau(\beta) &= \tau(b_0) + \tau(b_1)\tau(\alpha) + \cdots + \tau(b_r)\tau(\alpha)^r \\ &= \sigma(b_0) + \sigma(b_1)\tau(\alpha) + \cdots + \sigma(b_r)\tau(\alpha)^r.\end{aligned}$$

Thus knowing $\tau(\alpha) = \alpha_i$ uniquely determines τ on all of L.

The proof of the existence part of (ii) is more abstract.

Thanks to Lemma 9, σ is an isomorphism of fields $K \to \sigma(K)$. Consequently $\sigma(\mu_\alpha) \in \sigma(K)[X]$ is irreducible over $\sigma(K)$.

Also, $\sigma(\mu_\alpha)$ is monic and $\sigma(\mu_\alpha)(\alpha_i) = 0$. Hence $\sigma(\mu_\alpha)$ is the minimal polynomial of α_i over $\sigma(K)$. Therefore, using Theorem 3 twice, we have

$$L = K(\alpha) \cong K[X]/(\mu_\alpha) \cong \sigma(K)[X]/(\sigma(\mu_\alpha)) \cong \sigma(K)(\alpha_i). \qquad (*)$$

(The middle isomorphism comes from simply applying the isomorphism σ to the coefficients of the polynomials.)

Composing all the isomorphisms from (*) and the inclusion map $\sigma(K)(\alpha_i) \to \mathbb{C}$ gives an embedding $\tau_i\colon L \to \mathbb{C}$. This embedding extends σ because the first isomorphism in (*) is the identity on K, the middle isomorphism restricts to σ on K, and the third isomorphism is the identity on $\sigma(K)$. Furthermore, we have $\tau(\alpha) = \alpha_i$ because the isomorphisms from (*) map α as follows:

$$\alpha \mapsto X + \langle \mu_\alpha \rangle \mapsto X + \langle \sigma(\mu_\alpha) \rangle \mapsto \alpha_i.$$

Thus we have constructed $\tau_i \colon L \to \mathbb{C}$ extending σ and satisfying $\tau_i(\alpha) = \alpha_i$. $\square$

In order make use of this lemma, we need to know how many roots the polynomial $\sigma(\mu_\alpha)$ has in $\mathbb{C}$. We know that this polynomial has degree $[L : K]$, so there are $[L : K]$ complex roots "counted with multiplicity", but we have to rule out multiple roots. It turns out that this is possible because μ_α is irreducible over K.

This is done by the following lemma. Something close to this lemma will be proved in Galois Theory (in the language of Galois Theory, the statement is "every irreducible polynomial over a field of characteristic zero is separable.") As mentioned above, the difficulty is using irreducibility to rule out repeated roots.

Lemma 12. *Let K be a number field and let $\sigma \colon K \to \mathbb{C}$ be an embedding. Let $f \in K[X]$ be an irreducible polynomial over K of degree n. Then $\sigma(f)$ has exactly n distinct roots in $\mathbb{C}$.*

Proposition 13. *A number field K has exactly $[K : \mathbb{Q}]$ embeddings.*

Proof. Write $K = \mathbb{Q}(\alpha_1, \ldots, \alpha_m)$ and $K_i = K(\alpha_1, \ldots, \alpha_i)$ for $0 \leq i \leq m$. The proof is by induction on i.

The base case $K_0 = \mathbb{Q}$ holds, because $\mathbb{Q}$ has one embedding.

For $i \geq 1$: We have $K_i = K_{i-1}(\alpha_i)$. Let μ_i be the minimal polynomial of α_i over K_{i-1}.

For each embedding σ of K_{i-1}, we are going to count the number of ways to extend it to an embedding of K_i. By Lemma 11, the number of embeddings of K_i extending σ is equal to the number of complex roots of $\sigma(\mu_i)$. By Lemma 12, there are exactly $\deg(\mu_i)$ such roots. By Theorem 3, $\deg(\mu_i) = [K_i : K_{i-1}]$. So we conclude that there are $[K_i : K_{i-1}]$ embeddings of K_i extending σ.

By induction, K_{i-1} has $[K_{i-1} : \mathbb{Q}]$ embeddings. The previous paragraph applies to each of them, so the total number of embeddings of K_i is

$$[K_i : K_{i-1}][K_{i-1} : \mathbb{Q}]$$

which is equal to $[K_i : \mathbb{Q}]$ by the Tower Law. $\square$

Real and complex embeddings.

Definition. Let $\sigma \colon K \to \mathbb{C}$ be an embedding of a number field. We say that σ is a **real embedding** if $\sigma(K) \subseteq \mathbb{R}$ and σ is a **complex embedding** if $\sigma(K) \not\subseteq \mathbb{R}$.

e.g. a quadratic field $\mathbb{Q}(\sqrt{d})$: there are two embeddings, σ_1 mapping $\sqrt{d}$ to $\sqrt{d}$ and σ_2 mapping $\sqrt{d}$ to $-\sqrt{d}$.

If $d > 0$, then $\sqrt{d}$ and $-\sqrt{d}$ are both real, so both embeddings of $\mathbb{Q}(\sqrt{d})$ are real. If $d < 0$, then $\sqrt{d}$ and $-\sqrt{d}$ are both non-real, so both embeddings of $\mathbb{Q}(\sqrt{d})$ are complex.

For number fields of higher degree, the same number field may have some real embeddings and some complex embeddings.

6. Signature, norm and trace

Signature of a number field.

Last lecture we classified embeddings of a number field as real or complex, depending on whether their image lies in $\mathbb{R}$ or not.

Note that complex embeddings come in conjugate pairs: if σ is an embedding of K, then

$$\bar{\sigma}(\alpha) = \overline{\sigma(\alpha)}$$

is also an embedding of K (where the bar denotes complex conjugation). If σ is a complex embedding, then σ and $\bar{\sigma}$ are different. If σ is a real embedding, then $\sigma = \bar{\sigma}$.

Definition. The **signature** of K is (r, s) where $r =$ number of real embeddings of K, $s = \frac{1}{2} \times$ number of complex embeddings of K.

e.g. Signature of a real quadratic field is $(2, 0)$
Signature of an imaginary quadratic field is $(0, 1)$

Let K be a number field with signature (r, s).

Counting up all the embeddings and applying Proposition 13, we see that $[K : \mathbb{Q}] = r + 2s$.

We often label the embeddings as $\sigma_1, \ldots, \sigma_r$ (the real embeddings), $\sigma_{r+1}, \ldots, \sigma_{r+s}$, $\overline{\sigma_{r+1}}, \ldots, \overline{\sigma_{r+s}}$ (the complex embeddings).

Outline of course.

Now that we know what a number field is, it might be helpful to outline what we will do with them in the course.

(1) Number fields and tools for working with them. Embeddings which we have just defined, norm and trace which we are about to discuss.
(2) Algebraic integers. Inside each number field we define a "ring of integers" like $\mathbb{Z}[i]$ inside $\mathbb{Q}(i)$, or indeed $\mathbb{Z}$ inside $\mathbb{Q}$. (But unlike those simple examples, it turns out that in order to define rings of integers, we need to start with the number field and then decide which elements of that field behave like integers.)
(3) Factorisation in rings of integers. In general, the ring of integers of a number field is not a unique factorisation domain. However, they do have "unique factorisation of ideals" into prime ideals. We will explain what this means and prove it. Then we will look at the Dedekind–Kummer theorem which tells us how to find the prime ideals in a ring of integers.
(4) The class group. The class group of a number field is a finite abelian group defined as approximately "ideals in the ring of integers quotiented by elements of the ring of integers." Since we have unique factorisation, the class group measures how badly unique factorisation fails. Big theorem: the class group is finite (proof partially examinable). Method for calculating the class group.

(5) Units in the ring of integers. Units are integer whose inverse is also an integer (e.g. the units in $\mathbb{Z}$ are ± 1). They don't have any prime factors, so aren't taken care of in our earlier consideration of factorisation. Big theorem: structure of the units in a number field (proof not examinable).
(6) Diophantine equations. We will apply factorisation and units in number fields to study integer solutions of two types of equation: Mordell's equation $y^2 = x^3 + k$ and Pell's equation $x^2 - dy^2 = 1$.

Norm and trace.

Let K be a number field. We use linear algebra to define two functions $K \to \mathbb{Q}$ which can be helpful in transforming questions about elements of K into simpler questions about rational numbers.

Recall that K is $\mathbb{Q}$-vector space. For any element $\alpha \in K$ multiplication by α is a $\mathbb{Q}$-linear map $m_{K,\alpha} \colon K \to K$:

$$m_{K,\alpha}(\beta) = \alpha\beta.$$

Definition. The **trace** of α is the trace of the linear map $m_{K,\alpha}$.

The **norm** of α is the determinant of the linear map $m_{K,\alpha}$.

These are written $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha)$ and $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$ respectively.

The notation (subscript $K/\mathbb{Q}$) reminds us that $\mathrm{Tr}_{K/\mathbb{Q}}$ and $\mathrm{Nm}_{K/\mathbb{Q}}$ are functions $K \to \mathbb{Q}$.

Note that you could generalise this: instead of always having $\mathbb{Q}$ as the base field, you could define $\mathrm{Tr}_{L/K}$ and $\mathrm{Nm}_{L/K}$ for any extension of number fields L/K. These would be functions $L \to K$. The definition is essentially the same but we won't need this generalisation in this module.

e.g. Let $K = \mathbb{Q}(\sqrt{d})$. What are the norm and trace of $\alpha = a + b\sqrt{d}$?

Write $m_{K,\alpha} \colon K \to K$ as a matrix with respect to the basis $\{1, \sqrt{d}\}$. We get

$$m_{K,\alpha}(1) = a \cdot 1 + b \cdot \sqrt{d}, \quad m_{K,\alpha}(\sqrt{d}) = bd \cdot 1 + a \cdot \sqrt{d}$$

so the matrix of $m_{K,\alpha}$ (with respect to this basis) is

$$\begin{pmatrix} a & bd \\ b & a \end{pmatrix}.$$

Thus

$$\mathrm{Tr}_{K/\mathbb{Q}}(a + b\sqrt{d}) = 2a.$$
$$\mathrm{Nm}_{K/\mathbb{Q}}(a + b\sqrt{d}) = a^2 - b^2 d,$$

Lemma 14. *The trace is additive and the norm is multiplicative. In other words, for all α, β in K, we have*

$$\mathrm{Tr}_{K/\mathbb{Q}}(\alpha + \beta) = \mathrm{Tr}_{K/\mathbb{Q}}(\alpha) + \mathrm{Tr}_{K/\mathbb{Q}}(\beta),$$
$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha\beta) = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)\,\mathrm{Nm}_{K/\mathbb{Q}}(\beta).$$

Proof. Observe that $m_{K,\alpha+\beta} = m_{K,\alpha} + m_{K,\beta}$ and $m_{K,\alpha\beta} = m_{K,\alpha} m_{K,\beta}$. Thus the lemma follows from the properties of trace and determinant of linear maps. □

Characteristic polynomials.

Let V be a $\mathbb{Q}$-vector space and let $f\colon V \to V$ be a $\mathbb{Q}$-linear map. Recall that the **characteristic polynomial** of f is the polynomial

$$\chi_f(X) = \det(XI - f) \in \mathbb{Q}[X].$$

This polynomial is monic of degree $n = \dim(V)$. We can read off the determinant and trace of f from the coefficients of the characteristic polynomial: if $\chi_f = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, then

$$\mathrm{Tr}(f) = -a_{n-1}, \quad \det(f) = (-1)^n a_0. \tag{*}$$

Write $\chi_{K,\alpha}$ for the characteristic polynomial of the $\mathbb{Q}$-linear map $m_{K,\alpha}$. Using (*), we can read off the norm and trace of $\alpha \in K$ from the coefficients of $\chi_{K,\alpha}$.

The next lemmas tell us how to work out the characteristic polynomial $\chi_{K,\alpha}$. Firstly, if $K = \mathbb{Q}(\alpha)$, then it is equal to the minimal polynomial.

Lemma 15. *Let $K = \mathbb{Q}(\alpha)$. Then the characteristic polynomial of $m_{K,\alpha}\colon K \to K$ is equal to the minimal polynomial of α over $\mathbb{Q}$.*

Proof. Let $\chi_{K,\alpha}(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$ denote the characteristic polynomial of $m_{K,\alpha}$. By the Cayley–Hamilton theorem, $\chi_{K,\alpha}(m_{K,\alpha}) = 0$. In other words,

$$m_{K,\alpha}^n + a_{n-1}m_{K,\alpha}^{n-1} + \cdots + a_1 m_{K,\alpha} + a_0 = 0 \tag{*}$$

(in the ring of $\mathbb{Q}$-linear maps $K \to K$).

Evaluating the map (*) at $1 \in K$, and noting that $m_{K,\alpha}^i(1) = \alpha^i$, we get

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

or in other words $\chi_{K,\alpha}(\alpha) = 0$.

Furthermore $\chi_{K,\alpha} \in \mathbb{Q}[X]$, $\chi_{K,\alpha}$ is monic and $\deg(\chi_{K,\alpha}) = [K : \mathbb{Q}] = \deg(\mu_{\mathbb{Q},\alpha})$ (the latter holds by Theorem 3 because $K = \mathbb{Q}(\alpha)$). Hence by Lemma 1, $\chi_{K,\alpha} = \mu_{\mathbb{Q},\alpha}$. □

When K is bigger than $\mathbb{Q}(\alpha)$, we can use the following lemma to work out the characteristic polynomial.

Lemma 16. *Let L/K be an extension of number fields and let $\alpha \in K$. Let $\chi_{K,\alpha}$ and $\chi_{L,\alpha}$ be the characteristic polynomials of $m_{K,\alpha}\colon K \to K$ and $m_{L,\alpha}\colon L \to L$ respectively. Then*

$$\chi_{L,\alpha} = \chi_{K,\alpha}^{[L:K]}.$$

Proof. By the Tower Law, we can find a $\mathbb{Q}$-basis for L of the form

$$\{k_i\ell_j : 1 \le i \le r, 1 \le j \le s\}$$

where $\{k_1, \ldots, k_r\}$ is a $\mathbb{Q}$-basis for K and $\ell_1, \ldots, \ell_s$ is a K-basis for L, $r = [K : \mathbb{Q}]$ and $s = [L : K]$.

Let $M_{K,\alpha}$ be the matrix of the linear map $m_{K,\alpha}$ with respect to the basis $\{k_1, \ldots, k_r\}$. Some calculations show that the matrix for $m_{L,\alpha}$ with respect to the basis $\{k_i \ell_j\}$ is block diagonal with blocks that are copies of $M_{K,\alpha}$:

$$M_{L,\alpha} = \begin{pmatrix} M_{K,\alpha} & 0 & \cdots & 0 \\ 0 & M_{K,\alpha} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & M_{K,\alpha} \end{pmatrix}$$

There is one block for each ℓ_j i.e. s blocks. This is consistent with the fact that $M_{K,\alpha}$ is an $s \times s$ matrix and $M_{L,\alpha}$ is an $rs \times rs$ matrix.

The characteristic polynomial of a block diagonal matrix is the product of the characteristic polynomials of the blocks (because the same thing holds for determinants). Thus

$$\chi_{L,\alpha}(X) = \chi_{K,\alpha}(X)^s. \qquad \square$$

7. Conjugates, algebraic integers

More on characteristic polynomials.

Corollary 17. *Let L/K be an extension of number fields. If $\alpha \in K$, then*

$$\mathrm{Tr}_{L/\mathbb{Q}}(\alpha) = [L:K]\,\mathrm{Tr}_{K/\mathbb{Q}}(\alpha),$$
$$\mathrm{Nm}_{L/\mathbb{Q}}(\alpha) = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)^{[L:K]}.$$

Proof. Let $r = [K:\mathbb{Q}]$ and $s = [L:K]$. Write

$$\chi_{K,\alpha}(x) = X^r + a_{r-1}X^{r-1} + \cdots + a_1X + a_0,$$
$$\chi_{L,\alpha}(x) = X^{rs} + b_{rs-1}X^{rs-1} + \cdots + b_1X + b_0.$$

By Lemma 16, we have $\chi_{L,\alpha} = \chi_{K,\alpha}^s$. Expanding this out, we get

$$\chi_{L,\alpha} = X^{rs} + sa_{r-1}X^{rs-1} + \cdots + a_0^s.$$

Consequently

$$\mathrm{Tr}_{L/\mathbb{Q}}(\alpha) = -b_{rs-1} = -sa_{r-1} = s\,\mathrm{Tr}_{K/\mathbb{Q}}(\alpha),$$
$$\mathrm{Nm}_{L/\mathbb{Q}}(\alpha) = (-1)^{rs}b_0 = (-1)^{rs}a_0^s = ((-1)^ra_0)^s = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)^s. \qquad \square$$

By combining Lemmas 15 and 16, we can work out the characteristic polynomial of an arbitrary $\alpha \in K$ in terms of the minimal polynomial over $\mathbb{Q}$:

$$\chi_{K,\alpha} = \mu_{\mathbb{Q},\alpha}^{[K:\mathbb{Q}(\alpha)]}$$

It can be useful to apply this in reverse in order to work out the minimal polynomial. By choosing a basis for K, we can work out the characteristic polynomial $\chi_{K,\alpha}$ (especially when $[K:\mathbb{Q}] = 3$, this is a reasonable calculation to do by hand). This must be a power of an irreducible polynomial, which will be the minimal polynomial of α. Checking whether a polynomial is a power of an irreducible polynomial is easier than checking if it is irreducible.

Conjugates.

Definition. Let α be an algebraic number. The **conjugates** of α are the roots (in $\mathbb{C}$) of $\mu_{\mathbb{Q},\alpha}$, the minimal polynomial of α over $\mathbb{Q}$.

By Lemma 11, the conjugates of α are $\sigma_1(\alpha), \ldots, \sigma_n(\alpha)$ where $\sigma_1, \ldots, \sigma_n$ are the embeddings of $\mathbb{Q}(\alpha)$.

We can express the norm and trace in terms of the conjugates of α.

Lemma 18. *Let $\sigma_1, \ldots, \sigma_n$ denote the embeddings $K \to \mathbb{C}$. Then for any $\alpha \in K$,*

$$\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) = \sum_{i=1}^{n} \sigma_i(\alpha),$$
$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha).$$

Proof. First suppose that $K = \mathbb{Q}(\alpha)$. By Lemma 15, $\chi_{K,\alpha}$ = minimal polynomial of α over $\mathbb{Q}$. Hence

$$\{\sigma_1(\alpha), \ldots, \sigma_n(\alpha)\} = \text{conjugates of } \alpha = \text{roots of } \mu_{\mathbb{Q},\alpha}$$
$$= \text{roots of } \chi_{K,\alpha} = \text{eigenvalues of } m_{K,\alpha}.$$

By Lemma 12, $\mu_{\mathbb{Q},\alpha}$ has no repeated roots in $\mathbb{C}$, so we don't need to worry about multiplicities of roots/eigenvalues. Hence

$$\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) = \mathrm{Tr}(m_{K,\alpha}) = \text{sum of eigenvalues of } m_{K,\alpha} = \sum_{i=1}^{n} \sigma_i(\alpha),$$

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \det(m_{K,\alpha}) = \text{product of eigenvalues of } m_{K,\alpha} = \prod_{i=1}^{n} \sigma_i(\alpha).$$

Now consider a field K which strictly contains $\mathbb{Q}(\alpha)$. Let $s = [K : \mathbb{Q}(\alpha)]$ and $r = [\mathbb{Q}(\alpha) : \mathbb{Q}]$. Thanks to Lemmas 11 and 12, for each embedding of $\mathbb{Q}(\alpha)$, there are s embeddings of K extending it. Hence the values $\sigma_1(\alpha), \ldots, \sigma_n(\alpha)$ consist of each of the conjugates $\tau_1(\alpha), \ldots, \tau_r(\alpha)$, each repeated s times (where $\tau_1, \ldots, \tau_r$ are the embeddings of $\mathbb{Q}(\alpha)$). Thus

$$\sum_{i=1}^{n} \sigma_i(\alpha) = s \cdot \sum_{i=1}^{r} \tau_i(\alpha) = s \cdot \mathrm{Tr}_{\mathbb{Q}(\alpha)/\mathbb{Q}}(\alpha) = \mathrm{Tr}_{K/\mathbb{Q}}(\alpha)$$

where the second equality holds because we have already proved the lemma for $\mathbb{Q}(\alpha)$, and the third by Corollary 17.

The same argument works for norm, considering products instead of sums. □

Algebraic integers.

We have finished understanding number fields and tools for working with them *as fields*. Now we want to talk about a version of "integers" inside number fields, where we will have interesting properties relating to factorisation and primes.

We have already seen one example: in $\mathbb{Q}(i)$, we have the Gaussian integers

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

So maybe we could generalise this, and define the "integers" in $\mathbb{Q}(\alpha)$ to be

$$\mathbb{Z}[\alpha] = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_r\alpha^r : r \in \mathbb{N}, a_i \in \mathbb{Z}\}.$$

Unfortunately, this doesn't work: it depends on α, even when the field $\mathbb{Q}(\alpha)$ stays the same. For example, $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}(\sqrt{8})$ but $\mathbb{Z}[\sqrt{2}] \neq \mathbb{Z}[\sqrt{8}]$ (and $\mathbb{Z}[\sqrt{\frac{1}{2}}]$ is worse – it contains $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$.)

Instead, we will imitate the definition of an algebraic number as a root of a polynomial with coefficients in $\mathbb{Q}$. Maybe we could just replace $\mathbb{Q}$ with $\mathbb{Z}$ and define an algebraic integer to be the roots of a polynomial with coefficients in $\mathbb{Z}$?

That's no good, because $\frac{1}{2}$ is a root of $2X - 1 \in \mathbb{Z}[X]$ and $\frac{1}{2}$ should not be anything like an integer. More generally, every algebraic number is a root of some polynomial with coefficients in $\mathbb{Z}$ because we can multiply by a constant to clear denominators.

We will avoid this problem by restricting attention to monic polynomials (we can't multiply a monic polynomial by a constant and have it stay monic).

Definition. An algebraic number is an **algebraic integer** if it is a root of some monic polynomial with coefficients in $\mathbb{Z}$.

e.g. Any integer $n \in \mathbb{Z}$ is a root of $X - n$

$\sqrt{d}$ (where $d \in \mathbb{Z}$) is an algebraic integer because it is a root of $X^2 - d$.

$\frac{-1+\sqrt{-3}}{2} = \exp(2\pi i/3)$ is an algebraic integer because it is a root of $X^3 - 1$. Its minimal polynomial is $X^2 + X + 1$, also with integer coefficients.

The above definition is useful for showing that particular numbers are algebraic integers, but not useful for proving that numbers are *not* algebraic integers because it requires checking all polynomials in $\mathbb{Z}[X]$. Instead we use the following lemma.

Lemma 19. *An algebraic number is an algebraic integer if and only if its minimal polynomial over $\mathbb{Q}$ has coefficients in $\mathbb{Z}$.*

We will prove this lemma next time. For now, we use it to give some examples: $\frac{1}{2}$ has minimal polynomial $X - \frac{1}{2}$ so is not an algebraic integer. Similarly, any element of $\mathbb{Q} \setminus \mathbb{Z}$ is not an algebraic integer.

In contrast with the example of $\frac{-1+\sqrt{-3}}{2}$, $\frac{1+\sqrt{3}}{2}$ has minimal polynomial $X^2 - X - \frac{1}{2}$ so it is not an algebraic integer.

8. Algebraic integers

We begin by proving the following lemma, which we stated last time.

Lemma 19. *An algebraic number is an algebraic integer if and only if its minimal polynomial over $\mathbb{Q}$ has coefficients in $\mathbb{Z}$.*

The proof uses Gauss's Lemma from Algebra 2, which we now recall.

Definition. Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{Z}[X]$ be a polynomial with integer coefficients. We say that f is **primitive** if $\mathrm{HCF}(a_0, a_1, \ldots, a_n) = 1$.

Lemma (Gauss's Lemma)**.** *A primitive polynomial is irreducible in $\mathbb{Z}[X]$ if and only if it is irreducible in $\mathbb{Q}[X]$.*

(The key property that makes Gauss's Lemma work is that $\mathbb{Z}$ is a unique factorisation domain.)

Proof of Lemma 19. If $\mu_{\mathbb{Q},\alpha}$ has integer coefficients, then that gives a monic polynomial in $\mathbb{Z}[X]$ which has α as a root so α is an algebraic integer.

The main thing we have to prove is the other direction. Let α be an algebraic number.

By definition, there exists a monic polynomial $f(X) \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$. Choose f *of smallest degree* satisfying these conditions.

Let $\mu_\alpha(X) \in \mathbb{Q}[X]$ be the minimal polynomial of α over $\mathbb{Q}$. By Lemma 1, μ_α divides f in $\mathbb{Q}[X]$. Hence $\deg(f) \geq \deg(\mu_\alpha)$.

Assume for contradiction that $\deg(f) > \deg(\mu_\alpha)$. Then the fact that μ_α divides f shows that f is reducible in $\mathbb{Q}[X]$. Since f is monic, it is primitive. Hence by Gauss's Lemma, f is reducible in $\mathbb{Z}[X]$, i.e.

$$f = f_1 f_2 \text{ where } f_1, f_2 \in \mathbb{Z}[X] \text{ and } \deg(f_1), \deg(f_2) < \deg(f).$$

The leading coefficient of f is the product of the leading coefficients of f_1 and f_2. Thus these are integers whose product is 1, so both f_1 and f_2 have leading coefficient ± 1. Changing the sign if necessary, we may ensure that f_1 and f_2 are both monic.

Since $f(\alpha) = 0$, either $f_1(\alpha) = 0$ or $f_2(\alpha) = 0$. Thus either f_1 or f_2 gives us a monic polynomial in $\mathbb{Z}[X]$ with α as a root, contradicting the fact that f has the smallest degree.

Thus in fact $\deg(f) = \deg(\mu_\alpha)$. Since μ_α divides f and μ_α and f are both monic, this implies that $\mu_\alpha = f$ and so $\mu_\alpha \in \mathbb{Z}[X]$. □

Algebraic integers and finitely generated abelian groups.

Now we will show that the algebraic integers form a ring. The strategy resembles the proof that the algebraic numbers form a field, but is a bit harder.

The proof that algebraic numbers form a field relied on the idea of finite extensions of $\mathbb{Q}$. For algebraic integers, this notion is replaced by "rings finitely generated as an abelian group."

When we talk about a ring as an abelian group, we mean with the operation of addition. It is important to note that "finitely generated as an abelian group" is a much stronger property than "finitely generated as a ring."

Lemma 20. *Let α be an algebraic integer of degree n. Then $\mathbb{Z}[\alpha]$ is generated as an additive abelian group by $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$.*

Proof. Let $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{Z}[X]$ be the minimal polynomial of α.

We prove by induction on m that $\alpha^m \in \mathbb{Z}.1 + \mathbb{Z}.\alpha + \cdots + \mathbb{Z}.\alpha^{n-1}$ for all non-negative integers n.

For $m < n$, this is trivial. For $m \geq n$, since $f(\alpha) = 0$, we get

$$\alpha^m = \alpha^{m-n}\alpha^n = \alpha^{m-n}(-a_{n-1}\alpha^{n-1} - \cdots - a_1\alpha - a_0).$$

(The fact that f is monic is crucial here!) Thus α^m is a $\mathbb{Z}$-linear combination of smaller powers of α, and hence by induction it is a $\mathbb{Z}$-linear combination of $1, \alpha, \ldots, \alpha^{n-1}$.

Since the powers of α generate $\mathbb{Z}[\alpha]$ as an abelian group by definition, we conclude that $1, \alpha, \ldots, \alpha^{n-1}$ generate $\mathbb{Z}[\alpha]$ as an abelian group. □

The following lemma talks about all additive subgroups of $\mathbb{C}$, not just rings. So it is more general than we need for now. But we will need the full strength of the lemma later in the course.

Lemma 21. *(Integral Stability Lemma) Let H be a non-trivial finitely generated additive subgroup of $\mathbb{C}$. If $\alpha \in \mathbb{C}$ satisfies $\alpha H \subseteq H$, then α is an algebraic integer.*

Proof. Choose a finite set of generators $\{\beta_1, \ldots, \beta_n\}$ for H. Since $\alpha H \subseteq H$, each $\alpha\beta_j$ can be written as a $\mathbb{Z}$-linear combination of the generators:

$$\alpha\beta_j = \sum_{i=1}^{n} A_{ij}\beta_i \qquad (*)$$

where $A_{ij} \in \mathbb{Z}$ for $1 \leq i, j \leq n$.

Now A is an $n \times n$ square matrix with entries in $\mathbb{Z}$. Let $\underline{v}$ be the column vector $(\beta_1, \ldots, \beta_n)^t \in \mathbb{C}^n$. From (*), we get

$$A^T\underline{v} = \alpha\underline{v}$$

and so α is an eigenvalue of A^T. Thus α is a root of the characteristic polynomial of A^T, which is a monic polynomial with integer coefficients. So α is an algebraic integer. □

Lemma 22. *Let $\alpha \in \mathbb{C}$. Then α is an algebraic integer if and only if $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group.*

Proof. If α is an algebraic integer, then Lemma 20 tells us that $\mathbb{Z}[\alpha]$ is generated as an abelian group by the finite set $1, \alpha, \ldots, \alpha^{n-1}$.

Conversely, if $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group, then we can apply Lemma 21 to $H = \mathbb{Z}[\alpha]$. Since $\alpha H \subseteq H$, we conclude that α is an algebraic integer. □

The algebraic integers form a ring.

Notation. We write $\overline{\mathbb{Z}} = \{\alpha \in \mathbb{C} : \alpha \text{ is an algebraic integer}\}$.

We want to prove that the algebraic integers form a ring. When we proved that the algebraic numbers form a field, the method was to show that if $\alpha, \beta \in \overline{\mathbb{Q}}$ then $\alpha + \beta$ and $\alpha\beta$ are contained in a finite extension of $\mathbb{Q}$, and hence are algebraic numbers. In the same way, instead of considering the ring of algebraic integers all at once, we focus in on just two of them and show that if $\alpha, \beta \in \overline{\mathbb{Z}}$ then $\alpha + \beta$ and $\alpha\beta$ are contained in a ring which is finitely generated as an abelian group. This relies on both directions of Lemma 22.

Lemma 23. *$\overline{\mathbb{Z}}$ is a ring.*

Proof. Let $\alpha, \beta \in \overline{\mathbb{Z}}$. We have to show that $\alpha + \beta$, $\alpha\beta$ and $-\alpha \in \overline{\mathbb{Z}}$.

The easy one is $-\alpha$. (It was so easy it was skipped out in the lecture.) By Lemma 22, $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group. Now $\mathbb{Z}[-\alpha] = \mathbb{Z}[\alpha]$ so by the reverse direction of Lemma 22, $-\alpha \in \overline{\mathbb{Z}}$.

For α and β, we start by using Lemma 22 to say that $\mathbb{Z}[\alpha]$ and $\mathbb{Z}[\beta]$ are finitely generated as abelian groups. Let $\theta_1, \ldots, \theta_r$ be generators of $\mathbb{Z}[\alpha]$ and let $\phi_1, \ldots, \phi_s$ be generators of $\mathbb{Z}[\beta]$ as abelian groups.

Write $\mathbb{Z}[\alpha, \beta]$ for the smallest ring containing α and β, that is,

$$\mathbb{Z}[\alpha, \beta] = \Big\{ \sum_{i,j=0}^{m} c_{ij}\alpha^i\beta^j : m \in \mathbb{N}, c_{ij} \in \mathbb{Z} \Big\}.$$

Each $\alpha^i\beta^j$ is a $\mathbb{Z}$-combination of θs multiplied by a $\mathbb{Z}$-combination of ϕs. Thus it is a $\mathbb{Z}$-combination of $\theta_k\phi_\ell$s. Hence $\{\theta_k\phi_\ell : 1 \le k \le r, 1 \le \ell \le s\}$ generates $\mathbb{Z}[\alpha, \beta]$ as an abelian group (this is like the Tower Law for fields).

Thus $\mathbb{Z}[\alpha, \beta]$ is finitely generated as an abelian group. Every subgroup of a finitely generated abelian group is finitely generated, so in particular $\mathbb{Z}[\alpha + \beta]$ and $\mathbb{Z}[\alpha\beta]$ are finitely generated as abelian groups. Hence by the reverse direction of Lemma 22, $\alpha + \beta$ and $\alpha\beta$ are algebraic integers.

So $\overline{\mathbb{Z}}$ is closed under the ring operations of addition, multiplication and additive inverses. It also contains 0 and 1, so it is a ring. □

9. Rings of integers

Ring of integers of a number field.

Definition. If K is a number field, the **ring of integers** of K is $\overline{\mathbb{Z}} \cap K$, written $\mathcal{O}_K$.

Since $\overline{\mathbb{Z}}$ and K are both rings, so is their intersection $\mathcal{O}_K$.

e.g. The ring of integers of $\mathbb{Q}$ is $\mathbb{Z}$, because the minimal polynomial of $a \in \mathbb{Q}$ is $X - a$.

In order to avoid confusion with algebraic integers, we sometimes call an element of $\mathbb{Z}$ a **rational integer**.

Here are some basic properties of algebraic integers.

Lemma 24. *K is the field of fractions of $\mathcal{O}_K$.*

More strongly, every $\alpha \in K$ can be written as a/b where $a \in \mathcal{O}_K$ and $b \in \mathbb{Z} \setminus \{0\}$.

Proof. Let the minimal polynomial of α be $f(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_1 X + c_0$, where $a_0, \ldots, a_{n-1} \in \mathbb{Q}$. Let m be the lowest common multiple of the denominators of $a_0, \ldots, a_{n-1}$ (when we write them as fractions in lowest terms). Then

$$g(X) = X^n + mc_{n-1}X^{n-1} + m^2 c_{n-2}X^{n-2} + \cdots + m^{n-1}c_1 X + m^n c_0$$

is a monic polynomial with coefficients in $\mathbb{Z}$. We have $g(m\alpha) = 0$, and so $m\alpha$ is an algebraic integer. (In fact g is the minimal polynomial of $m\alpha$ but we don't need to know this to prove the lemma.)

Note that $m \neq 0$, so we can write $\alpha = (m\alpha)/\alpha$ as required. □

Being able to write elements of K as fractions with a rational integer as denominator is convenient because it makes it easier to find common denominators, or to multiply up by a rational number and reduce to a question about elements of $\mathcal{O}_K$.

Lemma 25. *If $\alpha \in \mathcal{O}_K$, then $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$ and $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha)$ are rational integers.*

Proof. Let $\chi_{K,\alpha}$ be the characteristic polynomial of $m_{K,\alpha}$. By Lemmas 15 and 16, $\chi_{K,\alpha}$ is a power of $\mu_{\mathbb{Q},\alpha}$ so $\chi_{K,\alpha}$ has integer coefficients.

$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$ and $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha)$ are coefficients of $\chi_{K,\alpha}$ (multiplied by ± 1) so they are in $\mathbb{Z}$. □

Lemma 25 is useful for checking that certain numbers are *not* algebraic integers. We can't use it to prove that a number *is* an algebraic integer, because the converse is false – except when K is a quadratic field.

Lemma 26. *Let $K =$ be a quadratic field. If $\alpha \in K$ satisfies $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}$ and $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}$, then $\alpha \in \mathcal{O}_K$.*

Proof. The characteristic polynomial of $\alpha \in K$ is

$$\chi_{K,\alpha}(X) = X^2 - \mathrm{Tr}_{K/\mathbb{Q}}(\alpha) \cdot X + \mathrm{Nm}_{K/\mathbb{Q}}(\alpha).$$

So if $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}$ and $\mathrm{Tr}_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}$, then $\chi_{K,\alpha}$ is a monic polynomial with rational integer coefficients which vanishes at α. (This relies on the fact that

$\deg(\chi_{K,\alpha}) = [K:\mathbb{Q}] = 2$ so $\chi_{K,\alpha}$ has no other coefficients in between the norm and trace.) Thus α is an algebraic integer. □

Ring of integers of a quadratic field.

Proposition 27. *Let $d \neq 1$ be a square-free integer and let $K = \mathbb{Q}(\sqrt{d})$. Then*

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

(Note that $d \not\equiv 0 \bmod 4$ because it is square-free.)

Proof. Any element of $\mathbb{Q}(\sqrt{d})$ can be written as $\alpha = a + b\sqrt{d}$ with $a, b \in \mathbb{Q}$. By Lemma 25, if $\alpha \in \mathcal{O}_K$ then

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = a^2 - db^2 \in \mathbb{Z}, \quad \mathrm{Tr}_{K/\mathbb{Q}}(\alpha) = 2a \in \mathbb{Z}.$$

We deduce that $4db^2 = (2b)^2 d \in \mathbb{Z}$. Since d is square-free, this implies that $2b \in \mathbb{Z}$.

So if $\alpha = a + b\sqrt{d} \in \mathcal{O}_K$, then $a, b \in \mathbb{Z}$ or $\mathbb{Z} + \frac{1}{2}$. In other words

$$\alpha = \beta + \gamma \text{ where } \beta \in \mathbb{Z}[\sqrt{d}], \quad \gamma \in \Big\{0, \frac{1}{2}, \frac{\sqrt{d}}{2}, \frac{1+\sqrt{d}}{2}\Big\}$$

for some $\beta \in \mathbb{Z}[\sqrt{d}]$. Now $\mathbb{Z}[\sqrt{d}] \subseteq \mathcal{O}_K$, so (using the fact that $\mathcal{O}_K$ is closed under addition and subtraction), $\alpha \in \mathcal{O}_K$ if and only if $\gamma \in \mathcal{O}_K$.

Now $\frac{1}{2}$ is not an algebraic integer, so $\beta + \frac{1}{2}$ is never in $\mathcal{O}_K$.

Likewise, $\frac{\sqrt{d}}{2}$ has minimal polynomial $X^2 - d/4$ (or norm $-d/4$) so $\frac{\sqrt{d}}{2}$ is not an algebraic integer and $\beta + \frac{\sqrt{d}}{2}$ is never in $\mathcal{O}_K$.

This leaves us to check $\frac{1+\sqrt{d}}{2}$. The minimal polynomial of $\frac{1+\sqrt{d}}{2}$ is $X^2 - X + \frac{1-d}{4}$. So $\frac{1+\sqrt{d}}{2}$ is an algebraic integer if and only if $\frac{1-d}{4} \in \mathbb{Z}$, or equivalently, if and only if $d \equiv 1 \bmod 4$.

Thus we have shown that if $d \equiv 2$ or $3 \bmod 4$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{d}]$.

If $d \equiv 1 \bmod 4$, we have shown that

$$\mathcal{O}_K = \{a + b\sqrt{d} + c\tfrac{1+\sqrt{d}}{2} : a, b \in \mathbb{Z}, c = 0 \text{ or } 1\}. \qquad (*)$$

This contains $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ because $\mathcal{O}_K$ is a ring but we still have to check that $\mathcal{O}_K \subseteq \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$.

Thanks to (*), it suffices to show that $\sqrt{d} \in \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$. This is true because

$$\sqrt{d} = 2\Big(\tfrac{1+\sqrt{d}}{2}\Big) - 1. \qquad \square$$

e.g. Since $-1 \equiv 3 \bmod 4$,$\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$ i.e. the Gaussian integers,

Since $-3 \equiv 1 \bmod 4$, we have

$$\mathcal{O}_{\mathbb{Q}(\sqrt{-3})} = \mathbb{Z}[\tfrac{1+\sqrt{-3}}{2}] = \mathbb{Z}[\zeta_3]$$

(where $\zeta_3 = \frac{-1+\sqrt{-3}}{2}$ is a primitive cube root of unity).

10. Discriminant of a basis

We were able to calculate the ring of integers of a quadratic field directly from the definition, but this is not easy for number fields of degree greater than 2. Instead we will rely on a tool called the discriminant. The discriminant will also be an important tool in many of the calculations that will come up later in the course, such as class groups, as well as being an important theoretical tool in some of the proofs.

The discriminant is a number which measures the size of the ring of integers of a number field (in a more refined sense than the degree). We begin by defining the discriminant of a *basis* of a number field, which varies depending on the basis we choose; we will subsequently pick a special kind of basis and use that to define the discriminant of the number field itself.

There are two equivalent formulae for the discriminant of a basis. Here is the first.

Definition. Let K be a number field with $n = [K : \mathbb{Q}]$. Let $\sigma_1, \ldots, \sigma_n$ be the embeddings of K and let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbb{Q}$-basis for K.

The **discriminant** of $\{\alpha_1, \ldots, \alpha_n\}$ is defined to be

$$\Delta_K(\alpha_1, \ldots, \alpha_n) = \det(\sigma_i(\alpha_j))^2$$

(usually we will just write $\Delta(\alpha_1, \ldots, \alpha_n)$ with no K). In other words, $\Delta(\alpha_1, \ldots, \alpha_n)$ is the square of the determinant of the matrix with entries $\sigma_i(\alpha_j)$ i.e.

$$\begin{pmatrix} \sigma_1(\alpha_1) & \sigma_1(\alpha_2) & \cdots & \sigma_1(\alpha_n) \\ \sigma_2(\alpha_1) & \sigma_2(\alpha_2) & \cdots & \sigma_2(\alpha_n) \\ \vdots & \vdots & & \vdots \\ \sigma_n(\alpha_1) & \sigma_n(\alpha_2) & \cdots & \sigma_n(\alpha_n) \end{pmatrix}.$$

Note that there are $[K : \mathbb{Q}]$ embeddings by the Corollary to Proposition 13, so this is indeed a square matrix and its determinant makes sense.

Note also that $\det(M^2) = (\det(M))^2$ so it does not matter whether we square the matrix then take the determinant, or take the determinant then square (the latter is much easier computationally!)

Why is this a measure of "size"? Suppose that K has signature $(n, 0)$ (all its embeddings are real). You might find it helpful to think of real quadratic fields in particular. Then $(\sigma_1, \ldots, \sigma_n)$ is a $\mathbb{Q}$-linear map $K \to \mathbb{R}^n$. It maps our basis elements to the vectors $(\sigma_1(\alpha_i), \sigma_2(\alpha_i), \ldots, \sigma_n(\alpha_i)) \in \mathbb{R}^n$ (i.e. the columns of the matrix above). The volume of the parallelepiped in $\mathbb{R}^n$ with edges $v_1, \ldots, v_n$ is given by the determinant of the matrix which has $v_1, \ldots, v_n$ as columns (up to sign). So $\Delta(\alpha_1, \ldots, \alpha_n)$ is the square of the volume of the parallelepiped formed from these vectors. Thus in some way it measures the "volume" of the basis.

When the field has complex embeddings as well as real, there is not quite such a clear geometric picture, but we can still generalise this to think of the discriminant as a volume.

Why do we need to square the determinant in order to define the discriminant? Here are several reasons:

(1) If we swap two of the embeddings, or two of the α_i, then that swaps two of the rows or columns of the matrix, so it multiplies the determinant by ± 1. Thus squaring gives us a value which is independent of the orderings.
(2) $\Delta(\alpha_1, \ldots, \alpha_n) \in \mathbb{Q}$ but $\det(\sigma_i(\alpha_j))$ need not be in $\mathbb{Q}$. This is not obvious: the entries of the matrix are algebraic numbers so all we can immediately see is that $\Delta(\alpha_1, \ldots, \alpha_n)$ is an algebraic number. In order to prove that $\Delta(\alpha_1, \ldots, \alpha_n) \in \mathbb{Q}$, we shall use the second definition of discriminant below. (It is also possible to use Galois theory to prove this directly from the first definition.)
(3) We have to square the determinant to match the second formula (which we are about to give)!

Example: if $K = \mathbb{Q}(\sqrt{d})$, we can calculate

$$\Delta(1, \sqrt{d}) = \det \begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{pmatrix}^2 = (-2\sqrt{d})^2 = 4d.$$

Thanks to Proposition 27, if $d \equiv 1 \bmod 4$, then it is often useful to use the basis $\{1, \frac{1+\sqrt{d}}{2}\}$ (it generates the ring of integers). Its discriminant is

$$\Delta(1, \tfrac{1+\sqrt{d}}{2}) = \det \begin{pmatrix} 1 & \frac{1+\sqrt{d}}{2} \\ 1 & \frac{1-\sqrt{d}}{2} \end{pmatrix}^2 = (-\sqrt{d})^2 = d.$$

Note that these are equal to the discriminants of the quadratic polynomials $X^2 - d$ and $X^2 - X + \frac{1-d}{4}$ (that is, the minimal polynomials of $\sqrt{d}$ and $\frac{1+\sqrt{d}}{2}$ respectively).

Second formula for the discriminant.

Let K be a number field of degree n. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbb{Q}$-basis for K. Then

$$\Delta(\alpha_1, \ldots, \alpha_n) = \det(\mathrm{Tr}_{K/\mathbb{Q}}(\alpha_i \alpha_j)).$$

Note that the determinant is not squared this time! (This expression is already "quadratic" because it involves multiplying together two αs – this can be helpful for remembering which formula is squared and which is not.) This time the matrix is symmetric.

We can check that the second formula gives the same values when applied to the bases of a quadratic field which we considered previously:

$$\Delta(1, \sqrt{d}) = \det \begin{pmatrix} \mathrm{Tr}(1) & \mathrm{Tr}(\sqrt{d}) \\ \mathrm{Tr}(\sqrt{d}) & \mathrm{Tr}(d) \end{pmatrix} = \det \begin{pmatrix} 2 & 0 \\ 0 & 2d \end{pmatrix} = 4d,$$

$$\Delta(1, \tfrac{1+\sqrt{d}}{2}) = \det \begin{pmatrix} \mathrm{Tr}(1) & \mathrm{Tr}(\frac{1+\sqrt{d}}{2}) \\ \mathrm{Tr}(\frac{1+\sqrt{d}}{2}) & \mathrm{Tr}(\frac{1+d+2\sqrt{d}}{4}) \end{pmatrix} = \det \begin{pmatrix} 2 & 1 \\ 1 & \frac{1+d}{2} \end{pmatrix} = d.$$

The matrix $(\mathrm{Tr}_{K/\mathbb{Q}}(\alpha_i \alpha_j))$ has entries in $\mathbb{Q}$, so the second formula tells us that $\Delta(\alpha_1, \ldots, \alpha_n)$ is always in $\mathbb{Q}$.

The second formula is usually better for calculating the discriminant because you begin by working out the traces and then you just have to calculate the determinant of a matrix with rational entries, whereas the first formula requires

you to calculate the discriminant of a matrix whose entries are algebraic numbers – usually a more difficult calculation. The first definition of discriminant will more often be useful in proofs where you don't have to calculate a specific example.

Proof that the two formulae for discriminant are equivalent.
Let M be the matrix with entries $\sigma_i(\alpha_j)$. Then

$$\Delta(\alpha_1, \ldots, \alpha_n) = \det(M)^2 = \det(M^t)\det(M) = \det(M^t M).$$

The ij-th entry of $M^t M$ is

$$\sum_{k=1}^{n} M^t_{ik} M_{kj} = \sum_{k=1}^{n} M_{ki} M_{kj} = \sum_{k=1}^{n} \sigma_k(\alpha_i)\sigma_k(\alpha_j) = \sum_{k=1}^{n} \sigma_k(\alpha_i\alpha_j) = \mathrm{Tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j)$$

where the penultimate inequality uses the fact that σ_k is a field homomorphism, and the last equality is Lemma 18. □

Discriminants and change of basis.

The discriminant depends on the choice of basis of K. It is important to understand how the discriminant changes when you change the basis: it gets multiplied by the square of the determinant of the change-of-basis matrix.

Lemma 28. *Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be $\mathbb{Q}$-bases for K. Let the change-of-basis matrix from $\{\beta_1, \ldots, \beta_n\}$ to $\{\alpha_1, \ldots, \alpha_n\}$ be C i.e.*

$$\beta_j = \sum_{j=1}^{n} C_{ij}\alpha_i$$

with $C_{ij} \in \mathbb{Q}$. Then

$$\Delta(\beta_1, \ldots, \beta_n) = \det(C)^2\, \Delta(\alpha_1, \ldots, \alpha_n).$$

Proof. We can prove this using either formula for the discriminant. The first definition is slightly easier.

Let A and B be the matrices with entries $\sigma_i(\alpha_j)$ and $\sigma_i(\beta_j)$ respectively. Let C be the matrix with entries C_{ij}. Then

$$B_{ij} = \sigma_i(\beta_j) = \sigma_i\Big(\sum_{k=1}^{n} C_{kj}\alpha_k\Big) = \sum_{k=1}^{n} C_{kj}\sigma_i(\alpha_k) = \sum_{k=1}^{n} A_{ik}C_{kj}$$

(using the facts that σ_i is a field homomorphism, and that it restricts to the identity on $\mathbb{Q}$). Hence $B = AC$ as matrices and so

$$\Delta(\beta_1, \ldots, \beta_n) = \det(B)^2 = \Big(\det(A)\det(C)\Big)^2 = \det(C)^2\, \Delta(\alpha_1, \ldots, \alpha_n). \qquad \square$$

11. Discriminants and integral bases

11.1. **Properties of the discriminant.** Last time we showed that the discriminants of two different bases for the same number field are related by the square of the determinant of the change-of-basis matrix:

$$\Delta(\beta_1, \ldots, \beta_n) = \det(C)^2 \, \Delta(\alpha_1, \ldots, \alpha_n)$$

where $\beta_j = \sum_{j=1}^{n} C_{ij}\alpha_i$ and $C_{ij} \in \mathbb{Q}$. The following lemma gives us a way of working out $|\det(C)|$ which is often very useful in conjunction with Lemma 28. (We will want to apply this with $G = \mathcal{O}_K$, $H =$ some subgroup of $\mathcal{O}_K$.)

Lemma 29. *Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be $\mathbb{Q}$-bases for K. Let $C_{ij} \in \mathbb{Q}$ be the entries of the change of basis matrix:*

$$\beta_j = \sum_{i=1}^{n} C_{ij}\alpha_i.$$

Let $G = \mathbb{Z}.\alpha_1 + \cdots + \mathbb{Z}.\alpha_n$ and $H = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n \subseteq K$.
Suppose that $H \subseteq G$. (In other words, $\beta_1, \ldots, \beta_n \in G$.)
Then H has finite index in G and $[G : H] = |\det(C)|$.

It is unfortunate that the notation $[G : H]$ for the index of a subgroup clashes with the notation $[L : K]$ for the degree of a field extension. Both are very standard notations. We will never want to talk about the index of a subgroup of a field (in characteristic zero, the index of one field as a subgroup of another is always infinite) so hopefully this will not cause confusion.

We omit the proof of Lemma 29. You can find it in the lecture notes from last year if you are interested. It is very closely related to the results in Algebra I about the structure of finitely generated abelian groups, with the key idea in the proof being Smith Normal Form.

Combining Lemmas 28 and 29 we get the following.

Corollary 30. *Let $G = \mathbb{Z}.\alpha_1 + \cdots + \mathbb{Z}.\alpha_n$ and $H = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$ with $H \subseteq G$ as in Lemma 29. Then*

$$\Delta(\beta_1, \ldots, \beta_n) = [G : H]^2 \, \Delta(\alpha_1, \ldots, \alpha_n)$$

(I stated Corollary 30 incorrectly in the lecture, with the αs and βs the wrong way round.)

We can also use Lemma 28 to prove another fundamental property of the discriminant.

Lemma 31. *For any $\mathbb{Q}$-basis $\{\alpha_1, \ldots, \alpha_n\}$ of K, the discriminant $\Delta(\alpha_1, \ldots, \alpha_n)$ is non-zero.*

In order to prove Lemma 31, we will use the Primitive Element Theorem. This was mentioned as a non-examinable aside back in lecture 4 – but I forgot then that I would need to use it now. So now we formally state the Primitive Element Theorem and declare the statement (but not the proof) to be examinable.

Theorem (Primitive Element Theorem)**.** *Let K be a number field. There exists $\alpha \in K$ such that $K = \mathbb{Q}(\alpha)$.*

Proof of Lemma 31. We will prove this first for a special choice of basis. By the Primitive Element Theorem, $K = \mathbb{Q}(\alpha)$ for some $\alpha \in K$. By Theorem 3, there is a $\mathbb{Q}$-basis for K of the form $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Now $\Delta(1, \alpha, \ldots, \alpha^{n-1})$ is the square of

$$\det \begin{pmatrix} 1 & \sigma_1(\alpha) & \cdots & \sigma_1(\alpha^{n-1}) \\ 1 & \sigma_2(\alpha) & \cdots & \sigma_2(\alpha^{n-1}) \\ \vdots & \vdots & & \vdots \\ 1 & \sigma_n(\alpha) & \cdots & \sigma_n(\alpha^{n-1}) \end{pmatrix} = \det \begin{pmatrix} 1 & \sigma_1(\alpha) & \cdots & \sigma_1(\alpha)^{n-1} \\ 1 & \sigma_2(\alpha) & \cdots & \sigma_2(\alpha)^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \sigma_n(\alpha) & \cdots & \sigma_n(\alpha)^{n-1} \end{pmatrix}.$$

This is a special kind of matrix called a Vandermonde matrix and it is well-known that its determinant is

$$\prod_{1 \le i < j \le n} (\sigma_i(\alpha) - \sigma_j(\alpha)).$$

Thus

$$\Delta(1, \alpha, \ldots, \alpha^{n-1}) = \prod_{1 \le i < j \le n} (\sigma_i(\alpha) - \sigma_j(\alpha))^2.$$

By Lemmas 11 and 12, the $\sigma_i(\alpha)$ are pairwise distinct so each factor $\sigma_i(\alpha) - \sigma_j(\alpha)$ is non-zero. Hence the product is non-zero.

Now consider an arbitrary $\mathbb{Q}$-basis $\beta_1, \ldots, \beta_n$ of K. The change-of-basis matrix from $1, \alpha, \ldots, \alpha^{n-1}$ to $\beta_1, \ldots, \beta_n$ has non-zero determinant, so Lemma 28 implies that $\Delta(\beta_1, \ldots, \beta_n) \neq 0$. □

Note that we only defined the discriminant for a basis of K, but we could apply the same formulae to any set of n elements $\{\alpha_1, \ldots, \alpha_n\} \subseteq K$. (The proof that the two formulae give the same value still works.) In fact, $\{\alpha_1, \ldots, \alpha_n\}$ forms a basis of K if and only if $\Delta(\alpha_1, \ldots, \alpha_n) \neq 0$.

Integral bases.

Definition. Let K be a number field. An **integral basis** for K is a set of elements $\alpha_1, \ldots, \alpha_m \in \mathcal{O}_K$ such that

(a) $\mathcal{O}_K = \mathbb{Z}.\alpha_1 + \cdots + \mathbb{Z}.\alpha_m$; and
(b) $\alpha_1, \ldots, \alpha_m$ are $\mathbb{Z}$-linearly independent i.e. if $c_1\alpha_1 + \cdots + c_m\alpha_m = 0$ with $c_1, \ldots, c_m \in \mathbb{Z}$, then $c_1 = \cdots = c_m = 0$.

This is like the definition for basis of a vector space (spans and linearly independent), but with the base field replaced by $\mathbb{Z}$.

Lemma 32. *Let $\{\alpha_1, \ldots, \alpha_m\}$ be an integral basis for a number field K. Then $\{\alpha_1, \ldots, \alpha_m\}$ is a basis for K as a $\mathbb{Q}$-vector space. In particular, $m = [K : \mathbb{Q}]$.*

Proof. $\mathbb{Z}$-linear independence is equivalent to $\mathbb{Q}$-linear independence: if $\{\alpha_1, \ldots, \alpha_m\}$ were $\mathbb{Q}$-linearly dependent, then we could multiply up by a common denominator for the coefficients to get a non-trivial $\mathbb{Z}$-linear relation between them.

Lemma 24 tells us that, if $\beta \in K$, then $\beta = \gamma/d$ with $\gamma \in \mathcal{O}_K$ and $d \in \mathbb{Z}$. From the definition of integral basis, we can write $\gamma = c_1\alpha_1 + \cdots + c_m\alpha_m$ with $c_1, \ldots, c_m \in \mathbb{Z}$. Then

$$\beta = \frac{c_1}{d}\alpha_1 + \cdots + \frac{c_m}{d}\alpha_m$$

so $\{\alpha_1, \ldots, \alpha_m\}$ spans K as a $\mathbb{Q}$-vector space. $\square$

For example, Proposition 27 gives us an integral basis for $\mathbb{Q}(\sqrt{d})$ where d is a square-free integer:

- $\{1, \sqrt{d}\}$ if $d \equiv 2, 3 \bmod 4$;
- $\{1, \frac{1+\sqrt{d}}{2}\}$ if $d \equiv 1 \bmod 4$.

Note that if $\{\alpha_1, \ldots, \alpha_n\}$ is a $\mathbb{Q}$-basis for K and $\alpha_1, \ldots, \alpha_n \in \mathcal{O}_K$, this is *not enough* to establish that $\{\alpha_1, \ldots, \alpha_n\}$ is an integral basis. For example, if $d \equiv 1 \bmod 4$, then $\{1, \sqrt{d}\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{d})$ consisting of algebraic integers but it is not an integral basis because $\frac{1+\sqrt{d}}{2} \in \mathcal{O}_K$ but not in $\mathbb{Z}.1 + \mathbb{Z}.\sqrt{d}$. (If we have a $\mathbb{Q}$-basis for K consisting of algebraic integers, then the "uniquely" part of the definition of integral basis is always satisfied. But, as here, the basis might fail to generate $\mathcal{O}_K$ over $\mathbb{Z}$.)

It is not obvious that an integral basis exists for every number field. Before we prove this, let's pause to reflect on why it is not obvious – this is just some remarks, not examinable. The structure theory of finitely generated abelian groups (from Algebra 1) tells us that every torsion-free finitely generated abelian group is isomorphic to $\mathbb{Z}^n$ for some n and hence possesses a $\mathbb{Z}$-basis. The group $(\mathcal{O}_K, +)$ is torsion-free (because number fields have characteristic 0). However it is not obvious that $(\mathcal{O}_K, +)$ is finitely generated – this is true, but we will only discover it as a corollary of the existence of an integral basis.

$\mathbb{Z}[\frac{1}{2}] \subseteq \mathbb{Q}$ is an example of a subring of a number field which is not finitely generated as an abelian group and so does not have an integral basis, demonstrating that we are really going to have to use some properties of algebraic integers to show that $(\mathcal{O}_K, +)$ is finitely generated.

One key property of algebraic integers was Lemma 22: if α is an algebraic integer, then $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group. However we can't apply this to $\mathcal{O}_K$ because $\mathcal{O}_K$ need not be of the form $\mathbb{Z}[\alpha]$ for any α (there is no analogue for the Primitive Element Theorem for rings of integers). Instead we have to work much harder, making use of the discriminant and Corollary 30.

12. Existence of integral bases

Theorem 33. *Every number field K possesses an integral basis.*

Proof. First note that there exists a $\mathbb{Q}$-basis $\{\beta_1, \ldots, \beta_n\}$ of K consisting of algebraic integers. Indeed, if we take any $\mathbb{Q}$-basis of K, then by Lemma 24, we can multiply each of the basis elements by a non-zero rational integer to obtain something in $\mathcal{O}_K$.

Since each $\beta_i \in \mathcal{O}_K$, the traces $\mathrm{Tr}_{K/\mathbb{Q}}(\beta_i\beta_j)$ are in $\mathbb{Z}$ and so $\Delta(\beta_1, \ldots, \beta_n) \in \mathbb{Z}$. Furthermore, $\Delta(\beta_1, \ldots, \beta_n) \neq 0$ by Lemma 31. So $|\Delta(\beta_1, \ldots, \beta_n)|$ is a positive integer.

Hence there is a *minimum* value for $|\Delta(\beta_1, \ldots, \beta_n)|$ among all $\mathbb{Q}$-bases of K with elements in $\mathcal{O}_K$.

Choose $\{\beta_1, \ldots, \beta_n\}$ to be a basis which achieves this minimum value for $|\Delta(\beta_1, \ldots, \beta_n)|$. We will show that this basis is an integral basis.

Suppose not (for contradiction).

Since $\beta_1, \ldots, \beta_n$ are linearly independent over $\mathbb{Q}$, they are also linearly independent over $\mathbb{Z}$. Hence in order to not be an integral basis, we must have

$$\mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n \neq \mathcal{O}_K.$$

Thus there is some $\gamma \in \mathcal{O}_K$ such that $\gamma \notin \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$.

Let

$$H = \mathbb{Z}.\beta_1 + \mathbb{Z}.\beta_2 + \cdots + \mathbb{Z}.\beta_n,$$
$$G = \mathbb{Z}.\beta_1 + \mathbb{Z}.\beta_2 + \cdots + \mathbb{Z}.\beta_n + \mathbb{Z}.\gamma.$$

By definition, G is a finitely generated abelian group. Furthermore G is torsion-free (i.e. if $\alpha \in G$ and $m \in \mathbb{Z}$ with $\alpha \neq 0, m \neq 0$ then $m\alpha \neq 0$). Consequently by the structure theory of finitely generated abelian groups (from Algebra 1), G is isomorphic to $\mathbb{Z}^m$ for some m and hence has a $\mathbb{Z}$-basis $\{\alpha_1, \ldots, \alpha_m\}$.

Now $\alpha_1, \ldots, \alpha_m$ span K as a $\mathbb{Q}$-vector space because they generate $\beta_1, \ldots, \beta_n$, and they are $\mathbb{Q}$-linearly independent because they are $\mathbb{Z}$-linearly independent. Hence $\{\alpha_1, \ldots, \alpha_m\}$ is a $\mathbb{Q}$-basis for K and $m = n$. (This is the same argument as in the proof of Lemma 32.)

By Corollary 30, we have

$$\Delta(\beta_1, \ldots, \beta_n) = [G : H]^2 \Delta(\alpha_1, \ldots, \alpha_n).$$

Since $\gamma \in G \setminus H$, $[G : H] > 1$. Also $|\Delta(\beta_1, \ldots, \beta_n)| > 0$ so we get

$$|\Delta(\beta_1, \ldots, \beta_n)| > |\Delta(\alpha_1, \ldots, \alpha_n)|.$$

But $\alpha_1, \ldots, \alpha_n \in G \subseteq \mathcal{O}_K$

Thus $\{\alpha_1, \ldots, \alpha_n\}$ is a $\mathbb{Q}$-basis for K, contained in $\mathcal{O}_K$, with strictly smaller $|\Delta|$ than $\{\beta_1, \ldots, \beta_n\}$. This contradicts our choice of $\beta_1, \ldots, \beta_n$ with $|\Delta(\beta_1, \ldots, \beta_n)|$ as small as possible. □

Discriminant of an integral basis.

An important observation is that all integral bases for a given number field have the same discriminant.

Lemma 34. *Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be integral bases for K. Then*

$$\Delta(\alpha_1, \ldots, \alpha_n) = \Delta(\beta_1, \ldots, \beta_n).$$

Proof. From the definition of integral basis, we have

$$\mathcal{O}_K = \mathbb{Z}.\alpha_1 + \cdots + \mathbb{Z}.\alpha_n = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n.$$

Hence by Corollary 30,

$$\Delta(\beta_1, \ldots, \beta_n) = [\mathcal{O}_K : \mathcal{O}_K]^2 \, \Delta(\alpha_1, \ldots, \alpha_n).$$

We are done because $[\mathcal{O}_K : \mathcal{O}_K] = 1$. □

Consequently the following definition makes sense.

Definition. Let K be a number field. The **discriminant** of K, written Δ_K, is the discriminant of any integral basis of K.

The discriminant of K is always a non-zero integer.

e.g. according to calculations from lecture 10, the discriminant of $\mathbb{Q}(\sqrt{d})$ (for $d \neq 1$ a square-free integer) is as follows:

$$\Delta_{\mathbb{Q}(\sqrt{d})} = \Delta(1, \sqrt{d}) = 4d \qquad \text{if } d \equiv 2, 3 \bmod 4,$$

$$\Delta_{\mathbb{Q}(\sqrt{d})} = \Delta(1, \tfrac{1+\sqrt{d}}{2}) = d \qquad \text{if } d \equiv 1 \bmod 4.$$

Finding an integral basis.

Corollary 30 implies the following easy sufficient criterion for recognising an integral basis.

Lemma 35. *Let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbb{Q}$-basis for K such that $\alpha_1, \ldots, \alpha_n \in \mathcal{O}_K$. If $\Delta(\alpha_1, \ldots, \alpha_n)$ is square-free, then $\{\alpha_1, \ldots, \alpha_n\}$ is an integral basis for K.*

Proof. Let $H = \mathbb{Z}.\alpha_1 + \cdots + \mathbb{Z}.\alpha_n$. Let $\{\beta_1, \ldots, \beta_n\}$ be an integral basis for K (which exists by Theorem 33). Then $H \subseteq \mathcal{O}_K = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$ so by Corollary 30,

$$\Delta(\alpha_1, \ldots, \alpha_n) = [\mathcal{O}_K : H]^2 \, \Delta(\beta_1, \ldots, \beta_n).$$

(Sorry I have used αs and βs the opposite way round to Corollary 30!)

Here $[\mathcal{O}_K : H]^2$ is a square and $\Delta(\beta_1, \ldots, \beta_n)$ is an integer, while $\Delta(\alpha_1, \ldots, \alpha_n)$ is square-free, so $[\mathcal{O}_K : H] = 1$. Thus $H = \mathcal{O}_K$, and so $\alpha_1, \ldots, \alpha_n$ form an integral basis. □

This is only a one-way implication. For example, we saw that if d is square-free and congruent to 2 or 3 mod 4, then $\{1, \sqrt{d}\}$ is an integral basis for $\mathbb{Q}(\sqrt{d})$ but $\Delta(1, \sqrt{d}) = 4d$ is not square-free (because it is divisible by 4). Lemma 35 is only useful when Δ_K is itself square-free, which often doesn't hold.

In general, we can give an algorithm based on the proof of Theorem 33 to find an integral basis (and hence calculate the ring of integers and the discriminant of the number field).

Start with some $\mathbb{Q}$-basis $\{\alpha_1, \ldots, \alpha_n\}$ for K consisting of algebraic integers. Suppose $\Delta(\beta_1, \ldots, \beta_n)$ is not square-free (so we cannot apply Lemma 35).

We want to either find $\gamma \in \mathcal{O}_K$ which is not in $\mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$, or prove there is no such γ.

The following lemma gives us a way to do this by checking only finitely many potential γs.

Lemma 36. *Let $\{\beta_1, \ldots, \beta_n\}$ be a $\mathbb{Q}$-basis for K with $\beta_1, \ldots, \beta_n \in \mathcal{O}_K$. If $\{\beta_1, \ldots, \beta_n\}$ is not an integral basis, then there exists a prime p such that*

(1) $p^2 \mid \Delta(\alpha_1, \ldots, \alpha_n)$; and

(2) there exist $u_1, \ldots, u_n \in \mathbb{Z}$, not all zero, satisfying $0 \leq u_i < p$ for all i and

$$\frac{u_1\alpha_1 + \cdots + u_n\alpha_n}{p} \in \mathcal{O}_K.$$

(Proof next time.)

13. Finding an integral basis

We begin by proving the lemma from the end of the last lecture.

Lemma 36. *Let $\{\beta_1, \ldots, \beta_n\}$ be a $\mathbb{Q}$-basis for K with $\beta_1, \ldots, \beta_n \in \mathcal{O}_K$. If $\{\beta_1, \ldots, \beta_n\}$ is not an integral basis, then there exists a prime p such that*

(1) $p^2 \mid \Delta(\alpha_1, \ldots, \alpha_n)$; and

(2) there exist $u_1, \ldots, u_n \in \mathbb{Z}$, not all zero, satisfying $0 \leq u_i < p$ for all i and

$$\frac{u_1\alpha_1 + \cdots + u_n\alpha_n}{p} \in \mathcal{O}_K.$$

Proof. Let $H = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$. Now $H \subseteq \mathcal{O}_K$ but $H \neq \mathcal{O}_K$ so $[\mathcal{O}_K : H] \neq 1$, and we can pick a prime p which divides $[\mathcal{O}_K : H]$. By Corollary 30,

$$\Delta(\alpha_1, \ldots, \alpha_n) = [\mathcal{O}_K : H]^2 \, \Delta_K$$

and so $p^2 \mid \Delta(\alpha_1, \ldots, \alpha_n)$.

Now $\mathcal{O}_K/H$ is a finite abelian group whose order is divisible by p. Cauchy's theorem on finite groups states that $\mathcal{O}_K/H$ contains an element of order p. (Cauchy's theorem is a general theorem about finite groups; in the case of abelian groups, it can easily be deduced from the structure theory of finite abelian groups.) Thus we can choose a coset $\gamma + H \in \mathcal{O}_K/H$ which has order p.

Here $\gamma \in \mathcal{O}_K$, while the fact that $p(\gamma + H) = 0 + H$ tells us that $p\gamma \in H$ so

$$p\gamma = x_1\alpha_1 + \cdots + x_n\alpha_n \text{ for some } x_1, \ldots, x_n \in \mathbb{Z}.$$

Also $\gamma + H \neq 0 + H$, that is, $\gamma \notin H$, so the x_i are not all multiples of p.

In order to get the right range, let u_i be the remainder of x_i divided by p, that is,

$$x_i = u_i + py_i \text{ where } u_i, y_i \in \mathbb{Z} \text{ and } 0 \leq u_i < p.$$

Since x_i are not all multiples of p, the u_i are not all zero. finally

$$\frac{u_1\beta_1 + \cdots + u_n\beta_n}{p} = \gamma - (y_1\beta_1 + \cdots + y_n\beta_n) \in \mathcal{O}_K \qquad \square$$

Algorithm to find an integral basis.

(1) Pick a $\mathbb{Q}$-basis $\{\beta_1, \ldots, \beta_n\}$ for K, such that $\beta_1, \ldots, \beta_n \in \mathcal{O}_K$.

(2) Calculate $\Delta(\beta_1, \ldots, \beta_n)$.

(3) List all primes p such that $p^2 \mid \Delta(\beta_1, \ldots, \beta_n)$.

(4) For each p in the list, look at all numbers of the form

$$\gamma = \frac{u_1\beta_1 + \cdots + u_n\beta_n}{p}$$

with $u_i \in \mathbb{Z}$, $0 \leq u_i < p$ not all zero. Check whether γ is an algebraic integer.

(5) If some γ is an algebraic integer, then find a $\mathbb{Z}$-basis for the subgroup G of $\mathcal{O}_K$ generated by $\beta_1, \ldots, \beta_n$ and γ. (Usually, it will be possible to replace one of the β_i by γ and get a $\mathbb{Z}$-basis. In order to check that $\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_n, \gamma$ is a $\mathbb{Z}$-basis for G, it is enough to check that you can write β_i as a $\mathbb{Z}$-linear combination of the other βs and γ.)

Go back to step 2 of the algorithm with this new basis. (Actually, you can skip step 2 because $[G : H] = p$ so $\Delta(\text{new basis}) = \Delta(\beta_1, \ldots, \beta_n)/p^2$.)

(6) If you did not find any γ in step 4 which was an algebraic integer, then you have found an integral basis (thanks to Lemma 36).

This algorithm is guaranteed to terminate because $|\Delta(\beta_1, \ldots, \beta_1)|$ gets smaller each time round.

Shortcut using Eisenstein's criterion.

Step 4 of the algorithm above can certainly be implemented on a computer, but it is a lot of work to carry it out by hand. There is a shortcut which is often useful, if K is generated by an element whose minimal polynomial satisfies Eisenstein's criterion. Recall the statement of Eisenstein's criterion form Algebra 2.

Definition. Let $f(X) = a_n X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0$ be a polynomial in $\mathbb{Z}[X]$. We say that f satisfies **Eisenstein's criterion at a prime** p if $p \nmid a_n$, $p \mid a_i$ for $0 \leq i \leq n-1$ and $p^2 \nmid a_0$.

Proposition 37. *Let $K = \mathbb{Q}(\alpha)$ where the minimal polynomial of α satisfies Eisenstein's criterion at p. Let $n = [K : \mathbb{Q}]$. Then:*

(i) p^{n-1} divides Δ_K.

(ii) $\frac{u_1 + u_2\alpha + u_3\alpha^2 + \cdots + u_n\alpha^{n-1}}{p}$, for $u_i \in \mathbb{Z}$, $0 \leq u_i < p$ and u_i not all zero, is never an algebraic integer.

(We are looking at the basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$ for K.)

This proposition can be proved using the methods of this course but it is a bit long so we will skip the proof (for part (ii), we already have the tools to prove it – and it appeared as "unseen" material in an exam question last year; for part (i), the easiest proof uses the Dedekind–Kummer theorem which we will study later). You need to know the statement of the proposition and be able to use it to replace step 4 of the algorithm.

One example in which this shortcut is useful is cyclotomic fields. Let $\zeta = \exp(2\pi i/p)$ where p is an odd prime number and let $K = \mathbb{Q}(\zeta)$. According to example sheet 1, $\{1, \zeta, \ldots, \zeta^{p-1}\}$ is a $\mathbb{Q}$-basis for K and as in example sheet 2 Q5(iv) we can calculate

$$\Delta(1, \zeta, \ldots, \zeta^{p-2}) = (-1)^{(p-1)/2} p^{p-2}.$$

Let $\omega = \zeta - 1$. (We choose this value because we know its minimal polynomial satisfies Eisenstein's criterion at p.) Now

$$\mathbb{Z}.1 + \mathbb{Z}.\omega + \cdots + \mathbb{Z}.\omega^{p-2} = \mathbb{Z}.1 + \mathbb{Z}.\zeta + \cdots + \mathbb{Z}.\zeta^{p-2}$$

so

$$\Delta(1, \omega, \ldots, \omega^{p-2}) = \Delta(1, \zeta, \ldots, \zeta^{p-2}) = (-1)^{(p-1)/2} p^{p-2}.$$

The only prime factor o f$\Delta(1, \omega, \ldots, \omega^{p-2}$ is p and so by Lemma 36, in order to find $\mathcal{O}_K$ we only need to check whether

$$\frac{u_1 + u_2\omega + \cdots + u_{p-1}\omega^{p-2}}{p}$$

is an algebraic integer for $u_1, \ldots, u_{p-1} \in \mathbb{Z}$, not all zero, $0 \leq u_i < p$. By Proposition 37, this never happens. So $\{1, \omega, \ldots, \omega^{p-2}\}$ is an integral basis for K.

Since $\mathbb{Z}.1 + \mathbb{Z}.\omega + \cdots + \mathbb{Z}.\omega^{p-2} = \mathbb{Z}.1 + \mathbb{Z}.\zeta + \cdots + \mathbb{Z}.\zeta^{p-2}$, it follows that $\{1, \zeta, \ldots, \zeta^{p-2}\}$ is also an integral basis for K, and

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

Example sheet 2 Q5(v) and (vi) give most of what you need to prove Proposition 37 in this special case. The general case is just a little more complicated – look at last year's exam for an outline of the proof!

Another example: $K = \mathbb{Q}(\alpha)$ with $\alpha = \sqrt[3]{17}$. Then $\{1, \alpha, \alpha^2\}$ is a $\mathbb{Q}$-basis for K consisting of algebraic integers. Using example sheet 2, we can calculate

$$\Delta(1, \alpha, \alpha^2) = -3^3 \times 17^2.$$

The primes whose squares divide this are: 3 and 17.

The minimal polynomial of α is $X^3 - 17$ and this is Eisenstein at 17, so we don't need to check 17.

At 3, we find that

$$\gamma = \frac{1 + 2\alpha + \alpha^2}{3}$$

is an algebraic integer. (If I got the calculation right, the minimal polynomial of γ is $X^3 - X^2 - 11X - 12$.)

Now

$$\alpha^2 = 3\gamma - 1 - 2\alpha \in \mathbb{Z}.1 + \mathbb{Z}.\alpha + \mathbb{Z}.\gamma$$

so $\{1, \alpha, \gamma\}$ is a $\mathbb{Z}$-basis for $\mathbb{Z}.1 + \mathbb{Z}.\alpha + \mathbb{Z}.\alpha^2 + \mathbb{Z}.\gamma$.

Now go back to step 2 with the basis $\{1, \alpha, \gamma\}$. It has discriminant

$$\Delta(1, \alpha, \gamma) = (-3^3 \times 17^2)/3^2 = -3 \times 17^2.$$

This is not divisible by 3^2, and we already know that we don't have to check 17.

Thus $\{1, \alpha, \gamma\}$ is an integral basis for K.

This ends the discussion about discriminants and integral bases. Next lecture we will look at factorisation and ideals, starting with some more Algebra 2 revision.

14. Factorisation in integral domains

We are going to talk about factorisation in the ring of integers of a number field. We begin by recalling several definitions from Algebra 2.

Throughout this lecture, R will denote an integral domain. (An **integral domain** is a non-zero ring in which, if $xy = 0$, then $x = 0$ or $y = 0$.)

Units.

Definition. "$x \mid y$" ("x **divides** y in R") means that there exists $z \in R$ such that $y = xz$.

Definition. An element $x \in R$ is a **unit** if there exists $y \in R$ such that $xy = 1$. (In other words, $x \mid 1$.)

The set of units in R forms an abelian group under multiplication. We write $R^\times$ for this group. (That's a "times" symbol in the superscript, because it's a group under multiplication. Some people call this group R^* with an asterisk.)

The following lemmas are useful for working with units in the ring of integers of a number field.

Lemma 38. *Let $\mathcal{O}_K$ be the ring of integers of a number field. For every $\alpha \in \mathcal{O}_K$, $\alpha \mid \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$ in $\mathcal{O}_K$.*

Proof. If $\alpha = 0$ then $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = 0$ and it's obvious.

Otherwise, consider $\beta = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)/\alpha$. Since K is a field, we have $\beta \in K$. We want to show that $\beta \in \mathcal{O}_K$.

Let $\sigma_1, \ldots, \sigma_n \colon K \to \mathbb{C}$ denote the embeddings of $\mathbb{C}$. Now

$$\sigma_1(\beta) = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)/\sigma_1(\alpha) = \sigma_2(\alpha)\sigma_3(\alpha)\cdots\sigma_n(\alpha)$$

by Lemma 18.

Since $\sigma_2(\alpha), \ldots, \sigma_n(\alpha)$ each have the same minimal polynomial (over $\mathbb{Q}$) as α, they are all algebraic integers. Hence their product, namely $\sigma_1(\beta)$, is an algebraic integer.

Now $\sigma_1(\beta)$ has the same minimal polynomial as β, so β is an algebraic integer. Since $\beta \in K$, we conclude that $\beta \in \mathcal{O}_K$. □

Corollary 39. *Let $\mathcal{O}_K$ be the ring of integers of a number field. An element $\alpha \in \mathcal{O}_K$ is a unit if and only if $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \pm 1$.*

Proof. If x is a unit, then x, x^{-1} are both in $\mathcal{O}_K$ so $\mathrm{Nm}_{K/\mathbb{Q}}(x)$ and $\mathrm{Nm}_{K/\mathbb{Q}}(x^{-1})$ are both rational integers. Since

$$\mathrm{Nm}_{K/\mathbb{Q}}(x)\,\mathrm{Nm}_{K/\mathbb{Q}}(x^{-1}) = \mathrm{Nm}_{K/\mathbb{Q}}(xx^{-1}) = 1$$

we conclude that $\mathrm{Nm}_{K/\mathbb{Q}}(x) = \pm 1$.

Conversely, by Lemma 38, if $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \pm 1$, then $\alpha \mid 1$ in $\mathcal{O}_K$ so α is a unit. □

Factorisation.

Definition. Elements $x, y \in R$ are **associates** if there exists a unit $z \in R^\times$ such that $x = yz$.

Definition. An element $x \in R$ is:

- **irreducible** if it is non-zero, not a unit and whenever we can write $x = ab$ with $a, b \in R$, then either a is a unit or b is a unit;
- **prime** if it is non-zero, not a unit and whenever $x \mid ab$ with $a, b \in R$, either $x \mid a$ or $x \mid b$.

Definition. An integral domain R is a **unique factorisation domain (UFD)** if, for every non-zero non-unit $a \in R$:

(i) a can be written in the form $a = x_1 x_2 \cdots x_n$ for some irreducible elements $x_1, \ldots, x_n \in R$;
(ii) given another factorisation $a = y_1 y_2 \cdots y_m$ into irreducibles, we must have $m = n$ and after permuting $y_1, \ldots, y_m$, each y_i is an associate of the corresponding x_i.

Definition. An integral domain R is a **principal ideal domain (PID)** if every ideal in R is of the form $\langle a \rangle$ for some $a \in R$.

The following facts were proved in Algebra 2.

Facts. *In any integral domain, every prime element is irreducible.*
In a UFD, every irreducible element is prime.
Every PID is a UFD.

There are UFDs which are not PIDs, for example $\mathbb{Z}[X]$. However we will prove later that if the *ring of integers of a number field* is a UFD, then it is a PID.

A classic example of an integral domain which is not a UFD is the ring of integers of $K = \mathbb{Q}(\sqrt{-5})$. The ring of integers is $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$. We have two factorisations of 6 in $\mathbb{Z}[\sqrt{-5}]$:

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

We can prove that $2, 3, 1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are all irreducible by considering their norms.

For example, $\mathrm{Nm}_{K/\mathbb{Q}}(1 + \sqrt{-5}) = 6$ so if $1 + \sqrt{-5} = ab$ with $a, b \in \mathbb{Z}[\sqrt{-5}]$, then either a, b have norms $\pm 1, \pm 6$ or $\pm 2, \pm 3$ (in some order). The equations

$$\mathrm{Nm}_{K/\mathbb{Q}}(x + y\sqrt{-5}) = x^2 + 5y^2 = \pm 2$$

have no solutions in rational integers x, y, so $\mathbb{Z}[\sqrt{-5}]$ contains no elements of norm ± 2. Hence the norms of a and b must be ± 1, ± 6.

If $\mathrm{Nm}_{K/\mathbb{Q}}(a) = \pm 1$, then by Corollary 39, a is a unit in $\mathbb{Z}[\sqrt{-5}]$. Thus in any factorisation $1 + \sqrt{-5} = ab$, either a or b is a unit so $1 + \sqrt{-5}$ is irreducible.

Similar arguments show that $1 - \sqrt{-5}$, 2, 3 are irreducible in $\mathbb{Z}[\sqrt{-5}]$.

To show that our two factorisations of 6 in $\mathbb{Z}[\sqrt{-5}]$ are not just associates of each other, observe that

$$\mathrm{Nm}_{K/\mathbb{Q}}(1+\sqrt{-5}) = \mathrm{Nm}_{K/\mathbb{Q}}(1-\sqrt{-5}) = 6$$

while $\mathrm{Nm}_{K/\mathbb{Q}}(2) = 4$, so (thanks to Corollary 39) any associate of 2 has norm ± 4. Thus 2 is not an associate of either $1+\sqrt{-5}$ or $1-\sqrt{-5}$, so we have written down two genuinely different factorisations of 6.

Thus $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Looking at the factorisation above, we see that $2, 3, 1+\sqrt{-5}$ and $1-\sqrt{-5}$ are irreducible but not prime in $\mathbb{Z}[\sqrt{-5}]$.

Since $\mathbb{Z}[\sqrt{-5}]$ is not a UFD, it cannot be a PID. An example of a non-principal ideal in $\mathbb{Z}[\sqrt{-5}]$ is $I = \langle 2, 1+\sqrt{-5}\rangle$. Indeed, if $I = \langle a\rangle$ then a divides both 2 and $1+\sqrt{-5}$. Since 2 and $1+\sqrt{-5}$ are irreducible but not associates of each other, the only elements which divide both of them are units. If $I = \langle a\rangle$ and a is a unit, then $1 \in I$. However, one can check that every element of $I = \langle 2, 1+\sqrt{-5}\rangle$ has the form $x + y\sqrt{-5}$ with $x \equiv y \bmod 2$ so $1 + 0\sqrt{-5} \notin I$.

15. Norm of an ideal

A note on fonts: in Algebraic Number Theory, it is traditional to denote ideals by Fraktur letters such as $\mathfrak{a}, \mathfrak{b}, \mathfrak{p}, \mathfrak{q}$. These are difficult to write clearly on the board, so my letters on the board denoting ideals will look rather different.

Norm of an ideal.

We will start to talk about special properties of ideals in the ring of integers of a number field.

(There are no Lemmas 40 and 41 because I skipped them out from the numbering in the lecture.)

Lemma 42. *Let K be a number field. Let $\mathfrak{a}$ be a non-zero ideal in $\mathcal{O}_K$. Then $\mathcal{O}_K/\mathfrak{a}$ is finite.*

Proof. Pick any $\alpha \in \mathfrak{a} \setminus \{0\}$ and let $N = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$. Then N is a non-zero integer because α is a non-zero algebraic integer. Furthermore by Lemma 38, we have $\alpha \mid N$ so $N \in \mathfrak{a}$. Then $\langle N \rangle \subseteq \mathfrak{a}$ and so $\mathcal{O}_K/\langle N \rangle$ surjects onto $\mathcal{O}_K/\mathfrak{a}$.

Because of the existence of an integral basis, $\mathcal{O}_K$ is isomorphic as an abelian group to $\mathbb{Z}^n$. Hence $\mathcal{O}_K/\langle N \rangle$ is isomorphic as an abelian group to $\mathbb{Z}/N\mathbb{Z}^n$ and this is finite. This implies that $\mathcal{O}_K/\mathfrak{a}$ is finite. □

Lemma 42 is a very special property for subrings of a number field – very few other integral domains have this property. (One example which does is $F[X]$ where F is a finite field. It turns out that you can do a lot of things very similar to Algebraic Number Theory in $F[X]$ or its finite extensions – called *function field arithmetic*.)

Thanks to Lemma 42, the following definition makes sense.

Definition. Let K be a number field and let $\mathfrak{a}$ be a non-zero ideal in $\mathcal{O}_K$. The **norm** of $\mathfrak{a}$, written $\mathrm{Nm}(\mathfrak{a})$, is defined to be $\#(\mathcal{O}_K/\mathfrak{a})$ (or in other words the index $[\mathcal{O}_K : \mathfrak{a}]$).

Note that $\mathrm{Nm}(\mathfrak{a})$ is always a positive integer – you can think of it as a measure of the size of an ideal.

We have defined the "norm of an element of K" and the "norm of an ideal in $\mathcal{O}_K$." The definitions look very different, but they are compatible in the case of principal ideals, as the following lemma shows (except that the norm of an ideal is always positive, while the norm of an element may be positive or negative, so we need to take the absolute value of the latter).

Lemma 43. *Let K be a number field and let $\alpha \in \mathcal{O}_K \setminus \{0\}$. Then*

$$\mathrm{Nm}(\langle \alpha \rangle) = |\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)|.$$

Proof. Choose an integral basis $\beta_1, \ldots, \beta_n$ for K. Let C be the matrix (with entries in $\mathbb{Q}$) representing "multiplication by α" with respect to this basis. Thus

$$\alpha\beta_j = \sum_{i=1}^{n} C_{ij}\beta_i.$$

Now $\alpha\beta_1, \ldots, \alpha\beta_n$ form a $\mathbb{Z}$-basis for the ideal $\langle\alpha\rangle$ and C is the the change-of-basis matrix from $\{\alpha\beta_1, \ldots, \alpha\beta_n\}$ to $\{\beta_1, \ldots, \beta_n\}$. By Lemma 29 (applied to $\mathcal{O}_K = \mathbb{Z}.\beta_1 + \cdots + \mathbb{Z}.\beta_n$ and $\mathfrak{a} = \mathbb{Z}.\alpha\beta_1 + \cdots + \mathbb{Z}.\alpha\beta_n$) we have

$$|\det(C)| = [\mathcal{O}_K : \langle\alpha\rangle] = \mathrm{Nm}(\mathfrak{a}).$$

Meanwhile the definition of norm of an element says that

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \det(C). \qquad \square$$

Lemma 44. *Let $\mathfrak{a}, \mathfrak{b}$ be non-zero ideals in $\mathcal{O}_K$. If $\mathfrak{a} \subseteq \mathfrak{b}$ and $\mathrm{Nm}(\mathfrak{a}) = \mathrm{Nm}(\mathfrak{b})$, then $\mathfrak{a} = \mathfrak{b}$.*

Proof. Since $\mathfrak{a} \subseteq \mathfrak{b}$, we have

$$\mathrm{Nm}(\mathfrak{a}) = [\mathcal{O}_K : \mathfrak{a}] = [\mathcal{O}_K : \mathfrak{b}][\mathfrak{b} : \mathfrak{a}] = \mathrm{Nm}(\mathfrak{b})[\mathfrak{b} : \mathfrak{a}].$$

Hence if $\mathrm{Nm}(\mathfrak{a}) = \mathrm{Nm}(\mathfrak{b})$, then $[\mathfrak{b} : \mathfrak{a}] = 1$ or in other words $\mathfrak{a} = \mathfrak{b}$. $\square$

Lemmas 43 and 44 are both useful on their own. They are also very useful when combined together (taking $\mathfrak{a} = \langle\alpha\rangle$), to get the following consequence:

If $\alpha \in \mathfrak{b}$ and $|\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| = \mathrm{Nm}(\mathfrak{b})$, then $\mathfrak{b} = \langle\alpha\rangle$.

This is valuable as a method of proving that an ideal is principal: you just have to find an element in the ideal $\mathfrak{b}$ which has norm equal to $\pm\mathrm{Nm}(\mathfrak{b})$.

Lemma 45. *Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be a non-zero ideal. Then $\mathrm{Nm}(\mathfrak{a}) \in \mathfrak{a}$.*

Proof. Lagrange's theorem for the finite group $(\mathcal{O}_K/\mathfrak{a}, +)$ tells us that

$$\#(\mathcal{O}_K/\mathfrak{a}) \cdot (1 + \mathfrak{a}) = 0 + \mathfrak{a} \text{ in } \mathcal{O}_K/\mathfrak{a}.$$

Unwrapping this statement about cosets, we get $\mathrm{Nm}(\mathfrak{a}) = \#(\mathcal{O}_K/\mathfrak{a}) \in \mathfrak{a}$. $\square$

Product of ideals.

We want to talk about "factorisation of ideals" in $\mathcal{O}_K$. First we need to define the product of ideals.

Let R be a ring and let $\mathfrak{a}, \mathfrak{b}$ be ideals in R. The set $\{ab : a \in \mathfrak{a}, b \in \mathfrak{b}\}$ is not necessarily an ideal because it might not be closed under addition (to find an example of this, both $\mathfrak{a}$ and $\mathfrak{b}$ need to be non-principal). Instead we define

$$\mathfrak{ab} = \{a_1b_1 + \cdots + a_mb_m : m \in \mathbb{N},\ a_1, \ldots, a_m \in \mathfrak{a},\ b_1, \ldots, b_m \in \mathfrak{b}\}.$$

This is an ideal in R.

If $\mathfrak{a} = \langle a_1, \ldots, a_r\rangle$ and $\mathfrak{b} = \langle b_1, \ldots, b_s\rangle$, then

$$\mathfrak{ab} = \langle a_ib_j : 1 \le i \le r,\ 1 \le j \le s\rangle.$$

e.g. We will calculate the square of the ideal $\mathfrak{a} = \langle 2, 1 + \sqrt{-5}\rangle$ in $\mathbb{Z}[\sqrt{-5}]$.

$$\begin{aligned}\mathfrak{a}^2 = \langle 2, 1 + \sqrt{-5}\rangle^2 &= \langle 2 \cdot 2,\ 2(1 + \sqrt{-5}),\ (1 + \sqrt{-5})(1 + \sqrt{-5})\rangle \\ &= \langle 4,\ 2 + 2\sqrt{-5},\ -4 + 2\sqrt{-5}\rangle.\end{aligned}$$

We can simplify this: $(2 + 2\sqrt{-5}) + (-4 + 2\sqrt{-5}) = -2 \in \mathfrak{a}^2$ so $2 \in \mathfrak{a}^2$ and hence $\langle 2\rangle \subseteq \mathfrak{a}^2$. Meanwhile 2 divides all of 4, $2 + 2\sqrt{-5}$, $-4 + 2\sqrt{-5}$ so $\mathfrak{a}^2 \subseteq \langle 2\rangle$.

Thus $\mathfrak{a}^2 = \langle 2\rangle$.

This example shows that the square of a non-principal ideal can be principal. Of course a product of ideals isn't always principal, but you can often simplify it to a smaller number of generators than you initially write down (in the ring of integers of a number field, every ideal can be generated by at most 2 elements).

Prime ideals.

Now we define prime ideals, the analogue for ideals of prime elements. I don't think this was in Algebra 2.

To motivate the definition, think about rewriting the definition of a "prime element" $p \in R$ in terms of the principal ideal $\langle p \rangle$: $p \in R$ (not 0 or a unit) is prime if and only if

$$\text{for all } x, y \in R, \text{ if } xy \in \langle p \rangle, \text{ then } x \in \langle p \rangle \text{ or } y \in \langle p \rangle.$$

We generalise this, replacing $\langle p \rangle$ by any ideal.

Definition. Let R be a ring. An ideal $\mathfrak{p} \subseteq R$ is **prime** if $\mathfrak{p} \neq R$ and for all $x, y \in R$, if $xy \in \mathfrak{p}$, then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$.

Observe that for $p \neq 0$, the principal ideal $\langle p \rangle$ is prime if and only if p is a prime element. (It's a historical quirk that the element 0 is not prime, but the ideal $\langle 0 \rangle$ can be prime – in fact $\langle 0 \rangle$ is prime if and only if R is an integral domain.)

There is an alternative equivalent definition for prime ideals, which looks even more like the definition of prime elements. Before stating this, we define what it means for one ideal to divide another.

Whenever we form a product of ideals $\mathfrak{ab}$, we have $\mathfrak{ab} \subseteq \mathfrak{a}$. Thus multiplying ideals makes them smaller as sets, so the following definition is reasonable.

Definition. Let $\mathfrak{a}, \mathfrak{b} \subseteq R$ be ideals. We say that $\mathfrak{a}$ **divides** $\mathfrak{b}$ (written $\mathfrak{a} \mid \mathfrak{b}$) if $\mathfrak{b} \subseteq \mathfrak{a}$.

As further justification for this definition, consider principal ideals. For any $\alpha, \beta \in R$, we have

$$\alpha \mid \beta \quad \Leftrightarrow \quad \langle \beta \rangle \subseteq \langle \alpha \rangle \quad \Leftrightarrow \quad \langle \alpha \rangle \mid \langle \beta \rangle.$$

Note that, in a general ring R, it need not be true that if $\mathfrak{a} \mid \mathfrak{b}$ then there exists an ideal $\mathfrak{c}$ satisying $\mathfrak{b} = \mathfrak{ac}$. (Eventually we will show that this does hold in the ring of integers of a number field.)

16. Prime and maximal ideals

Using the notion of divisibility of ideals which we defined last time, we give the second equivalent definition of prime ideals – which is just the definition of prime elements, replacing elements by ideals everywhere.

Lemma 46. *Let R be any ring and let $\mathfrak{p}$ be a ideal in R, not equal to R. Then $\mathfrak{p}$ is a prime ideal if and only if, for all ideals $\mathfrak{a}, \mathfrak{b} \subseteq R$, whenever $\mathfrak{p} \mid \mathfrak{ab}$, then $\mathfrak{p} \mid \mathfrak{a}$ or $\mathfrak{p} \mid \mathfrak{b}$.*

We omit the proof because it is pure algebra.

Maximal ideals.

Maximal ideals are another useful type of ideal, closely related to prime ideals. A maximal ideal is one of the largest possible ideals in the ring R (excluding R itself).

Definition. Let R be a ring. An ideal $\mathfrak{a} \subseteq R$ is **maximal** if $\mathfrak{a} \neq R$ and there is no ideal $\mathfrak{b}$ satisfying $\mathfrak{a} \subsetneq \mathfrak{b} \subsetneq R$.

Prime and maximal ideals are closely related to properties of the quotient ring.

Lemma 47. *Let R be a ring and let $\mathfrak{a} \subseteq R$ be an ideal.*

(i) $\mathfrak{a}$ is a prime ideal if and only if $R/\mathfrak{a}$ is an integral domain.
(ii) $\mathfrak{a}$ is a maximal ideal if and only if $R/\mathfrak{a}$ is a field.

Again we omit the proof because it is pure algebra.

Corollary. *In any ring R, every maximal ideal is prime.*

The converse of the corollary is false in general, but almost true for the ring of integers of a number field (we just have to leave out zero). In order to prove this, we first need a lemma about integral domains.

Lemma 48. *A finite integral domain is a field.*

Proof. (This proof is pure algebra, but it is fundamental to the properties of $\mathcal{O}_K$, so it is examinable.)

Let R be a finite integral domain and let $x \in R \setminus \{0\}$. The map $m_x \colon R \to R$ given by $m_x(y) = xy$ is injective because R is an integral domain. Since R is finite, this implies that m_x is a bijection $R \to R$. Thus there exists $y \in R$ such that $m_x(y) = 1$. In other words, y is a multiplicative inverse for x. □

Corollary 49. *In the ring of integers of a number field, every non-zero prime ideal is maximal.*

Proof. Let $\mathfrak{a}$ be a non-zero prime ideal in $\mathcal{O}_K$. By Lemma 42, $\mathcal{O}_K/\mathfrak{a}$ is finite. Since $\mathfrak{a}$ is a prime ideal, $\mathcal{O}_K/\mathfrak{a}$ is an integral domain. Hence by Lemma 48, $\mathcal{O}_K/\mathfrak{a}$ is a field and so $\mathfrak{a}$ is a maximal ideal. □

Corollary 49 is not quite as special as Lemma 42, even though we used the latter in the proof. For example one-variable polynomial rings $K[X]$ satisfy Corollary 49 where K is any field, but not the two-variable polynomial rings $K[X, Y]$. From the perspective of Algebraic Geometry, this corollary can be interpreted as saying that "$\mathcal{O}_K$ is a one-dimensional geometric object" (in a very abstract sense).

Unique factorisation of ideals.

We have seen that the ring of integers of a number field is not always a UFD. One of the central results of this module is that it does have unique factorisation *of ideals into prime ideals.*

Theorem 50. *Let $\mathcal{O}_K$ be the ring of integers of a number field and let $\mathfrak{a} \subseteq \mathcal{O}_K$ be an ideal, not equal to $\langle 0 \rangle$ or $\mathcal{O}_K$. Then:*

(i) there exist prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r$;

(ii) if we have another list of prime ideals $\mathfrak{q}_1, \ldots, \mathfrak{q}_s$ such that $\mathfrak{a} = \mathfrak{q}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s$, then $r = s$ and $\mathfrak{q}_1, \ldots, \mathfrak{q}_s$ are a permutation of $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$.

Note that the uniqueness condition is simpler than for a UFD: because we are talking about ideals instead of elements, we don't need to mention associates (if x, y are associates, then they generate the same ideal).

We will prove this theorem next week. First, we want to explore some more concrete things we can do with factorisation of ideals in $\mathcal{O}_K$.

Prime ideals of a number field.

We want to describe the prime ideals in $\mathcal{O}_K$.

Definition. We shall use the phrase **rational prime** to mean a prime in $\mathbb{Z}$ (just the usual meaning of prime number) – similar to how we sometimes say "rational integers" to avoid confusion with algebraic integers. The purpose of this is to avoid any possible confusion with "prime ideals in $\mathcal{O}_K$" (or even "prime elements in $\mathcal{O}_K$"). Sorry if it causes more confusion!

Proposition 51. *Let K be a number field. Let $\mathfrak{p}$ be a non-zero prime ideal in $\mathcal{O}_K$. Then $\mathrm{Nm}(\mathfrak{p}) = p^n$ for some rational prime p and some positive integer n. Furthermore, $\mathfrak{p} \mid \langle p \rangle$ and $\mathfrak{p} \nmid \langle q \rangle$ for any rational prime $q \neq p$.*

Proof. By Corollary 49, $\mathfrak{p}$ is a maximal ideal so $\mathcal{O}_K/\mathfrak{p}$ is a field. Furthermore $\mathcal{O}_K/\mathfrak{p}$ is finite by Lemma 42. By a result from Algebra 2, the order of any finite field is a prime power. So

$$\mathrm{Nm}(\mathfrak{p}) = \#\mathcal{O}_K/\mathfrak{p} = p^n$$

for some p and n.

By Lemma 45, we deduce that $p^n \in \mathfrak{p}$. If $n > 1$, we write $p^n = pp^{n-1}$ and use the definition of prime ideal to deduce that either $p \in \mathfrak{a}$ or $p^{n-1} \in \mathfrak{p}$. If $p^{n-1} \in \mathfrak{p}$, then we can repeat the process; eventually we conclude that $p \in \mathfrak{p}$ or in other words $\mathfrak{p} \mid \langle p \rangle$.

Finally we want to show that there is no other rational prime $q \neq p$ such that $\mathfrak{p} \mid \langle q \rangle$. Assume for contradiction that such a prime exists. Then $q \in \mathfrak{p}$. By Euclid's algorithm, we can find $x, y \in \mathbb{Z}$ such that $xp + yq = 1$. Since $p, q \in \mathfrak{p}$, we deduce that $1 \in \mathfrak{p}$ and so $\mathfrak{p} = \mathcal{O}_K$. But this contradicts the fact that $\mathfrak{p}$ is a prime ideal. □

This lemma tells us that, in order to list all prime ideals in $\mathcal{O}_K$, it suffices to go through the rational primes and, for each rational prime p, determine the prime ideals of $\mathcal{O}_K$ which divide $\langle p \rangle$.

Using unique factorisation of ideals, we can write $\langle p \rangle$ as a product of prime ideals in $\mathcal{O}_K$

$$\langle p \rangle = \mathfrak{p}_1^{e_1} \mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_r^{e_r}.$$

The $\mathfrak{p}_i$ which appear in this factorisation are the only prime ideals of $\mathcal{O}_K$ which divide $\langle p \rangle$. Thus, for each rational prime p, we get a finite list of prime ideals of $\mathcal{O}_K$ which divide $\langle p \rangle$.

For example, in the Gaussian integers $\mathbb{Z}[i]$, each rational prime p has either one or two prime ideal factors in $\mathbb{Z}[i]$, depending on the value of $p \bmod 4$.

In the next lecture, we will state a theorem telling us how $\langle p \rangle$ factorises into prime ideals in $\mathcal{O}_K$, and look at a couple of examples of applying it.

17. Dedekind–Kummer theorem

Dedekind–Kummer theorem: statement.

By Proposition 51, each non-zero prime ideal of $\mathcal{O}_K$ divides exactly one rational prime. Thus in order to determine the prime ideals of $\mathcal{O}_K$, we just have to determine the prime ideals which divide each rational prime.

The Dedekind–Kummer theorem tells us how to find these ideals. The statement of the theorem may look rather long, but it gives a very clear recipe which we can apply in practice.

Notation. If p is a rational prime, we write $\mathbb{F}_p$ for "the field with p elements" i.e. $\mathbb{Z}/p\mathbb{Z}$. We use this notation to emphasise that it is a field (of course $\mathbb{Z}/p\mathbb{Z}$ is only a field when p is prime).

Theorem (Dedekind–Kummer)**.** *Let $K = \mathbb{Q}(\alpha)$ be a number field where α is an algebraic integer. Let p be a rational prime which does not divide $[\mathcal{O}_K : \mathbb{Z}[\alpha]]$.*

Let $f(X) \in \mathbb{Z}[X]$ be the minimal polynomial of α, and let $\bar{f}(X) \in \mathbb{F}_p[X]$ denote the reduction of f modulo p. Let the factorisation of $\bar{f}$ into monic irreducible polynomials be

$$\bar{f} = \bar{f}_1^{e_1} \bar{f}_2^{e_2} \cdots \bar{f}_r^{e_r}$$

where $\bar{f}_1(X), \ldots, \bar{f}_r(X) \in \mathbb{F}_p[X]$ are pairwise distinct.

For each i, choose a polynomial $f_i(X) \in \mathbb{Z}[X]$ such that $\bar{f}_i = f_i$ modulo p.

Let $\mathfrak{p}_i$ denote the ideal $\langle p, f_i(\alpha)\rangle$ in $\mathcal{O}_K$.

Then:

(i) the $\mathfrak{p}_i$ are distinct prime ideals of $\mathcal{O}_K$;
(ii) $\langle p\rangle = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$;
(iii) $\mathrm{Nm}(\mathfrak{p}_i) = p^{\deg(\bar{f}_i)}$.

Most of the words of this theorem are just carefully defining notation for the factorisation of f mod p.

There is one condition which it is important not to forget: $p \nmid [\mathcal{O}_K : \mathbb{Z}[\alpha]]$. We want to apply the Dedekind–Kummer theorem with $\mathcal{O}_K = \mathbb{Z}[\alpha]$ whenever possible, because then $[\mathcal{O}_K : \mathbb{Z}[\alpha]] = 1$ and so the condition $p \nmid [\mathcal{O}_K : \mathbb{Z}[\alpha]]$ is satisfied for every prime p.

However, it is not always possible to choose α such that $\mathcal{O}_K = Z[\alpha]$ and then we have to exclude the finitely many primes which divide $[\mathcal{O}_K : \mathbb{Z}[\alpha]]$ when using the Dedekind–Kummer theorem.

The proof involves a lot of ring homomorphisms and applications of the Third Isomorphism Theorem for rings (it's in last year's notes). The central step is showing that there are ring isomorphisms

$$\mathcal{O}_K/\langle p\rangle \leftarrow \mathbb{Z}[X]/\langle p, f(X)\rangle \rightarrow \mathbb{F}_p[X]/\langle \bar{f}\rangle,$$

so the prime ideals of the left and right rings are in bijection with each other. This bijection matches $\langle p, f_i(\alpha)\rangle$ with $\langle \bar{f}_i\rangle$. Checking that the exponents in the prime factorisation match up requires some calculations with norms.

Example of the Dedekind–Kummer theorem.

$K = \mathbb{Q}(\sqrt{-10})$

Since $-10 \equiv 2 \bmod 4$, we have $\mathcal{O}_K = \mathbb{Z}[\sqrt{-10}]$. so we can apply Dedekind–Kummer with $\alpha = \sqrt{-10}$ for every rational prime p. The minimal polynomial of $\sqrt{-10}$ is $f(X) = X^2 + 10$.

- $p = 2$: $f(X) \equiv X^2 \bmod 2$.
  In the notation of the Dedekind–Kummer theorem, we have $r = 1$, $f_1(X) = X$, $e_1 = 2$.
  So the only prime ideal of $\mathcal{O}_K$ dividing $\langle 2\rangle$ is $\mathfrak{p}_1 = \langle 2, f_1(\alpha)\rangle = \langle 2, \sqrt{-10}\rangle$.
  By (iii), $\mathrm{Nm}(\langle 2, \sqrt{-10}\rangle) = 2^{\deg(X)} = 2^1 = 2$.
  By (iv), $\langle 2\rangle = \langle 2, \sqrt{-10}\rangle^2$.

- $p = 3$: $f(X) \equiv X^2 - 2 \bmod 3$. This is irreducible (because it is quadratic, it suffices to check that it has no roots; it has no roots because 2 is not a quadratic residue mod 3).
  Thus in the notation of the Dedekind–Kummer theorem, we have $r = 1$, $f_1(X) = X^2 - 2$, $e_1 = 1$.
  Hence the only prime ideal of $\mathcal{O}_K$ dividing $\langle 3\rangle$ is $\langle 3, \alpha^2 - 2\rangle = \langle 3, -12\rangle = \langle 3\rangle$.
  In fact, we could have deduced this without any calculations using (iv): since $r = e_1 = 1$, (iv) tells us that $\langle p\rangle = \mathfrak{p}_1^{e_1} = \mathfrak{p}_1$.
  This gives us a general conclusion (valid for any p and K):
  If $\bar{f}(X)$ is irreducible in $\mathbb{F}_p[X]$, then $\langle p\rangle$ is a prime ideal of $\mathcal{O}_K$.

- $p = 5$: $\langle 5\rangle = \langle 5, \sqrt{-10}\rangle^2$ (similar to $p = 2$).

- $p = 7$: $f(X) \equiv X^2 - 4 \equiv (X + 2)(X - 2) \bmod 7$.
  In the notation of the Dedekind–Kummer theorem, we have $r = 2$, $f_1(X) = X - 2$, $f_2(X) = X + 2$, $e_1 = e_2 = 1$.
  Since $f(X)$ has two distinct irreducible factors, there are two prime ideals of $\mathcal{O}_K$ which divide $\langle 7\rangle$, namely

  $$\langle 7, f_1(\alpha)\rangle = \langle 7, 2 + \sqrt{-10}\rangle \text{ and } \langle 7, f_2(\alpha)\rangle = \langle 7, -2 + \sqrt{-10}\rangle.$$

  By (iv),

  $$\langle 7\rangle = \langle 7, 2 + \sqrt{-10}\rangle \langle 7, -2 + \sqrt{-10}\rangle.$$

  (You can check this product by hand!)

Another example of Dedekind–Kummer.

$K = \mathbb{Q}(\sqrt{-7})$

Since $-7 \equiv 1 \bmod 4$, we have $\mathcal{O}_K = \mathbb{Z}[\alpha]$ where $\alpha = \frac{1+\sqrt{-7}}{2}$. The minimal polynomial of α is $f(X) = X^2 - X + 2$.

For $p = 2$, we have $f(X) \equiv X^2 - X \equiv X(X - 1) \bmod 2$.

Thus $\langle 2\rangle = \mathfrak{p}\mathfrak{q}$ where $\mathfrak{p} = \langle 2, \alpha\rangle = \langle 2, \frac{1+\sqrt{-7}}{2}\rangle$ and $\mathfrak{q} = \langle 2, \alpha - 1\rangle = \langle 2, \frac{-1+\sqrt{-7}}{2}\rangle$.

In fact, $\mathrm{Nm}(\mathfrak{p}) = 2^{\deg(X)} = 2$ while $\mathrm{Nm}_{K/\mathbb{Q}}(\frac{1+\sqrt{-7}}{2}) = (\frac{1}{2})^2 + 7(\frac{1}{2})^2 = 2$ so

$$\mathfrak{p} = \Big\langle \frac{1+\sqrt{-7}}{2} \Big\rangle.$$

Similarly, $\mathfrak{q} = \langle \frac{-1+\sqrt{-7}}{2} \rangle$.

The Dedekind–Kummer theorem tells us that the ideals $\mathfrak{p}$ and $\mathfrak{q}$ are distinct. We could also see this directly because if $\mathfrak{p} = \mathfrak{q}$, then this ideal would contain $\alpha - (\alpha - 1) = 1$, contradicting the fact that it must be a proper ideal of $\mathcal{O}_K$.

We would have got the wrong answer if we tried to use $\alpha = \sqrt{-7}$ instead of $\frac{1+\sqrt{-7}}{2}$! The Dedekind–Kummer theorem is not valid for $p = 2$ and $\alpha = \sqrt{-7}$, because

$$[\mathcal{O}_K : \mathbb{Z}[\sqrt{-7}]] = 2,$$

so this is not a contradiction. Rather it is a warning that the condition $p \nmid [\mathcal{O}_K : \mathbb{Z}[\sqrt{-7}]]$ is important (and that making sure you use the correct ring of integers for a quadratic field is also important).

Indeed, the minimal polynomial of $\sqrt{-7}$ is $X^2 + 7 \equiv (X+1)^2 \bmod 2$, so if we (incorrectly) used Dedekind–Kummer for $\sqrt{-7}$ we would conclude that $\langle 2 \rangle$ is the square of a prime ideal, but we saw that in fact $\mathfrak{p}$ and $\mathfrak{q}$ are *distinct* prime ideals.

Note that, if $p \neq 2$, then $p \nmid [\mathcal{O}_K : \mathbb{Z}[\sqrt{17}]]$ so we can apply Dedekind–Kummer either for $\alpha = \frac{1+\sqrt{-7}}{2}$ or for $\alpha = \sqrt{-7}$ and both will give the right answer.

18. Fractional ideals

Ideal norms are multiplicative.

We state the following theorem now, even though we won't be able to prove it until after unique factorisation of ideals, because it is needed for some of the example sheet questions (but not the assessed ones I think).

Lemma 53. *Let $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}_K$ be non-zero ideals. Then*

$$\mathrm{Nm}(\mathfrak{ab}) = \mathrm{Nm}(\mathfrak{a})\,\mathrm{Nm}(\mathfrak{b}).$$

Fractional ideals.

In the next couple of lectures, we will prove the unique factorisation of ideals. This is the longest examinable proof in the module.

A key tool in the proof will be *fractional ideals*. Confusingly, fractional ideals are not a special kind of ideal (unlike most phrases of the form "adjective noun"), but rather a generalisation of ideals. The purpose of fractional ideals (like fractions!) is tha they can arise by "dividing" ideals.

As well as using them to prove unique factorisation of ideals, we will subsequently use fractional ideals to define the class group of a number field which is the key to understanding how unique factorisation of elements fails.

Definition. Let K be a number field. A **fractional ideal** of $\mathcal{O}_K$ is a subset $\mathfrak{a} \subseteq K$ satisfying the following conditions.

(a) if $x, y \in \mathfrak{a}$, then $x + y \in \mathfrak{a}$;
(b) $x\mathfrak{a} \subseteq \mathfrak{a}$ for every $x \in \mathcal{O}_K$;
(c) there exists some non-zero $x \in \mathcal{O}_K$ such that $x\mathfrak{a} \subseteq \mathcal{O}_K$.

Conditions (a) and (b) are the ordinary conditions from the definition of an ideal of $\mathcal{O}_K$. However, a fractional ideal doesn't have to be an ideal of $\mathcal{O}_K$ because it might not be contained in $\mathcal{O}_K$. (Also, it won't be an ideal of K because K is a field so its only ideals are 0 and K itself.) Condition (c) says that a fractional ideal is not too much bigger than $\mathcal{O}_K$ (for example, it implies that K itself is not a fractional ideal).

e.g. For $\mathbb{Z}$, $\langle \frac{1}{2} \rangle := \frac{1}{2}\mathbb{Z}$ is a fractional ideal which is not contained in $\mathbb{Z}$.

More generally, for any number field K and any $\alpha \in K$, we can form the principal fractional ideal $\langle \alpha \rangle := \alpha\mathcal{O}_K$. This will be an ideal of $\mathcal{O}_K$ if and only if $\alpha \in \mathcal{O}_K$.

The following is clear.

Lemma 54. *An ideal of $\mathcal{O}_K$ is a fractional ideal.*
A fractional ideal is an ideal of $\mathcal{O}_K$ if and only if it is contained in $\mathcal{O}_K$.

The following lemma justifies the idea that "fractional ideals are fractions of ideals."

Lemma 55. *A subset $\mathfrak{a} \subseteq K$ is a fractional ideal if and only if there exist an ideal $\mathfrak{b} \subseteq \mathcal{O}_K$ and an element $x \in \mathcal{O}_K$ such that $\mathfrak{a} = \frac{1}{x}\mathfrak{b}$.*

Proof. It is immediate from the definitions that, if $\mathfrak{b}$ is an ideal of $\mathcal{O}_K$ and $x \in \mathcal{O}_K$, then $\frac{1}{x}\mathfrak{b}$ is a fractional ideal.

Conversely, if $\mathfrak{a}$ is a fractional ideal then condition (c) gives us $x \in \mathcal{O}_K$ such that $x\mathfrak{a} \subseteq \mathcal{O}_K$. Thanks to conditions (a) and (b), $\mathfrak{b} = x\mathfrak{a}$ is an ideal of $\mathcal{O}_K$ and we have $\mathfrak{a} = \frac{1}{x}\mathfrak{b}$. □

Thanks to Lemma 38, we can actually arrange that $x \in \mathbb{Z}$ in Lemma 55 (thus we get an "ideal version" of Lemma 24).

We define the fractional ideal generated by a set and the product of fractional ideals in the same way as for ideals (note that, like in the definition of fractional ideals, we work with subsets of K but the closure properties are only for multiplication by $\mathcal{O}_K$):

Definition. If $\alpha_1, \ldots, \alpha_r \in K$, then

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

If $\mathfrak{a}$, $\mathfrak{b}$ are fractional ideals of $\mathcal{O}_K$, then

$$\mathfrak{a}\mathfrak{b} = \{\alpha_1\beta_1 + \cdots + \alpha_r\beta_r : r \in \mathbb{N}, \alpha_1, \ldots, \alpha_r \in \mathfrak{a}, \beta_1, \ldots, \beta_r \in \mathfrak{b}\}.$$

The set $\langle \alpha_1, \ldots, \alpha_r \rangle$, for any $\alpha_1, \ldots, \alpha_r \in K$, and the product of any fractional ideals are fractional ideals (you can prove this directly, or use Lemma 55).

One of the benefits of working with fractional ideals instead of ideals is that the non-zero fractional ideals form a group under this multiplication operation. In particular, every fractional ideal has an inverse with respect to multiplication. In order to prove this, we will need to prove unique factorisation first.

For now, we define a fractional ideal which will eventually turn out to be the inverse of $\mathfrak{a}$.

Definition. Let $\mathfrak{a}$ be a non-zero fractional ideal of $\mathcal{O}_K$. We define $\mathfrak{a}^{-1}$ to be

$$\mathfrak{a}^{-1} = \{x \in K : x\mathfrak{a} \subseteq \mathcal{O}_K\}.$$

The notation suggests that $\mathfrak{a}^{-1}$ should be an inverse to $\mathfrak{a}$, but that is not the definition! So we have to be careful not to use $\mathfrak{a}^{-1}$ as an inverse to $\mathfrak{a}$ until we have proved that it actually is an inverse.

e.g. if $\alpha \in K \setminus \{0\}$, then

$$\begin{aligned}\langle \alpha \rangle^{-1} = \{x \in K : x\langle \alpha \rangle \subseteq \mathcal{O}_K\} &= \{x \in K : x\alpha \in \mathcal{O}_K\} \\ &= \{x \in K : x \in \tfrac{1}{\alpha}\mathcal{O}_K\} = \left\langle \tfrac{1}{\alpha} \right\rangle.\end{aligned}$$

Lemma 56. *$\mathfrak{a}^{-1}$ is a fractional ideal of $\mathcal{O}_K$, and $\mathfrak{a}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$.*

Proof. If $x \in \mathfrak{a}^{-1}$ and $y \in \mathfrak{a}$, then the definition of $\mathfrak{a}^{-1}$ shows that $xy \in \mathcal{O}_K$. Hence $\mathfrak{a}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$.

To show that $\mathfrak{a}^{-1}$ is a fractional ideal: conditions (a) and (b) are simple algebraic checks. If we pick any $x \in \mathfrak{a} \setminus \{0\}$, then (from the previous paragraph) $x\mathfrak{a} \subseteq \mathcal{O}_K$ so condition (c) is satisfied. □

A lemma on maximal ideals.

We need one more lemma before we begin the proof of unique factorisation of ideals.

Lemma 57. *Every proper ideal in $\mathcal{O}_K$ is contained in a maximal ideal. ("Proper" means the ideal is a proper subset of $\mathcal{O}_K$, i.e. not equal to $\mathcal{O}_K$ itself.)*

Proof. We prove this by induction on $\mathrm{Nm}(\mathfrak{a})$.

Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be a proper ideal.

Suppose that $\mathfrak{a} \neq \{0\}$. If $\mathfrak{a}$ is maximal itself, then the lemma is trivially true.

Otherwise, $\mathfrak{a}$ is not zero and not a maximal ideal so there exists an ideal $\mathfrak{b}$ such that $\mathfrak{a} \subsetneq \mathfrak{b} \subsetneq \mathcal{O}_K$. By Lemma 44, $\mathrm{Nm}(\mathfrak{b}) < \mathrm{Nm}(\mathfrak{a})$. Hence by induction $\mathfrak{b}$ is contained in a maximal ideal. But then $\mathfrak{a}$ is also contained in this maximal ideal.

For $\{0\}$: choose $x \in \mathcal{O}_K$ not a unit (e.g. $x = 2$). Then $\{0\} \subseteq \langle x \rangle \subsetneq \mathcal{O}_K$ so by what we have just proved,

$$\{0\} \subseteq \langle x \rangle \subseteq \text{a maximal ideal.}$$ □

We will use this strategy of "induction on the norm" several times. It relies on the fact that $\mathcal{O}_K/\mathfrak{a}$ is finite for every proper ideal $\mathfrak{a}$ (so that the norm is well-defined). We said that this is a very special property of $\mathcal{O}_K$ which does not apply to many other rings. But you can actually replace it by a weaker property: $\mathcal{O}_K$ is a Noetherian ring, a concept which will be defined in Commutative Algebra and which allows you to do "induction on ideals" in a more abstract sense.

Thus Lemma 57 holds for every Noetherian ring, and even for every ring if you use the Axiom of Choice.

Proof of unique factorisation of ideals.

We are now ready to begin the proof of the unique factorisation of ideals in $\mathcal{O}_K$ (Theorem 50).

The proof goes through a number of steps. Several of these steps will prove things which "obviously should be true" but each step depends on the previous ones, often in a subtle way, so we have to be careful to prove them in the right order.

Step 1. *Every non-zero ideal $\mathfrak{a} \subseteq \mathcal{O}_K$ has the following property:*
(*) there exist non-zero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r \subseteq \mathfrak{a}$.

Proof. We prove this by induction on $\mathrm{Nm}(\mathfrak{a})$.

If $\mathfrak{a} = \mathcal{O}_K$, we can just pick $r = 1, \mathfrak{p}_1 =$ any non-zero prime ideal. If $\mathfrak{a}$ is prime, we can pick $r = 1, \mathfrak{p}_1 = \mathfrak{a}$.

So we may assume that $\mathfrak{a} \neq \mathcal{O}_K$ and $\mathfrak{a}$ is not prime. Then by the definition of prime ideal, we can pick $x, y \in \mathcal{O}_K$ such that $x, y \notin \mathfrak{a}$ but $xy \in \mathfrak{a}$.

Let $\mathfrak{b} = \langle \mathfrak{a}, x \rangle$ and $\mathfrak{c} = \langle \mathfrak{a}, y \rangle$. Then

$$\mathfrak{b}\mathfrak{c} = \mathfrak{a}^2 + x\mathfrak{a} + y\mathfrak{a} + \langle xy \rangle \subseteq \mathfrak{a}.$$

Since $\mathfrak{a} \subseteq \mathfrak{b}$ but $\mathfrak{a} \neq \mathfrak{b}$, we have $\mathrm{Nm}(\mathfrak{b}) < \mathrm{Nm}(\mathfrak{a})$. Hence by induction, there are prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r \subseteq \mathfrak{b}$.

Similarly $\mathrm{Nm}(\mathfrak{c}) < \mathrm{Nm}(\mathfrak{a})$ so by induction there are prime ideals $\mathfrak{q}_1, \ldots, \mathfrak{q}_s$ such that $\mathfrak{q}_1 \cdots \mathfrak{q}_s \subseteq \mathfrak{c}$.

Thus

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r \mathfrak{q}_1 \cdots \mathfrak{q}_s \subseteq \mathfrak{b}\mathfrak{c} \subseteq \mathfrak{a}. \qquad \square$$

Aside: we could replace the induction on $\mathrm{Nm}(\mathfrak{a})$ here by an argument using the fact that $\mathcal{O}_K$ is Noetherian.

19. Proof of unique factorisation of ideals

We continue the proof of the unique factorisation of ideals in $\mathcal{O}_K$. Step 2 is the hardest step. The reason this is hard is that we don't have an easy way to construct elements of $\mathfrak{p}^{-1}$ in order to show that it is bigger than $\mathcal{O}_K$.

Step 2. *If $\mathfrak{p} \subseteq \mathcal{O}_K$ is a non-zero prime ideal, then $\mathcal{O}_K \subsetneq \mathfrak{p}^{-1}$.*

Proof. Since $\mathfrak{p} \subseteq \mathcal{O}_K$, we have $\mathcal{O}_K \subseteq \mathfrak{p}^{-1}$. The hard part is to prove that $\mathfrak{p}^{-1} \neq \mathcal{O}_K$.

We want to apply Step 1 but applying it directly to $\mathfrak{p}$ is no use, because $\mathfrak{p}$ is already a prime ideal so Step 1 tells us nothing new about $\mathfrak{p}$. Instead, pick a non-zero element $\alpha \in \mathfrak{p}$ and apply Step 1 to $\langle \alpha \rangle$. We get non-zero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that

$$\mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_r \subseteq \langle \alpha \rangle.$$

Choose these prime ideals so that r is as small as possible.

Since $\langle \alpha \rangle \subseteq \mathfrak{p}$, this implies that

$$\mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_r \subseteq \mathfrak{p}.$$

Since $\mathfrak{p}$ is a prime ideal, Lemma 46 tells us that $\mathfrak{p}$ contains one of the $\mathfrak{p}_i$. WLOG $\mathfrak{p}_1 \subseteq \mathfrak{p}$. By Corollary 49, $\mathfrak{p}_1$ is maximal ideal of $\mathcal{O}_K$ and so $\mathfrak{p}_1 = \mathfrak{p}$.

Since we chose $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ so that r is as small as possible,

$$\mathfrak{p}_2 \mathfrak{p}_3 \cdots \mathfrak{p}_r \not\subseteq \langle \alpha \rangle.$$

Therefore we can choose $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_r$ which is not in $\langle \alpha \rangle$. But $\beta \mathfrak{p} \subseteq \mathfrak{p}\mathfrak{p}_2 \mathfrak{p}_3 \cdots \mathfrak{p}_r \subseteq \langle \alpha \rangle$. Therefore $\alpha^{-1}\beta \mathfrak{p} \subseteq \mathcal{O}_K$ i.e. $\alpha^{-1}\beta \in \mathfrak{p}^{-1}$. But since $\beta \notin \langle \alpha \rangle$, $\beta\alpha^{-1} \notin \mathcal{O}_K$. □

Aside: in Step 2, we used Corollary 49, so this no longer applies in an arbitrary Noetherian ring.

Step 3. *If $\mathfrak{a} \subseteq \mathcal{O}_K$ is a non-zero ideal and $\beta \in K$ is such that $\beta \mathfrak{a} \subseteq \mathfrak{a}$, then $\beta \in \mathcal{O}_K$.*

Proof. Since $(\mathfrak{a}, +)$ is a subgroup of the finitely generated abelian group $(\mathcal{O}_K, +)$, it is itself a finitely generated abelian group. Applying Lemma 21 to $H = \mathfrak{a}$, we deduce that β is an algebraic integer. In other words, $\beta \in \mathcal{O}_K$. □

Now we can put the previous steps together. The next step looks like a small strengthening of Step 2, but it is actually a big step forward because it uses Step 3 as well.

Step 4. *If $\mathfrak{p} \subseteq \mathcal{O}_K$ is a non-zero prime ideal and $\mathfrak{a} \subseteq \mathcal{O}_K$ is a non-zero ideal such that $\mathfrak{a} \subseteq \mathfrak{p}$, then $\mathfrak{a} \subsetneq \mathfrak{p}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$.*

Proof. Since $1 \in \mathfrak{p}^{-1}$, it is clear that $\mathfrak{a} \subseteq \mathfrak{p}^{-1}\mathfrak{a}$. Since $\mathfrak{a} \subseteq \mathfrak{p}$, $\mathfrak{p}^{-1} \subseteq \mathfrak{a}^{-1}$ and so $\mathfrak{p}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$.

The key point is proving that $\mathfrak{a} \neq \mathfrak{p}^{-1}\mathfrak{a}$. Assume for contradiction that $\mathfrak{a} = \mathfrak{p}^{-1}\mathfrak{a}$. Then for every $\beta \in \mathfrak{p}^{-1}$, we have $\beta \mathfrak{p} \subseteq \mathfrak{p}$. Hence by Step 3, $\beta \in \mathcal{O}_K$.

Thus $\mathfrak{p}^{-1} \subseteq \mathcal{O}_K$, contradicting Step 2. □

Now it is quite easy to prove that "$\mathfrak{p}^{-1}$" means what we expect, at least for prime ideals.

Step 5. *If $\mathfrak{p} \subseteq \mathcal{O}_K$ is a non-zero prime ideal, then $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$.*

Proof. We know that $\mathfrak{p}\mathfrak{p}^{-1}$ is a fractional ideal contained in $\mathcal{O}_K$, so $\mathfrak{p}\mathfrak{p}^{-1}$ is an ideal of $\mathcal{O}_K$. By Step 4, we have $\mathfrak{p} \subsetneq \mathfrak{p}\mathfrak{p}^{-1}$. Since $\mathfrak{p}$ is a maximal ideal, this implies that $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$. □

We will finish the proof of existence and uniqueness of factorisation of ideals in the next lecture. Now that we have established Step 5 (allowing us to "divide" by prime ideals and ensure that $\mathfrak{p}^{-1}$ cancels with $\mathfrak{p}$), the rest of the proof looks similar to the proof of existence and uniqueness of prime factorisations in $\mathbb{Z}$.

A remark about learning proofs and exam questions.

When learning proofs, you generally look for patterns and simple manipulations which get used again and again, parts which are just applying definitions so you don't need to remember them separately, and small steps you could work out for yourself if you know what you are aiming for. For example, in the proof of unique factorisation, a pattern which appeared a lot is arguing about which ideals are contained/strictly contained in other ideals.

Then you don't need to hold all these details in your memory, just the key steps which they link up. There's a fancy term for this in mathematical education – it's called "chunking."

Anyway, for this proof, I think that remembering exactly what is Step 1, Step 2 etc and the order in which they come would be unreasonable for an exam. However, the individual proofs of each step are short enough that you can apply the chunking strategy to them. So an exam question on this proof might look something like: "Here's the statement of Steps 2 and 3. Assuming these, prove Step 4." (The statement of Step 4 would also be given.)

20. Properties of ideals of number fields

End of proof of unique factorisation of ideals.

At last we can prove existence and uniqueness of factorisation of ideals.

Step 6 (Existence of factorisation into prime ideals)**.** *For any non-zero proper ideal $\mathfrak{a} \subseteq \mathcal{O}_K$, there exist prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r$.*

Proof. This strengthens Step 1 because $\mathfrak{a}$ is equal to the product, rather than containing it.

Again the proof uses induction on $\mathrm{Nm}(\mathfrak{a})$, and it is similar to the proof of existence of prime factorisations in $\mathbb{Z}$.

Let $\mathfrak{a}$ be a non-zero ideal in $\mathcal{O}_K$. By Lemma 57, $\mathfrak{a} \subseteq \mathfrak{p}$ for some maximal ideal $\mathfrak{p}$.

By Step 4, $\mathfrak{p}^{-1}\mathfrak{a}$ is an ideal of $\mathcal{O}_K$ which strictly contains $\mathfrak{a}$.

If $\mathfrak{p}^{-1}\mathfrak{a} = \mathcal{O}_K$, then by Step 5 we have

$$\mathfrak{p} = \mathfrak{p}\mathcal{O}_K = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{a} = \mathcal{O}_K\mathfrak{a} = \mathfrak{a}$$

so $\mathfrak{a}$ is a product of prime ideals (of the single ideal $\mathfrak{p}$).

Otherwise, we have $\mathcal{O}_K \subsetneq \mathfrak{p}^{-1}\mathfrak{a} \subsetneq \mathfrak{a}$. Then $\mathrm{Nm}(\mathfrak{p}^{-1}\mathfrak{a}) < \mathrm{Nm}(\mathfrak{a})$, so by induction $\mathfrak{p}^{-1}\mathfrak{a}$ is equal to a product of prime ideals i.e.

$$\mathfrak{p}^{-1}\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3 \cdots \mathfrak{p}_r$$

for some prime ideals $\mathfrak{p}_2, \ldots, \mathfrak{p}_r$. Multiplying on both sides by $\mathfrak{p}$ and using Step 5, we get

$$\mathfrak{p}\mathfrak{p}_2\mathfrak{p}_3 \cdots \mathfrak{p}_r = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{a} = \mathcal{O}_K\mathfrak{a} = \mathfrak{a}. \qquad \square$$

Step 7 (Uniqueness of factorisation into prime ideals)**.** *Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be a non-zero proper ideal and suppose that*

$$\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$$

where $\mathfrak{p}_1, \ldots, \mathfrak{p}_r, \mathfrak{q}_1, \ldots, \mathfrak{q}_s$ are prime ideals. Then $r = s$ and $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ form a permutation of $\mathfrak{q}_1, \ldots, \mathfrak{q}_s$.

Proof. The proof resembles the proof that prime factorisations in $\mathbb{Z}$ are unique, just written using ideals instead of elements and using Lemma 46.

We proceed by induction on r.

If $r = 1$, then $\mathfrak{p}_1 = \mathfrak{q}_1 \cdots \mathfrak{q}_s$ so $\mathfrak{p}_1 \subseteq \mathfrak{q}_i$ for all i. But $\mathfrak{p}_1$ is maximal by Corollary 49, so this forces $\mathfrak{q}_i = \mathfrak{p}_1$ for all i. Hence our original equation becomes $\mathfrak{p}_1 = \mathfrak{p}_1^s$. Multiplying both sides by $\mathfrak{p}_1{}^{-}1$ and using Step 5, we get

$$\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_1^{-1} = \mathfrak{p}_1^{s-1}$$

which forces $s = 1$ (otherwise $\mathfrak{p}_1^{s-1} \subseteq \mathfrak{p}_1$). This completes the proof when $r = 1$.

If $r > 1$, since $\mathfrak{p}_1 \mid \mathfrak{q}_1\mathfrak{q}_2 \cdots \mathfrak{q}_r$ and the $\mathfrak{q}_i$ are prime ideals, by Lemma 46, we must have $\mathfrak{p}_1 \mid \mathfrak{q}_j$ for some j. WLOG $\mathfrak{p}_1 \mid \mathfrak{q}_1$, that is, $\mathfrak{q}_1 \subseteq \mathfrak{p}_1$. Since $\mathfrak{q}_1$ is a maximal ideal, this implies that $\mathfrak{p}_1 = \mathfrak{q}_1$. Hence using Step 5 twice,

$$\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{p}_1^{-1}\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{p}_1^{-1}\mathfrak{a} = \mathfrak{q}_1^{-1}\mathfrak{a} = \mathfrak{q}_1^{-1}\mathfrak{q}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s = \mathfrak{q}_2 \cdots \mathfrak{q}_s.$$

By induction, we conclude that $r-1 = s-1$ and that $\mathfrak{p}_2, \ldots, \mathfrak{p}_r$ are a permutation of $\mathfrak{q}_2, \ldots, \mathfrak{q}_s$. $\square$

This completes the proof of Theorem 50.

Aside (non-examinable, for people with an interest in commutative algebra or algebraic geometry):

Looking back over the proof of Theorem 50, the properties of $\mathcal{O}_K$ which we used were:

(1) $\mathcal{O}_K$ is Noetherian (in Steps 1 and 6);
(2) every non-zero prime ideal is maximal (in Steps 2 and 7);
(3) if $\alpha \in K$ is a root of a monic polynomial in $\mathcal{O}_K[X]$, then $\alpha \in \mathcal{O}_K$ (this is called "integral closedness" and is the property which makes Step 3 work).

An integral domain with these properties is called a **Dedekind domain** and Theorem 50 works in any Dedekind domain. Besides $\mathcal{O}_K$, the other important example are rings of the form $K[X]$ where K is any field, and "finite extensions of $K[X]$" which have a geometrical interpretation as "the ring of functions on a smooth algebraic curve." ((2) says the rings correspond to 1-dimensional geometric objects, (3) that they are smooth.)

Group of fractional ideals.

Now we prove some other facts which are relatively easy to prove, now that we have the proof of unique factorisation of ideals. In particular, we complete the promise to show that the non-zero fractional ideals form a group under multiplication.

Associativity is obvious (as is commutativity). The identity element is $\mathcal{O}_K = \langle 1 \rangle$. We just have to check that every element has an inverse. Indeed, we check that $\mathfrak{a}^{-1}$ is the inverse of $\mathfrak{a}$ in this group (otherwise it would have been a very confusing choice of notation!)

Lemma 58. *Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be a non-zero fractional ideal. Then $\mathfrak{a}\mathfrak{a}^{-1} = \mathcal{O}_K$.*

Proof. By Lemma 55, we can write $\mathfrak{a} = \frac{1}{x}\mathfrak{b}$ for some $x \in \mathcal{O}_K$ and an ideal $\mathfrak{b} \subseteq \mathcal{O}_K$. Then we can write $\mathfrak{b}$ as a product of prime ideals: $\mathfrak{b} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r$.

Let $\mathfrak{c} = x\mathfrak{p}_1^{-1}\mathfrak{p}_2^{-1} \cdots \mathfrak{p}_r^{-1}$. Then

$$\mathfrak{a}\mathfrak{c} = (\frac{1}{x}\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r)(x\mathfrak{p}_1^{-1}\mathfrak{p}_2^{-1} \cdots \mathfrak{p}_r^{-1}) = \mathcal{O}_K$$

using Step 5 to cancel each of the products $\mathfrak{p}_1\mathfrak{p}_1^{-1}$, $\mathfrak{p}_2\mathfrak{p}_2^{-1}$ etc. This shows that $\mathfrak{c}$ is the inverse of $\mathfrak{a}$, and hence is enough to show that the fractional ideals form a group.

But we still want to show that the previous definition of the notation $\mathfrak{a}^{-1}$ actually gives the inverse.

Since $\mathfrak{a}\mathfrak{c} = \mathcal{O}_K$, the definition of $\mathfrak{a}^{-1}$ tells us that $\mathfrak{c} \subseteq \mathfrak{a}^{-1}$. In the other direction, using the facts that $\mathfrak{a}\mathfrak{c} = \mathcal{O}_K$ (just proved) and $\mathfrak{a}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$ (Lemma 56), we have

$$\mathfrak{a}^{-1} = \mathfrak{a}^{-1}\mathcal{O}_K = \mathfrak{a}^{-1}\mathfrak{a}\mathfrak{c} \subseteq \mathcal{O}_K\mathfrak{c} = \mathfrak{c}.$$

Thus $\mathfrak{c} = \mathfrak{a}^{-1}$. □

Recall that we defined $\mathfrak{a} \mid \mathfrak{b}$ to mean $\mathfrak{b} \subseteq \mathfrak{a}$. This looks rather different from the definition of divisibility of elements $a \mid b$ which says "there exists c such that $b = ac$." We can now verify that our definition of $\mathfrak{a} \mid \mathfrak{b}$ is actually equivalent to an "ideal version" of the definition for elements.

Lemma 59. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be non-zero ideals of $\mathcal{O}_K$ such that $\mathfrak{b} \subseteq \mathfrak{a}$ (i.e. $\mathfrak{a} \mid \mathfrak{b}$). Then there exists an ideal $\mathfrak{c} \subseteq \mathcal{O}_K$ such that $\mathfrak{b} = \mathfrak{ac}$.*

Proof. Let $\mathfrak{c} = \mathfrak{a}^{-1}\mathfrak{b}$. By Lemma 58,

$$\mathfrak{ac} = \mathfrak{a}\mathfrak{a}^{-1}\mathfrak{b} = \mathcal{O}_K\mathfrak{b} = \mathfrak{b}.$$

Since $\mathfrak{b} \subseteq \mathfrak{a}$, we have $\mathfrak{c} \subseteq \mathfrak{a}^{-1}\mathfrak{a} \subseteq \mathcal{O}_K$. Thus $\mathfrak{c}$ is an ideal of $\mathcal{O}_K$ (not just a fractional ideal). □

Ideal norms are multiplicative.

At last we have enough theory to prove Lemma 53: if $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}_K$ are non-zero ideals, then

$$\mathrm{Nm}(\mathfrak{ab}) = \mathrm{Nm}(\mathfrak{a})\,\mathrm{Nm}(\mathfrak{b}).$$

The proof of this is surprisingly difficult! It is also not particularly enlightening, so it is non-examinable (but the statement is examinable).

Proof of Lemma 53. (Non-examinable)

We can factorise $\mathfrak{b}$ into prime ideals and multiply by them one at a time, so it suffices to prove the lemma when $\mathfrak{b} = \mathfrak{p}$ is a prime ideal.

Overall strategy: we will write down an additive group homomorphism $\mathcal{O}_K \to \mathfrak{a}/\mathfrak{ap}$. Using the First Isomorphism Theorem for groups, we get an isomorphism $\mathcal{O}_K/\mathfrak{p} \cong \mathfrak{a}/\mathfrak{ap}$. Then

$$\mathrm{Nm}(\mathfrak{ap}) = [\mathcal{O}_K : \mathfrak{ap}] = [\mathcal{O}_K : \mathfrak{a}][\mathfrak{a} : \mathfrak{ap}] = \mathrm{Nm}(\mathfrak{a})[\mathfrak{a} : \mathfrak{ap}]$$

and we are done because the isomorphism establishes that $\mathrm{Nm}(\mathfrak{p}) = [\mathfrak{a} : \mathfrak{ap}]$.

Constructing the homomorphism: Since $\mathfrak{p} \neq \mathcal{O}_K$ and the fractional ideals form a group (so we can cancel the $\mathfrak{a}$s), we get $\mathfrak{a} \neq \mathfrak{ap}$. Also $\mathfrak{ap} \subseteq \mathfrak{a}$, so we can choose an element $\alpha \in \mathfrak{a} \setminus \mathfrak{ap}$.

Define a group homomorphism (of the additive groups) $\phi\colon \mathcal{O}_K \to \mathfrak{a}/\mathfrak{ap}$ by

$$\phi(x) = x\alpha + \mathfrak{ap}.$$

In order to apply the First Isomorphism Theorem, we need to calculate the kernel and image of ϕ.

Claim: ϕ is surjective.

Let $\mathfrak{b} = \langle \alpha, \mathfrak{ap} \rangle$. Then $\mathfrak{ap} \subsetneq \mathfrak{b} \subseteq \mathfrak{a}$. Multiplying by $\mathfrak{a}^{-1}$, we get $\mathfrak{p} \subsetneq \mathfrak{a}^{-1}\mathfrak{b} \subseteq \mathcal{O}_K$. Now $\mathfrak{p}$ is a maximal ideal (Corollary 49), so this implies that $\mathfrak{a}^{-1}\mathfrak{b} = \mathcal{O}_K$ i.e. $\mathfrak{a} = \mathfrak{b}$.

Indeed, if $y + \mathfrak{ap} \in \mathfrak{a}/\mathfrak{ap}$, then we have

$$y \in \mathfrak{a} = \mathfrak{b} = \langle \alpha, \mathfrak{ap} \rangle$$

so we can write $y = x\alpha + z$ where $x \in \mathcal{O}_K$ and $z \in \mathfrak{ap}$. Then $y + \mathfrak{ap} = \phi(x)$, so ϕ is surjective.

Claim: $\ker(\phi) = \mathfrak{p}$. Indeed,

$$\ker(\phi) = \{x \in \mathcal{O}_K : x\alpha \in \mathfrak{a}\mathfrak{p}\} = \mathcal{O}_K \cap \alpha^{-1}\mathfrak{a}\mathfrak{p}.$$

Now $\alpha^{-1}\mathfrak{a}\mathfrak{p}$ is a fractional ideal, so $\mathcal{O}_K \cap \alpha^{-1}\mathfrak{a}\mathfrak{p}$ is an ideal in $\mathcal{O}_K$. (Note: we can't just say "$\ker(\phi)$ is a kernel of a homomorphism, therefore an ideal" because that only works for ring homomorphisms and ϕ is a group homomorphism, not a ring homomorphism.)

Since $\alpha \in \mathfrak{a}$, $\mathfrak{p} \subseteq \ker(\phi)$. Since $\alpha \notin \mathfrak{a}\mathfrak{p}$, $\phi(1) = \alpha + \mathfrak{a}\mathfrak{p} \neq 0 + \mathfrak{a}\mathfrak{p}$ and so $\ker(\phi) \neq \mathcal{O}_K$. Thus $\ker(\phi)$ is an ideal and $\mathfrak{p} \subseteq \ker(\phi) \subsetneq \mathcal{O}_K$. Since $\mathfrak{p}$ is a maximal ideal, we must have $\ker(\phi) = \mathfrak{p}$.

Using both Claims, the First Isomorphism Theorem for groups tells us that

$$\mathcal{O}_K/\mathfrak{p} \cong \mathfrak{a}/\mathfrak{a}\mathfrak{p}$$

as additive groups, and so $[\mathcal{O}_K : \mathfrak{p}] = [\mathfrak{a} : \mathfrak{a}\mathfrak{p}]$. □

21. The class group

Definition of the class group.

The class group of a number field K is a finite abelian group which measures "how badly unique factorisation fails in $\mathcal{O}_K$."

Definition. Let K be a number field. Write:

- I_K = the group of non-zero fractional ideals of K;
- P_K = the group of non-zero principal fractional ideals of K

We showed that I_K was a group (under multiplication of fractional ideals) in the last lecture. P_K is a subgroup of I_K because $\langle\alpha\rangle\langle\beta\rangle = \langle\alpha\beta\rangle$ and $\langle\alpha\rangle\langle\alpha^{-1}\rangle = \mathcal{O}_K$. The group I_K is abelian because multiplication is commutative, hence every subgroup of I_K is a normal subgroup. Therefore the quotient group I_K/P_K makes sense.

Definition. The **class group** of K is the quotient group $\mathrm{Cl}(K) = I_K/P_K$.

The elements of $\mathrm{Cl}(K)$ are called **ideal classes**. If $\mathfrak{a}$ is a non-zero fractional ideal of K, we write $[\mathfrak{a}]$ for its class in $\mathrm{Cl}(K)$.

Observe that $[\mathfrak{a}] = [\mathfrak{b}]$ if and only if $\mathfrak{a} = \gamma\mathfrak{b}$ for some $\gamma \in K$.

We said that $\mathrm{Cl}(K)$ is finite but this is far from obvious. Notice that I_K and P_K are both infinite groups – indeed, they are non-finitely generated groups (by unique factorisation of ideals, the set of all non-zero prime ideals forms a minimal generating set for I_K). Hence it is not at all obvious that their quotient is finite. The finiteness of the class group is one of the deepest theorems of the course.

UFDs and PIDs.

The following lemma justifies the slogan that $\mathrm{Cl}(K)$ measures how badly $\mathcal{O}_K$ fails to be a UFD.

Lemma 60. *Let K be a number field. Then $\mathrm{Cl}(K) = \{1\}$ if and only if $\mathcal{O}_K$ is a UFD.*

Proof. From the definition of $\mathrm{Cl}(K)$, we see immediately that

$$\mathrm{Cl}(K) = \{1\} \quad \Leftrightarrow \quad P_K = I_K \quad \Leftrightarrow \quad \mathcal{O}_K \text{ is a PID.}$$

Every PID is a UFD. So what we have to prove is: if $\mathcal{O}_K$ is a UFD, then it is a PID.

Since every proper ideal of $\mathcal{O}_K$ is a product of prime ideals, it suffices to prove that every prime ideal of $\mathcal{O}_K$ is principal.

Let $\mathfrak{p}$ be a non-zero prime ideal of $\mathcal{O}_K$ and let $\alpha \in \mathfrak{p} \setminus \{0\}$. We can factorise

$$\alpha = \pi_1\pi_2\cdots\pi_r$$

where $\pi_1, \ldots, \pi_r \in \mathcal{O}_K$ are irreducible elements. (This doesn't use the fact that $\mathcal{O}_K$ is a UFD: it is possible in any integral domain.)

Now we use that $\mathcal{O}_K$ is a UFD: in a UFD, irreducible elements are prime. In any integral domain, the ideal generated by a prime element is a prime ideal. Hence the ideals $\langle\pi_i\rangle$ are prime ideals of $\mathcal{O}_K$.

We have

$$\mathfrak{p} \mid \langle \alpha \rangle = \langle \pi_1 \rangle \cdots \langle \pi_r \rangle.$$

By the definition of prime ideal, we deduce that $\mathfrak{p} \mid \langle \pi_i \rangle$ for some i. In other words $\langle \pi_i \rangle \subseteq \mathfrak{p}$.

(I got this the wrong way round in the lecture: $\langle \pi_i \rangle \subseteq \mathfrak{p}$ is correct. In the lecture, I wrote $\langle \pi_i \rangle \supseteq \mathfrak{p}$ then got a bit confused.)

In $\mathcal{O}_K$, every prime ideal is maximal. In particular $\langle \pi_i \rangle$ is maximal, so we conclude that $\mathfrak{p} = \langle \pi_i \rangle$. □

Minkowski's theorem.

A fundamental property of the class group is that it is finite. In order to prove this, we will use Minkowski's theorem. Minkowski's theorem is also important for calculating the class group of a particular number field, because it gives us a way to find ideals which represent every class in $\mathrm{Cl}(K)$.

Definition. Let K be a number field of signature (r, s) and degree $n = r + 2s$. Let Δ_K be the discriminant of K. The **Minkowski bound** of K is

$$B_K = \left(\frac{4}{\pi}\right)^s \frac{n!}{n^n} \sqrt{|\Delta_K|}.$$

Theorem 61. *Every ideal class in* $\mathrm{Cl}(K)$ *has a representative* $\mathfrak{a}$ *which is an ideal of* $\mathcal{O}_K$ *(not just a fractional ideal) and satisfies* $\mathrm{Nm}(\mathfrak{a}) \leq B_K$.

e.g. We compute the class group of $\mathbb{Q}(i)$.

For $K = \mathbb{Q}(i)$, the degree is $n = 2$, the signature is $(r, s) = (0, 1)$ and the discriminant is $\Delta_K = -4$. Hence the Minkowski bound is

$$B_K = \left(\frac{4}{\pi}\right)^1 \frac{2!}{2^2} \sqrt{|-4|} = \frac{2}{\pi} \times 2 = \frac{4}{\pi} < 2.$$

Hence by Theorem 61, every class in $\mathrm{Cl}(K)$ has a representative which is an ideal in $\mathcal{O}_K$ and has norm less than 2. The norm of an ideal is always an integer, so in fact every class has a representative of norm 1.

But the only ideal in $\mathcal{O}_K$ of norm 1 is $\mathcal{O}_K$ itself. Hence $\mathrm{Cl}(K)$ contains only one class, the trivial class. Thus $\mathbb{Z}[i]$ is a PID and hence a UFD.

This was proved in Intro to Number Theory by showing that $\mathbb{Z}[i]$ is a Euclidean domain, but now we have an alternative proof – which was very simple once we know Theorem 61.

e.g. We do another example of computing the class group: $K = \mathbb{Q}(\sqrt{-10})$.

Since $-10 \equiv 2 \bmod 4$, $\mathcal{O}_K = \mathbb{Z}[\sqrt{-10}]$ and $\Delta_K = -40$. The signature is $(0, 1)$. Thus the Minkowski bound is

$$B_K = \left(\frac{4}{\pi}\right)^1 \frac{2!}{2^2} \sqrt{|-40|} = \frac{2}{\pi} \sqrt{40} < \frac{2}{3} \times 7 < 5.$$

Thus by Theorem 61, every class in $\mathrm{Cl}(K)$ has a representative which is an ideal of norm < 5.

If $\mathfrak{a}$ is an ideal of norm 2, 3 or 4, then it's a product of prime ideals whose norms are 2, 3 or 4. Such prime ideals must divide $\langle 2\rangle$ or $\langle 3\rangle$.

Let's use the Dedekind–Kummer theorem to work out the prime ideals dividing $\langle 2\rangle$ and $\langle 3\rangle$. Taking $\alpha = \sqrt{-10}$, $f(X) = X^2 + 10$, Dedekind–Kummer gives:

- For $p = 2$: $f(X) \equiv X^2 \bmod 2$ so $\langle 2\rangle = \mathfrak{p}_2^2$ where $\mathfrak{p}_2$ is a prime ideal of norm 2.
- For $p = 3$: $f(X) \equiv X^2 - 2 \bmod 3$ which is irreducible since 2 is not a quadratic residue mod 3, so $\langle 3\rangle$ is a prime ideal of $\mathcal{O}_K$.

$\langle 3\rangle$ has norm 9, so it doesn't give us any ideals of norm < 5. Thus $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2]$ (see Lemma 62 in the next lecture for a more careful justification of this).

To test whether $\mathfrak{p}_2$ is principal, we look for elements of $\mathcal{O}_K$ of norm ± 2. The equation $\mathrm{Nm}(x+y\sqrt{-10}) = x^2+10y^2 = \pm 2$ has no solutions in $\mathbb{Z}$, so $\mathcal{O}_K$ contains no elements of norm ± 2. Hence $\mathfrak{p}_2$ is not principal.

On the other hand, from the Dedekind–Kummer calculation above, $\mathfrak{p}_2^2 = \langle 2\rangle$ is principal. So $[\mathfrak{p}_2] \neq [1]$ in $\mathrm{Cl}(K)$ but $[\mathfrak{p}_2]^2 = [\langle 2\rangle] = [1]$. Thus $[\mathfrak{p}_2]$ has order 2, so $\mathrm{Cl}(K) \cong \mathbb{Z}/2\mathbb{Z}$.

Filling in some more details in the last paragraph: $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2]$. In other words, $\mathrm{Cl}(K) = \{[\mathfrak{p}_2]^n : n \in \mathbb{Z}\}$. But when we work out $[1], [\mathfrak{p}_2], [\mathfrak{p}_2]^2, [\mathfrak{p}_3]^2, \ldots$, once we reach $[\mathfrak{p}_2]^2$ we are already back at $[1]$. Also $[\mathfrak{p}_2]^{-1} = [\mathfrak{p}_2]$. Thus $\mathrm{Cl}(K) = \{[1], [\mathfrak{p}_2]\}$ and it has to be the cyclic group of order 2.

22. More on the class group

Computing the class group.

Theorem 61 tells us that, when we compute the class group, we only need to look at ideals of norm $\leq B_K$ in order to hit every class in $\mathrm{Cl}(K)$. In fact, we only need to look at *prime* ideals of norm $\leq B_K$: these don't necessarily hit every ideal class, but they do generate $\mathrm{Cl}(K)$. For convenience, we state this as a lemma.

Lemma 62. *Let K be a number field. The group $\mathrm{Cl}(K)$ is generated by the classes of prime ideals in $\mathcal{O}_K$ of norm $\leq B_K$.*

Proof. By Theorem 61, every class in $\mathrm{Cl}(K)$ has a representative $\mathfrak{a} \subseteq \mathcal{O}_K$ such that $\mathrm{Nm}(\mathfrak{a}) \leq B_K$. We can write $\mathfrak{a}$ as a product of prime ideals:

$$\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r$$

from which we get

$$[\mathfrak{a}] = [\mathfrak{p}_1][\mathfrak{p}_2] \cdots [\mathfrak{p}_r].$$

Because ideal norms are multiplicative, $\mathrm{Nm}(\mathfrak{p}_i) \leq \mathrm{Nm}(\mathfrak{a}) \leq B_K$ for each i. □

Thus we have the following strategy for finding $\mathrm{Cl}(K)$:

(1) Calculate the Minkowski bound B_K.
(2) For each rational prime $p \leq B_K$, use Dedekind–Kummer to factorise $\langle p \rangle$ into prime ideals of $\mathcal{O}_K$. (When we do this, some of the prime ideals we obtain may have norm $> B_K$. We can throw these away.)
(3) Check whether each of the prime ideals we have found is principal.
(4) For any non-principal ideals, look for relations between their ideal classes, and eventually prove that we have found all the relations. This is an *ad hoc* process – if we only found a single non-principal prime ideal, then it is just a matter of finding the smallest power of that ideal which becomes principal. If there are multiple non-principal prime ideals of norm $\leq B_K$, then this may require more tricks.

e.g. A harder example: $K = \mathbb{Q}(\sqrt{-14})$.

Since $-14 \equiv 2 \bmod 4$, $\mathcal{O}_K = \mathbb{Z}[\sqrt{-14}]$ and $\Delta_K = -56$. The signature is $(0, 1)$. Thus the Minkowski bound is

$$B_K = \left(\frac{4}{\pi}\right)^1 \frac{2!}{2^2}\sqrt{|-56|} = \frac{2}{\pi}\sqrt{56} < \frac{2}{3} \times 8 = \frac{16}{3}.$$

This is only just over 5, and the inequality is quite weak because 56 is a long way from $8^2 = 64$. So it seems likely that $B_K < 5$. It would be a shame to make extra work for ourselves by having to factorise $\langle 5 \rangle$ and find relations involving its prime factors, so we get out the calculator and find

$$B_K \approx 4.76 < 5.$$

You could also prove this by hand, by noting that $56 < 56.25 = 7.5^2$.

Thus by Lemma 62, $\mathrm{Cl}(K)$ is generated by prime ideals dividing $\langle 2 \rangle$ or $\langle 3 \rangle$.

Using the Dedekind–Kummer theorem with $\alpha = \sqrt{-14}$, $f(X) = X^2 + 14$:

- $p = 2$: $f(X) \equiv X^2 \bmod 2$ so $\langle 2\rangle = \mathfrak{p}_2^2$ where $\mathfrak{p}_2 = \langle 2, \sqrt{-14}\rangle$.
- $p = 3$: $f(X) \equiv (X-1)(X+1) \bmod 3$ so $\langle 3\rangle = \mathfrak{p}_3\mathfrak{q}_3$ where $\mathfrak{p}_3 = \langle 3, -1 + \sqrt{-14}\rangle$ and $\mathfrak{q}_3 = \langle 3, 1 + \sqrt{-14}\rangle$.

The difference from $\mathbb{Q}(\sqrt{-10})$ is that this time $\langle 3\rangle$ factorises.

Now $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2]$, $[\mathfrak{p}_3]$ and $[\mathfrak{q}_3]$. We can check that none of these ideals are principal ($x^2 + 14y^2 = \pm 2$ or ± 3 have no solutions in $\mathbb{Z}$).

From the Dedekind–Kummer calculations, we know that

$$[\mathfrak{p}_2]^2 = [\langle 2\rangle] = [1], \quad [\mathfrak{p}_3][\mathfrak{q}_3] = [\langle 3\rangle] = [1].$$

The latter implies that

$$[\mathfrak{q}_3] = [\mathfrak{p}_3]^{-1}$$

so $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$. We still have to figure out if there is any relation between $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$.

One way to do this is to try looking for principal ideals whose norm is a product of small powers of 2 and 3. We find that $\mathcal{O}_K$ contains an element of norm 18:

$$\mathrm{Nm}_{K/\mathbb{Q}}(2 + \sqrt{-14}) = 4 + 14 = 18.$$

(How did we find this? Maybe just by luck/intelligent guesswork/trying lots of $x + y\sqrt{-14}$ until we find one whose norm is a product of powers of 2 and 3. If I set this as a question on an exam, there would be some sort of hint like: "Find an element of $\mathcal{O}_K$ of norm 18." or an earlier part of the question which asks to calculate the norm of $2 + \sqrt{-14}$.)

Let's factorise the principal ideal $\langle 2 + \sqrt{-14}\rangle$. Since its norm is a product of powers of 2 and 3, the only prime ideals which divide $\langle 2 + \sqrt{-14}\rangle$ must be factors of $\langle 2\rangle$ or $\langle 3\rangle$. That is,

$$\langle 2 + \sqrt{-14}\rangle = \mathfrak{p}_2^a\mathfrak{p}_3^b\mathfrak{q}_3^c.$$

(Note: we said earlier that $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$ are enough to generate $\mathrm{Cl}(K)$ because $[\mathfrak{q}_3] = [\mathfrak{p}_3]^{-1}$. That's a statement about *ideal classes*. Here we are doing a calculation with actual *ideals*, wo we need to keep $[\mathfrak{q}_3]$ in there.)

Comparing norms, and since $\mathrm{Nm}(\langle 2 + \sqrt{-14}\rangle) = 2 \times 3^2$, we get $a = 1$ and $b + c = 2$. We can't have $b = c = 1$ because then $\mathfrak{p}_3\mathfrak{q}_3 = \langle 3\rangle$ divides $\langle 2 + \sqrt{-14}\rangle$, but $3 \nmid (2 + \sqrt{-14})$ in $\mathcal{O}_K$. Thus $(b, c) = (2, 0)$ or $(0, 2)$.

Finally $2 + \sqrt{-14} \in \langle 3, -1 + \sqrt{-14}\rangle = \mathfrak{p}_3$, so $b > 0$. Thus

$$\langle 2 + \sqrt{-14}\rangle = \mathfrak{p}_2\mathfrak{p}_3^2.$$

In $\mathrm{Cl}(K)$, this tells us that

$$[\mathfrak{p}_2][\mathfrak{p}_3]^2 = [1]$$

or in other words $[\mathfrak{p}_2] = [\mathfrak{p}_3]^{-2}$. Hence $[\mathfrak{p}_3]$ alone is enough to generate $\mathrm{Cl}(K)$.

We have

$$[1] = [\mathfrak{p}_2]^2 = [\mathfrak{p}_3]^{-4}$$

and so $[1] = [\mathfrak{p}_3]^4$. So the order of $[\mathfrak{p}_3]$ in $\mathrm{Cl}(K)$ divides 4. On the other hand, since $\mathfrak{p}_2$ is not principal,

$$[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]^{-1} \neq [1]$$

so the order of $[\mathfrak{p}_3]$ does not divide 2. Thus the only possibility for the order of $[\mathfrak{p}_3]$ is 4. We deduce that $\mathrm{Cl}(K) \cong \mathbb{Z}/4\mathbb{Z}$, generated by $[\mathfrak{p}_3]$.

Class groups of quadratic fields (non-examinable).

We have calculated class groups for a few examples of imaginary quadratic fields $\mathbb{Q}(\sqrt{-d})$. When $d \to \infty$, $|\Delta_K| \to \infty$ and so $B_K \to \infty$. Hence there are more and more prime ideals which are generators of $\mathrm{Cl}(K)$.

These prime ideals are rarely principal: the equation $x^2 + dy^2 = \pm c$ doesn't have many solutions when d is large and c is small (compared with d). This also tells us that there are not many relations between these prime ideals, because relations involve finding some product of prime ideals which is principal (and whose norm is not too big). Thus we should expect the class group of $\mathbb{Q}(\sqrt{-d})$ to get bigger when d gets bigger.

That's not a rigorous argument, but it turns out to be true: $\#\,\mathrm{Cl}(\mathbb{Q}(\sqrt{-d})) \to \infty$ as $d \to \infty$. This is much harder than anything we prove in this module.

What about the smallest case, $\#\,\mathrm{Cl}(\mathbb{Q}(\sqrt{-d})) = \{1\}$? The above hard theorem tells us that there are only finitely many such d, but can we find them? Gauss found nine values of d in the early 19th century:

$$1, 2, 3, 7, 11, 19, 43, 67, 163$$

(I couldn't quite remember the list in the lecture!) He guessed that these are the only ones. Actually, number fields and ideals were not invented until much later, so Gauss asked an equivalent question about binary quadratic forms.

One can use the methods we have just developed to prove that all of these fields have trivial class group, but proving that there are no more is much harder. It was finally solved in the 1960s: the 9 listedabove are indeed the only ones.

The situation is different for real quadratic fields $\mathbb{Q}(\sqrt{d})$ $(d > 0)$. Again, as d gets larger, Lemma 62 gives more generators for $\mathrm{Cl}(K)$. But this time there is no obstacle to them being principal: the equation $x^2 - dy^2 = \pm c$ can have solutions even when d is big and c is small. So class groups for real quadratic fields do not grow in a predictable fashion.

There are lots of real quadratic fields with trivial class group: as far as we have calculated with computers, about 76% have trivial class group. It is a conjecture that there are infinitely many real quadratic fields with trivial class group, but this has not been proved!

Finiteness of the class group (examinable).

Starting from Minkowski's theorem on ideal classes, it is easy to prove that the class group is finite. We just need a lemma about ideals.

Lemma 63. *For any $B > 0$, there are only finitely many ideals in $\mathcal{O}_K$ of norm $\leq B$.*

Proof. For each positive integer $N \leq B$: Let $\mathfrak{a} \subseteq \mathcal{O}_K$ be an ideal of norm N. By Lemma 45, $N \in \mathfrak{a}$ so $\langle N \rangle \subseteq \mathfrak{a}$. Thus it is enough to show that there are only finitely many ideals in $\mathcal{O}_K$ containing $\langle N \rangle$.

By the Isomorphism Theorems for rings, there is a bijection

$$\{\text{ideals of } \mathcal{O}_K \text{ containing } \langle N \rangle\} \leftrightarrow \{\text{ideals of } \mathcal{O}_K/\langle N \rangle\}.$$

Since $\mathcal{O}_K/\langle N \rangle$ is a finite ring, it contains only finitely many ideals. □

Combining Theorem 61 and Lemma 63, we deduce that $\mathrm{Cl}(K)$ is finite (we will write this down formally at the start of the next lecture).

23. Proof of Minkowski's theorem

Finiteness of the class group.

We complete the proof that the class group is finite.

Theorem 64. *Let K be a number field. Then $\mathrm{Cl}(K)$ is finite.*

Proof. By Theorem 61, every ideal class has a representative of norm $\leq B_K$. By Lemma 63, there are only finitely many ideals with such a norm. So $\mathrm{Cl}(K)$ is finite. □

Consequently, the following definition makes sense.

Definition. For any number field K, the **class number** of K is the size of $\mathrm{Cl}(K)$ (it is often denoted h_K).

Proof of Minkowski's theorem (examinable).

We will spend the rest of this lecture, and all of the next lecture, proving Minkowski's theorem on ideal classes. Recall the theorem:

Theorem (Theorem 61)**.** *Every ideal class in $\mathrm{Cl}(K)$ has a representative $\mathfrak{a}$ which is an ideal of $\mathcal{O}_K$ (not just a fractional ideal) and satisfies $\mathrm{Nm}(\mathfrak{a}) \leq B_K$, where*

$$B_K = \left(\frac{4}{\pi}\right)^s \frac{n!}{n^n}\sqrt{|\Delta_K|}.$$

The proof relies on two other theorems of Minkowski:

Minkowski's theorem on ideal classes	
$\Uparrow$	(easy, examinable)
Minkowski's theorem on ideals	
$\Uparrow$	(harder, non-examinable)
Minkowski's theorem on lattices	

We first prove that Minkowski's theorem on ideals implies the theorem on ideal classes – this proof, and the statement of Minkowski's theorem on ideals, are examinable material for this module.

We will then outline the proof that Minkowski's theorem on lattices implies the theorem on ideals – this is non-examinable. A (slightly simplified) version of Minkowski's theorem on lattices was in Introduction to Number Theory, so we will skip the proof of the lattice theorem altogether.

Here's the statement of Minkowski's theorem on ideals.

Theorem 65 (Minkowski's theorem on ideals)**.** *Let K be a number field. Let $\mathfrak{a}$ be a non-zero ideal of $\mathcal{O}_K$. Then $\mathfrak{a}$ contains a non-zero element α such that*

$$|\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| \leq B_K \,\mathrm{Nm}(\mathfrak{a}).$$

We prove that this implies the theorem on ideal classes.

Proof of Theorem 61. (starting from Theorem 65)

There is one subtle point to watch out for in this proof – we apply Theorem 65 to a representative of the *inverse* of the ideal class we are interested in!

Let C be an ideal class in $\mathrm{Cl}(K)$. Pick a fractional ideal $\mathfrak{b} \in I_K$ which represents the *inverse* class C^{-1}. Thanks to Lemma 55, $\mathfrak{b} = \frac{1}{x}\mathfrak{c}$ for some ideal $\mathfrak{c} \subseteq \mathcal{O}_K$. Then $\mathfrak{c}$ also represents the ideal class C^{-1}.

By Theorem 65, we can find $\alpha \in \mathfrak{c} \setminus \{0\}$ such that $|\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| \leq B_K \,\mathrm{Nm}(\mathfrak{c})$.

Let $\mathfrak{a} = \alpha\mathfrak{c}^{-1}$. Then $[\mathfrak{a}] = [\mathfrak{c}]^{-1} = C$, $\mathfrak{a} \subseteq \mathcal{O}_K$ by the definition of $\mathfrak{c}^{-1}$, and

$$\mathrm{Nm}(\mathfrak{a}) = |\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| \cdot \mathrm{Nm}(\mathfrak{c})^{-1} \leq B_K. \qquad \square$$

Canonical embedding of a number field (non-examinable).

The proof of Minkowski's theorem on ideals (Theorem 65) is quite long. Surprisingly, it relies on geometry and analysis, even though we are proving a theorem which appears algebraic! That's how π appears in B_K. This method is called "geometry of numbers."

In order to use geometry to study a number field K, we need to embed K inside a real vector space. Let $n = [K : \mathbb{Q}]$. Then K is isomorphic to $\mathbb{Q}^n$, so we can embed it inside $\mathbb{R}^n$. We will do this using the real and complex embeddings of K.

K has n embeddings, but some of them might be complex so just using all n embeddings gives a map into $\mathbb{C}^n$ rather than $\mathbb{R}^n$. Instead we pair up the complex embeddings into complex conjugate pairs and use their real and imaginary parts:

Definition. Let K be a number field of signature (r, s). Label the real embeddings of K as $\sigma_1, \ldots, \sigma_r$ and the complex embeddings as $\sigma_{r+1}, \ldots, \sigma_{r+s}, \overline{\sigma_{r+1}}, \ldots, \overline{\sigma_{r+s}}$. The **canonical embedding** of K is the map $\iota\colon K \to \mathbb{R}^n$ defined by

$$\iota(\alpha) = (\sigma_1(\alpha), \ldots, \sigma_r(\alpha), \mathrm{Re}\,\sigma_{r+1}(\alpha), \mathrm{Im}\,\sigma_{r+1}(\alpha), \ldots, \mathrm{Re}\,\sigma_{r+s}(\alpha), \mathrm{Im}\,\sigma_{r+s}(\alpha)).$$

Since $n = r + 2s$, this does indeed map K into $\mathbb{R}^n$.

Lattices (non-examinable).

The image of the ring of integers under the canonical embedding forms a lattice in $\mathbb{R}^n$. This is defined as follows (a slight generalisation of the notion used in Introduction to Number Theory).

Definition. A **lattice** in $\mathbb{R}^n$ is a subgroup of $(\mathbb{R}^n, +)$ which is generated by a basis of $\mathbb{R}^n$.

e.g. $\mathbb{Z}^n$ is a lattice in $\mathbb{R}^n$ because it is generated by the standard basis.

Every lattice in $\mathbb{R}^n$ is isomorphic to $\mathbb{Z}^n$ as a group, but it might not be equal to $\mathbb{Z}^n$ as a different subgroup of $\mathbb{R}^n$. (e.g. $\{(x_1, \ldots, x_n) \in \mathbb{Z}^n : x_1 \equiv 0 \bmod 2\}$ is a lattice)

Definition. Let $L \subseteq \mathbb{R}^n$ be a lattice, generated by the basis $\{\underline{v}_1, \ldots, \underline{v}_n\}$. The **covolume** of L is the volume of the parallelepiped

$$\{x_1\underline{v}_1 + \cdots + x_n\underline{v}_n : x_1, \ldots, x_n \in \mathbb{R}, 0 \leq x_1, \ldots, x_n \leq 1\}.$$

The prefix "co" is here because this is not the volume of the lattice itself (that's zero because it is a countable set of points!); rather it is the volume of the "spaces in between the lattice."

Note that any lattice has many bases, and the covolume is the same whichever basis we choose. To prove this, we will use another way of defining the covolume: if C is the matrix with $\underline{v}_1, \ldots, \underline{v}_n$ as columns, then

$$\mathrm{covol}(L) = |\det(C)|.$$

(This is just the formula for the volume of a parallelepiped.) If we have two bases which generate the same lattice L, then the change-of-basis matrix between them has determinant ± 1, proving that the covolume of L is independent of the choice of basis.

Covolume of ideals in $\mathcal{O}_K$ (non-examinable).

For each ideal, its image under the canonical embedding is a lattice, and we can calculate its covolume in terms of the discriminant and the norm.

Lemma 66. *Let $\mathfrak{a}$ be a non-zero ideal in $\mathcal{O}_K$. Then $\iota(\mathfrak{a})$ is a lattice in $\mathbb{R}^n$ with covolume*

$$2^{-s}\sqrt{|\Delta_K|}\,\mathrm{Nm}(\mathfrak{a}).$$

Outline proof. Since $\mathfrak{a}$ is a subgroup of $\mathcal{O}_K$ of finite index $\mathrm{Nm}(\mathfrak{a})$, it suffices to prove that $\iota(\mathcal{O}_K)$ is a lattice with covolume

$$2^{-s}\sqrt{|\Delta_K|}$$

because $\mathrm{covol}(\iota(\mathfrak{a})) = [\mathcal{O}_K : \mathfrak{a}]\,\mathrm{covol}(\iota(\mathcal{O}_K))$. This formula for $\mathrm{covol}(\iota(\mathcal{O}_K))$ justifies our earlier claim that the discriminant measures a volume related to $\mathcal{O}_K$.

In order to prove the formula for $\mathrm{covol}(\iota(\mathcal{O}_K))$, let $\alpha_1, \ldots, \alpha_n$ be an integral basis for K. Let $C \in \mathrm{M}_{n\times n}(\mathbb{R})$ be the matrix with columns $\iota(\alpha_1), \ldots, \iota(\alpha_n)$.

We will calculate $\det(C)$. The calculation will show that $\det(C) \neq 0$, so $\iota(\alpha_1), \ldots, \iota(\alpha_n)$ form a basis of $\mathbb{R}^n$. Consequently the subgroup which they generate, namely $\iota(\mathcal{O}_K)$, is a lattice. Furthermore the covolume is given by $|\det(C)|$.

In order to calculate $\det(C)$, consider a different matrix $B \in \mathrm{M}_{n\times n}(\mathbb{C})$ with columns

$$\begin{pmatrix} \sigma_1(\alpha_i) \\ \vdots \\ \sigma_r(\alpha_i) \\ \sigma_{r+1}(\alpha_i) \\ \overline{\sigma_{r+1}}(\alpha_i) \\ \vdots \\ \sigma_{r+s}(\alpha_i) \\ \overline{\sigma_{r+s}}(\alpha_i) \end{pmatrix}.$$

(This is different from C because C involved taking real and imaginary parts, while B uses the complex embeddings directly.) By the definition of discriminant, we

have

$$\Delta_K = \det(B)^2.$$

To get from B to C, we multiply by a block diagonal matrix

$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \frac{1}{2} & \frac{1}{2} & & & \\ & & & -\frac{1}{2}i & \frac{1}{2}i & & & \\ & & & & & \ddots & & \\ & & & & & & \frac{1}{2} & \frac{1}{2} \\ & & & & & & -\frac{1}{2}i & \frac{1}{2}i \end{pmatrix}.$$

For each pair of complex embeddings, we have a block $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2}i & \frac{1}{2}i \end{pmatrix}$ because

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2}i & \frac{1}{2}i \end{pmatrix} \begin{pmatrix} z \\ \bar{z} \end{pmatrix} = \begin{pmatrix} \operatorname{Re} z \\ \operatorname{Im} z \end{pmatrix}.$$

The determinant of $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2}i & \frac{1}{2}i \end{pmatrix}$ is $\frac{1}{2}i$ so

$$\det(C) = (\tfrac{1}{2}i)^s \det(B).$$

Consequently

$$\operatorname{covol}(\iota(\mathcal{O}_K)) = |\det(C)| = 2^{-s}|\det(B)| = 2^{-s}\sqrt{|\Delta_K|}. \qquad \square$$

24. Proof of Minkowski's theorem, Part II

All of this lecture is non-examinable.

Last time, we defined the canonical embedding $\iota\colon \mathcal{O}_K \to \mathbb{R}^n$ and saw that, for any ideal $\mathfrak{a}$, $\iota(\mathfrak{a})$ is a lattice whose covolume we can calculate. In the formula

$$\operatorname{covol}(\iota(\mathfrak{a})) = 2^{-2}\sqrt{|\Delta_K|}\operatorname{Nm}(\mathfrak{a})$$

we see some but not all of the ingredients which appear in Minkowski's theorem on ideals.

Our aim is to show that $\mathfrak{a}$ contains a non-zero element whose norm is bounded in terms of the norm of the ideal. In other words, we want to show that the lattice $\iota(\mathfrak{a})$ intersects a set $S \subseteq \mathbb{R}^n$ consisting of elements of bounded norm.

Minkowski's theorem on lattices (non-examinable).

Suppose we have a lattice $L \subseteq \mathbb{R}^n$ and a compact set $S \subseteq \mathbb{R}^n$. If we make S big enough, can we guarantee that it contains an element of L? How big does it need to be?

This is a trick question: you can make S as big as you want without ever intersecting L by drawing a set which has holes round the lattice points! Even if you insist that S is no holes (i.e. it is simply connected), you still draw a set S which wiggles in and out around the lattice points. We can rule this out by insisting that S is convex.

Definition. A subset $S \subseteq \mathbb{R}^n$ is **convex** if for all $x, y \in S$ and all $t \in \mathbb{R}$ with $0 \leq t \leq 1$, we have $tx + (1-t)y \in S$.

It turns out that this is not enough. (You could take the line generated by one of the elements of the lattice, draw a very long, thin rectangle round that line and then translate that rectangle so that it lies between lattice points.) We need one more condition on S.

Definition. A subset $S \subseteq \mathbb{R}^n$ is **symmetric** if for all $x \in S$, we have $-x \in S$.

Now it looks like we have gone too far: A convex symmetric set S automatically contains $\frac{1}{2}x + \frac{1}{2}(-x) = 0$ for any $x \in S$, so $S \cap L$ is always non-empty. But 0 is not what we are looking for!

Thus the right question to ask is: if we make a convex symmetric set S large enough, can we guarantee that it contains an element of $L \setminus \{0\}$? How large we will need to make S obviously depends on how big the gaps between elements of L are, i.e. on $\operatorname{covol}(L)$.

The following theorem answers this question.

Theorem 67 (Minkowski's theorem on lattices)**.** *Let L be a lattice in $\mathbb{R}^n$. Let $S \subseteq \mathbb{R}^n$ be a compact, convex, symmetric set. If*

$$\operatorname{vol}(S) \geq 2^n \operatorname{covol}(L),$$

then S contains a non-zero element of L.

If you want to do Q11 on example sheet 4, you will need a slight variation of Theorem 67: S does not have to be compact, but then we have to replace the $\geq$ in the inequality for the covolume by $>$.

In Introduction to Number Theory, you proved Theorem 67 for lattices which are contained in $\mathbb{Z}^n$. One can reduce to that case by a linear transformation, so we will omit the proof of Theorem 67.

Proof of Minkowski's theorem on ideals (non-examinable).

Now $\iota(\mathfrak{a})$ is a lattice in $\mathbb{R}^n$, and we have calculated its covolume. In order to apply Theorem 67, we need to choose a compact, convex, symmetric set S. If $\underline{x} \in S \cap \iota(\mathfrak{a})$, then the fact that $\underline{x} \in \iota(\mathfrak{a})$ tells us that $\underline{x} = \iota(\alpha)$ for some $\alpha \in \mathfrak{a}$. So we want to choose S in such a way that $\iota(\alpha) \in S$ implies a bound for $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)$.

In a first attempt to do this, we extend $\mathrm{Nm}_{K/\mathbb{Q}}$ from a function on K to a continuous function on $\mathbb{R}^n$.

Define a function $N_{r,s} \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by

$$N_{r,s}(x_1, \ldots, x_r, y_1, z_1, \ldots, y_s, z_s) = |x_1| \cdots |x_r| \, (y_1^2 + z_1^2) \cdots (y_s^2 + z_s^2).$$

We have labelled the coordinates in this way in order to relate them to the canonical embedding of K: the x_is correspond to real embeddings of K, the y_is to the real part of complex embeddings, the z_is to the imaginary part of complex embeddings. This relation with the canonical embedding also explains why $N_{r,s}$ involves factors $y_i^2 + z_i^2$ which look like the norm of a complex number.

The significance of this function $N_{r,s}$ is that, for all $\alpha \in K$, we have

$$\begin{aligned} N_{r,s}(\iota(\alpha)) &= |\sigma_1(\alpha)| \cdots |\sigma_r(\alpha)| \, |\sigma_{r+1}(\alpha)|^2 \cdots |\sigma_{r+s}(\alpha)|^2 \\ &= |\sigma_1(\alpha)| \cdots |\sigma_r(\alpha)| \, |\sigma_{r+1}(\alpha)||\overline{\sigma_{r+1}}(\alpha)| \cdots |\sigma_{r+s}(\alpha)||\overline{\sigma_{r+s}}(\alpha)| \\ &= \prod_{i=1}^{n} |\sigma_i(\alpha)| = |\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)|. \end{aligned}$$

(The final step is Lemma 18).

Let

$$Y_{r,s}(T) = \{\underline{x} \in \mathbb{R}^n : N_{r,s}(\underline{x}) \leq T\}.$$

We want to show that $Y_{r,s}(B_K \, \mathrm{Nm}(\mathfrak{a}))$ contains a non-zero element of $\iota(\mathfrak{a})$.

Unfortunately we cannot deduce this directly from Theorem 67 because the set $Y_{r,s}(T)$ is usually not compact and, more importantly, not convex. In the lecture I drew pictures:

- for an imaginary quadratic field, $n = 2$, $(r, s) = (0, 1)$,
$$Y_{r,s}(T) = \{(y, z) \in \mathbb{R}^2 : y^2 + z^2 \leq T\}.$$
Thus $Y_{r,s}(T)$ is a disc, which is compact and convex. (This is the easy case!)
- for a real quadratic field, $n = 2$, $(r, s) = (2, 0)$,
$$Y_{r,s}(T) = \{(x_1, x_2) \in \mathbb{R}^2 : |x_1||x_2| \leq T\}.$$

This set is bounded by hyperbolae – it is the set A in Figure 10.1 of Stewart and Tall, p. 175. It is neither compact nor convex.

Instead, we choose a subset of $Y_{r,s}(T)$ which is compact and convex, and apply Theorem 67 to that. If we only want to prove the finiteness of the class group, we don't need to describe exactly which compact convex set we choose because we don't need an exact value for B_K, just that there exists some B_K – so we could just say that when we make T large enough, we know that $Y_{r,s}(T)$ will always contain a compact convex set which is large enough for Theorem 67 to apply.

On the other hand, in order to calculate class groups via Minkowski's theorem, we need a value for B_K. In order to get the best value we can, we choose the largest compact convex set we can inside $Y_{r,s}(T)$. In the real quadratic field case, a picture can be found in Stewart and Tall, Figure 10.1 (the new set is set B in the figure).

To define this new set, we define a new function $\phi_{r,s}\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by

$$\phi_{r,s}(x_1,\dots,x_r,y_1,z_1,\dots,y_s,z_s) = |x_1| + \cdots + |x_r| + 2\sqrt{y_1^2+z_1^2} + \cdots + 2\sqrt{y_s^2+z_s^2}.$$

(Adding the absolute values of the embeddings instead of multiplying them.) By the AM-GM inequality, we have

$$N_{r,s}(\underline{x})^{1/n} \leq \tfrac{1}{n}\,\phi_{r,s}(\underline{x}).$$

Consequently, the set

$$X_{r,s}(\lambda) = \{\underline{x} \in \mathbb{R}^n : T_{r,s}(\underline{x}) \leq \lambda\}$$

is contained in $Y_{r,s}((\frac{1}{n}\lambda)^n)$. Choosing $\lambda = n(B_K \operatorname{Nm}(\mathfrak{a}))^{1/n}$, we get $X_{r,s}(\lambda) \subseteq Y_{r,s}(B_K \operatorname{Nm}(\mathfrak{a}))$.

The set $X_{r,s}(\lambda)$ is compact, convex and symmetric. All that remains is to compute its volume. It turns out that

$$\operatorname{vol}(X_{r,s}(\lambda)) = 2^r\Big(\frac{\pi}{2}\Big)^s \frac{1}{n!}\,\lambda^n.$$

We will skip the calculation (which is quite fiddly, especially to get the correct powers of 2) but broadly speaking: the 2^r comes from integrating the x_is, because a bound on $|x_i|$ allows both positive or negative values of x_i, the $(\pi/2)^s$ comes from the circles defined by $2\sqrt{y_i^2+z_i^2}$, and the $1/n!\cdot\lambda^n$ comes from the fact that when we integrate a constant n times we get $1/n!\cdot t^n$.

Inserting our choice of λ into this formula, we get

$$\operatorname{vol}(X_{r,s}(\lambda)) = 2^r\Big(\frac{\pi}{2}\Big)^s \frac{n^n}{n!}\,B_K \operatorname{Nm}(\mathfrak{a}) = 2^{r+s}\sqrt{|\Delta_K|}\operatorname{Nm}(\mathfrak{a}) = 2^n \operatorname{covol}(\iota(\mathfrak{a})).$$

Hence by Theorem 67, $X_{r,s}(\lambda)$ contains a non-zero element $\underline{x} \in \iota(\mathfrak{a})$. Write $\underline{x} = \iota(\alpha)$ where $\alpha \in \mathfrak{a}$. Since $X_{r,s}(\lambda)$ is contained in (*), we have

$$|\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)| = N_{r,s}(\underline{x}) \leq B_K \operatorname{Nm}(\mathfrak{a})$$

as required.

(End of non-examinable material)

25. Mordell equation

This was an experimental lecture, and this is just a summary of what we did (there will be more careful notes for the next lecture).

We looked at question:

> Find all solutions $x, y \in \mathbb{Z}$ of the following equations:
> (a) $x^3 = y^2 + 1$;
> (b) $x^3 = y^2 + 13$;
> (c) $x^3 = y^2 + 23$.

We never actually looked at (c): it was a stretch goal.

To solve the equation $x^3 = y^2 + 1$, write

$$x^3 = (y + i)(y - i)$$

and use the fact that $\mathbb{Z}[i]$ is a UFD. Roughly, the argument goes like this: $y + i$ and $y - i$ are coprime and their product is a cube, so $y + i = \alpha^3$ and $y - i = \beta^3$ for some $\alpha, \beta \in \mathbb{Z}[i]$.

However justifying this turned out to be much more difficult than I had anticipated:

- In order to show that $y + i$ and $y - i$ are coprime, we say that any common factor would have to divide $(y + i) - (y - i) = 2i = (1 + i)^2$. But ruling out a common factor of $1 + i$ is more difficult. We'll see how to do it next time.
- Once we know they are coprime, we then look at prime factorisations (which we can do because $\mathbb{Z}[i]$ is a UFD). The problem here is units: the prime factorisations look like

  $$y + i = \delta\, p_1^{a_1} \cdots p_r^{a_r},$$
  $$y - i = \epsilon\, q_1^{b_1} \cdots q_s^{b_s},$$

  where $p_1, \ldots, p_r, q_1, \ldots, q_s$ are prime elements and δ, ϵ are units in $\mathbb{Z}[i]$. Since $y + i$ and $y - i$ are coprime, all the ps are distinct from all the qs. So the prime factorisation of x^3 is

  $$x^3 = \delta\epsilon\, p_1^{a_1} \cdots p_r^{a_r} q_1^{b_1} \cdots q_s^{b_s}$$

  which forces all the a_i and b_i to be multiples of 3. But in order to conclude that $y + i$, $y - i$ are cubes in $\mathbb{Z}[i]$, you still have to prove that δ, ϵ (which is true because the only units in $\mathbb{Z}[i]$ are $\pm 1, \pm i$ and each of these are cubes, but this was harder than I intended).

Anyway, once you get to $y + i = \alpha^3$, you can solve this: set $\alpha = u + iv$ and expand out the cube to get

$$y + i = u(u^3 - 3v) + iv(3u^2 - v^2).$$

Since $\{1, i\}$ are a $\mathbb{Q}$-basis for $\mathbb{Q}(i)$, the i coefficients on both sides are equal, leading to

$$1 = v(3u^2 - v^2).$$

Since $u, v \in \mathbb{Z}$, this forces $v = \pm 1$. Then:

- $v = +1$ leads to a contradiction.
- $v = -1$ leads to $u = 0$.

So the only solution is $(u, v) = (0, -1)$, giving $y = 0$ and $x = \pm 1$.

Now for

$$x^3 = y^2 + 13 = (y + \sqrt{-13})(y - \sqrt{-13}).$$

This time the class group of $\mathbb{Q}(\sqrt{-13})$ is $\mathbb{Z}/2\mathbb{Z}$, so $\mathbb{Z}[\sqrt{-13}]$ is not a UFD and the argument for showing that $y + \sqrt{-13}$ and $y - \sqrt{-13}$ are cubes won't work. But we can replace this by unique factorisation of ideals.

We have the equation of ideals

$$\langle x \rangle^3 = \langle y + \sqrt{-13} \rangle \langle y - \sqrt{-13} \rangle.$$

To solve this:

(1) Show that $\langle y + \sqrt{-13} \rangle$ and $\langle y - \sqrt{-13} \rangle$ are coprime (that is, there is no prime ideal which divides both of them).

Using unique factorisation of ideals into prime ideals, this implies that there are ideals $\mathfrak{a}, \mathfrak{b}$ such that

$$\langle y + \sqrt{-13} \rangle = \mathfrak{a}^3, \quad \langle y - \sqrt{-13} \rangle = \mathfrak{b}^3.$$

(Proof: every prime ideal in the prime factorisation of $\langle y + \sqrt{-13} \rangle$ and $\langle y - \sqrt{-13} \rangle$ must be raised to a power which is a multiple of 3.)

(2) Prove that $\mathfrak{a}$ is principal.

(3) Deduce an equation $y + \sqrt{-13} = (u + v\sqrt{-13})^3$ and solve it.

We'll do these steps in more detail in the next lecture.

Challenge before the next lecture: find a non-trivial solution to $x^3 = y^2 + 13$, without using a computer.

26. Mordell equation again and units

We will solve the equation

$$x^3 = y^2 + 13 = (y + \sqrt{-13})(y - \sqrt{-13}) \tag{1}$$

from last time.

Equation (1) implies the following equation of principal ideals:

$$\langle x \rangle^3 = \langle y + \sqrt{-13} \rangle \langle y - \sqrt{-13} \rangle. \tag{2}$$

We begin by showing that the ideals $\langle y + \sqrt{-13} \rangle$ and $\langle y - \sqrt{-13} \rangle$ are coprime.

Step 1. *$\langle y + \sqrt{-13} \rangle$ and $\langle y + \sqrt{-13} \rangle$ are coprime in $\mathbb{Z}[\sqrt{-13}]$.*

Suppose $\mathfrak{p}$ was a prime ideal which divides $\langle y + \sqrt{-13} \rangle$ and $\langle y - \sqrt{-13} \rangle$. Then

$$(y + \sqrt{-13}) - (y - \sqrt{-13}) = 2\sqrt{-13} \in \mathfrak{p}.$$

Hence $\mathrm{Nm}(\mathfrak{p})$ divides

$$\mathrm{Nm}_{K/\mathbb{Q}}(2\sqrt{-13}) = 4 \times 13 = 52.$$

Since $\mathrm{Nm}(\mathfrak{p})$ is a prime power (by Proposition 51), it must be 2, 4 or 13.

If $\mathrm{Nm}(\mathfrak{p}) = 13$, then 13 divides $\mathrm{Nm}_{K/\mathbb{Q}}(y + \sqrt{-13}) = y^2 + 13$. Consequently $13 \mid y$. But then $x^3 = y^2 + 13$ will be divisible by 13 but not by 13^2, which is impossible for a cube. Thus $\mathrm{Nm}(\mathfrak{p}) \neq 13$.

If $\mathrm{Nm}(\mathfrak{p}) = 2$ or 4, then 2 divides $\mathrm{Nm}_{K/\mathbb{Q}}(y + \sqrt{-13}) = y^2 + 13 = x^3$. So y is odd and x is even. Hence $y^2 \equiv 1 \bmod 8$ and $x^3 \equiv 1 \bmod 8$. This contradicts $x^3 = y^2 + 13 \equiv y^2 + 5 \bmod 8$.

We deduce that there are no prime ideals of $\mathcal{O}_K$ which divide both $\langle y + \sqrt{-13} \rangle$ and $\langle y + \sqrt{-13} \rangle$.

We will need the following lemma.

Lemma 68. *Let $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}_K$ be ideals which are coprime (have no common prime ideal factors). Suppose that*

$$\mathfrak{a}\mathfrak{b} = \mathfrak{c}^n$$

for some ideal $\mathfrak{c} \subseteq \mathcal{O}_K$ and some $n \in \mathbb{N}$. Then there are ideals $\mathfrak{a}', \mathfrak{b}' \subseteq \mathcal{O}_K$ such that

$$\mathfrak{a} = (\mathfrak{a}')^n, \quad \mathfrak{b} = (\mathfrak{b}')^n.$$

Proof. (On example sheet 4) This is an easy consequence of the unique factorisation of ideals (Theorem 50) – think about how you would prove the same result for rational integers. □

Step 2. *There is an ideal $\mathfrak{a} \subseteq \mathcal{O}_K$ such that $\langle y + \sqrt{-13} \rangle = \mathfrak{a}^3$.*

This follows immediately from Lemma 68, Step 1 and (2).

Step 3. *The ideal $\mathfrak{a}$ is principal.*

To prove this, we need to work out the class group. Using the Minkowski bound and Dedekind–Kummer theorem, we can calculate $\mathrm{Cl}(\mathbb{Q}(\sqrt{-13})) = \mathbb{Z}/2\mathbb{Z}$. From Step 2, we get the following equation in the class group:

$$[\mathfrak{a}]^3 = [\langle y + \sqrt{-13}\rangle] = [1].$$

Since the class group has order 2, $[\mathfrak{a}]^2 = [1]$ so we deduce that $[\mathfrak{a}] = [1]$.

Step 4. $y + \sqrt{-13} = (u + v\sqrt{-13})^3$ *for some* $u, v \in \mathbb{Z}$.

This is not quite an immediate consequence of Step 3 because we have to worry about units. More precisely, since $\mathfrak{a}$ is principal, we can write

$$\mathfrak{a} = \langle u + v\sqrt{-13}\rangle$$

for some $u, v \in \mathbb{Z}$. Then

$$\langle y + \sqrt{-13}\rangle = \langle u + v\sqrt{-13}\rangle^3$$

so

$$(u + v\sqrt{-13})^3 = \alpha(y + \sqrt{-13})$$

where α is a unit in $\mathbb{Z}[\sqrt{-13}]$. The only units in $\mathbb{Z}[\sqrt{-13}]$ are ± 1 (because a unit must have norm ± 1, and the only integer solutions to $x^2 + 13y^2 = \pm 1$ are $(\pm 1, 0)$). So we get

$$(u + v\sqrt{-13})^3 = \pm(y + \sqrt{-13}).$$

Multiplying u and v by -1 if necessary, we may assume WLOG that

$$(u + v\sqrt{-13})^3 = y + \sqrt{-13},$$

as we claimed in Step 4.

Step 5. *Find the possible values of u and v.*

Expanding out the equation from Step 4, we get

$$u^3 + 3u^2v\sqrt{-13} - 3 \times 13uv^2 - 13v^3\sqrt{-13} = y + \sqrt{-13}. \tag{3}$$

Because $\{1, \sqrt{-13}\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{-13})$, we can group together the terms to get

$$u(u^2 - 39v^2) = y, \tag{4}$$

$$v(3u^2 - 13v^2) = 1. \tag{5}$$

From (5), we deduce that $v = \pm 1$ (since $u, v \in \mathbb{Z}$).
If $v = +1$, then $3u^2 - 13 = 1$ so $3u^2 = 14$ which has no integer solutions.
If $v = -1$, then $3u^2 - 13 = -1$ so $3u^2 = 12$ so $u = \pm 2$.

Step 6. *Calculate x and y.*

Substituting $(u, v) = (\pm 2, -1)$ into (4), we get

$$y = u(u^2 - 39v^2) = \pm 2 \times (4 - 39) = \pm 70.$$

We could find x by substituting this back into the original equation:

$$x^3 = y^2 + 13 = 4913.$$

Of course, we could calculate $\sqrt[3]{4913}$, but I don't know $\sqrt[3]{4913}$ off by heart!

With a little more manipulation of algebraic numbers, we can avoid calculations involving big numbers. In fact, our original equation (1) can be rewritten as

$$x^3 = \mathrm{Nm}_{K/\mathbb{Q}}(y + \sqrt{-13}).$$

Since norms are multiplicative, this becomes

$$x^3 = \mathrm{Nm}_{K/\mathbb{Q}}(u + v\sqrt{-13})^3.$$

Both sides of this equation are in $\mathbb{Z}$, where the only cube root of 1 is 1 itself. Thus we get

$$x = \mathrm{Nm}_{K/\mathbb{Q}}(u + v\sqrt{-13}) = u^2 + 13v^2 = 4 + 13 = 17.$$

Conclusion.

Hence the only integer solutions to $y^2 = x^3 - 13$ are

$$x = 17, \quad y = \pm 70.$$

This equation has solutions which are rather large to find by manual brute force search. Of course a computer could have found them quickly by searching through values of x and y, but it could not prove that they are the only solutions. This method allows us to do both – find the solutions and prove that there are no more – entirely by hand.

Some harder examples.

This method can work for solving $x^3 = y^2 + k$ for many integers k, but we sometimes run into difficulties:

- For $x^3 = y^2 + 23$, we would look at ideals in $\mathbb{Q}(\sqrt{-23})$. The class group is $\mathbb{Z}/3\mathbb{Z}$ so the cube of every ideal is principal and Step 3 does not work.
- For $x^3 = y^2 - 2$, we get stuck at Step 4 because there are infinitely many units in $\mathcal{O}_K$ where $K = \mathbb{Q}(\sqrt{2})$. ($1 + \sqrt{2}$ is a unit in $\mathcal{O}_K$, and so are all of its powers.)

We will try to solve the second problem by studying more about units in $\mathcal{O}_K$.

Units in number fields.

Our final topic will be to understand the units in $\mathcal{O}_K$. Out of laziness I might say "unit of K" to mean "unit of $\mathcal{O}_K$" (the true units of K are not very interesting – just all the non-zero elements of K, because it is a field).

The following lemma shows that the property "α is a unit in $\mathcal{O}_K$" doesn't depend on which number field K we choose to look at (providing $\alpha \in K$ of course).

Lemma 69. *Let L/K be an extension of number fields and suppose that $\alpha \in K$. Then $\alpha \in \mathcal{O}_K^\times$ if and only if $\alpha \in \mathcal{O}_L^\times$.*

Proof. If $\alpha \in \mathcal{O}_K^\times$, then $\alpha \in \mathcal{O}_K \subseteq \mathcal{O}_L$ and $\alpha^{-1} \in \mathcal{O}_K \subseteq \mathcal{O}_L$ so $\alpha \in \mathcal{O}_L^\times$.

If $\alpha \in \mathcal{O}_L^\times$, then $\alpha, \alpha^{-1} \in \mathcal{O}_L$ so α, α^{-1} are both algebraic integers. We are given that $\alpha \in K$; since K is a field, this implies that $\alpha^{-1} \in K$ (note that $\alpha \neq 0$ because $\alpha \in \mathcal{O}_L^\times$). Thus α, α^{-1} are both algebraic integers contained in K i.e. they are both in $\mathcal{O}_K$. Hence $\alpha \in \mathcal{O}_K^\times$. □

Roots of unity.

The first examples of units in $\mathcal{O}_K$ are roots of unity.

Definition. Let K be a number field. We write

$$\mu_K = \{\zeta \in K : \zeta \text{ is a root of unity}\}.$$

If $\zeta \in \mu_K$, then ζ is a root of the monic polynomial $X^n - 1$ for some n, so $\zeta \in \mathcal{O}_K$. Also $\zeta^{-1} = \zeta^{n-1}$. Since $\mathcal{O}_K$ is closed under multiplication, we deduce that $\zeta^{-1} \in \mathcal{O}_K$. So $\zeta \in \mathcal{O}_K^\times$.

Lemma 70. $\mu_K \subseteq \mathcal{O}_K^\times$.

Proof. Let $\zeta \in \mu_K$. Then ζ is a root of $X^n - 1$ for some n, so ζ is an algebraic integer. Since $\zeta \in K$, we deduce that $\zeta \in \mathcal{O}_K$.

We have $\zeta^{-1} = \zeta^{n-1} \in \mathcal{O}_K$ (because $\mathcal{O}_K$ is a ring) so ζ is a unit in $\mathcal{O}_K$. □

Indeed, roots of unity are elements of $\mathcal{O}_K^\times$ of finite order w.r.t. multiplication; conversely, any element of $\mathcal{O}_K^\times$ of finite order is a root of unity.

We can easily work out μ_K in some cases.

Lemma 71. *If K has at least one real embedding, then $\mu_K = \{\pm 1\}$.*

Proof. Let σ be a real embedding of K. For $\zeta \in \mu_K$, we have $\zeta^n = 1$ for some n. Then $\sigma(\zeta) \in \mathbb{R}$ and $\sigma(\zeta)^n = 1$, which implies that $\sigma(\zeta) = \pm 1$.

But of course $\sigma(1) = 1$ and $\sigma(-1) = -1$. Since σ is injective, we must have $\zeta = \pm 1$. □

On the other hand, if all the embeddings of K are complex, then there is no shortcut to finding μ_K.

27. Dirichlet's Unit Theorem

Roots of unity in a number field.

Last time, we said that roots of unity are examples of units in $\mathcal{O}_K$. We can easily describe the structure of μ_K.

Lemma 72. *For any number field K, μ_K is a finite cyclic group under multiplication.*

Proof. It is clear that μ_K is closed under multiplication and under multiplicative inverses, and contains 1, so μ_K is a group under multiplication.

To prove that μ_K is finite: let ζ be a primitive n-th root of unity (i.e. $\zeta^n = 1$ but $\zeta^m \neq 1$ if $1 \leq m \leq n-1$). The minimal polynomial of ζ is the n-th cyclotomic polynomial $\Phi_n(X)$, which has degree $\phi(n)$ (Euler's totient function) – this was proved in Algebra 2. Hence $[\mathbb{Q}(\zeta) : \mathbb{Q}] = \phi(n)$. Furthermore, $\phi(n) \geq \sqrt{n}$ for all $n \geq 3$ (you can prove this using the formula for $\phi(n)$ in terms of the prime factorisation of n).

So if $\zeta \in K$, we must have

$$[K : \mathbb{Q}] > [\mathbb{Q}(\zeta) : \mathbb{Q}] \geq \sqrt{n}.$$

Since K is a fixed number field, this means that there are only finitely many possible values of n. In other words, there are only finitely many possible values for the order n of a root of unity in K.

Furthermore for each n, there are only finitely many n-th roots of unity. Hence K contains only finitely many roots of unity.

To show that μ_K is cyclic, we use another result from Algebra 2: any finite subgroup of the multiplicative group of a field is cyclic. □

Dirichlet's Unit Theorem.

Dirichlet's Unit Theorem gives us a description of $\mathcal{O}_K^\times$ as a group under multiplication. Firstly, it tells us that $\mathcal{O}_K^\times$ is a finitely generated abelian group (this is not obvious: after all, $K^\times$ is not finitely generated). The structure theory of finitely generated abelian groups then tells us that $\mathcal{O}_K^\times$ is isomorphic to the direct product of its torsion subgroup with $\mathbb{Z}^t$ for some t. The torsion subgroups is the group of roots of unity μ_K, whose structure we have just described. Dirichlet's Unit Theorem also tells us the number of copies of $\mathbb{Z}$ which occur in $\mathcal{O}_K^\times$.

Theorem 73. *Let K be a number field of signature (r, s). Let μ_K denote the set of roots of unity in K. Then $\mathcal{O}_K^\times$ is isomorphic to $\mu_K \times \mathbb{Z}^{r+s-1}$ as an abelian group (with the operation of multiplication).*

We will not prove this theorem. It uses Minkowski's theorem on lattices, but is somewhat harder than the theorem on the class group. There is a sketch of the proof in the last lecture of the 2018-19 notes if you are interested.

e.g. $K = \mathbb{Q}$: the signature is $(1, 0)$ so $r + s - 1 = 0$. Thus $\mathcal{O}_\mathbb{Q}^\times = \mu_\mathbb{Q} = \{\pm 1\}$. This matches what we already knew: $\mathbb{Z}^\times = \{\pm 1\}$.

Units in an imaginary quadratic field.

Let K be an imaginary quadratic field. The signature is $(0,1)$ so $r+s-1=0$. Hence Dirichlet's Unit Theorem tells us that $\mathcal{O}_K^\times = \mu_K$, and thus is finite.

(Note: $\mathbb{Q}$ and imaginary quadratic fields are the only number fields for which $r+s-1=0$ and so the only number fields for which $\mathcal{O}_K^\times$ is finite.)

We can work out $\mathcal{O}_K^\times$ exactly for each imaginary quadratic field. Note that the proof below does not rely on Dirichlet's Unit Theorem, so it actually proves Dirichlet's Unit Theorem for the case of imaginary quadratic fields.

Lemma 74. *Let d be a square-free positive integer and let $K = \mathbb{Q}(\sqrt{-d})$. Then $\mathcal{O}_K^\times = \mu_K$. More precisely,*

$$\mathcal{O}_K^\times = \begin{cases} \{\pm 1, \pm i\} & \text{if } d = 1 \\ \{\pm 1, \pm\zeta, \pm\zeta^2\} & \text{if } d = 3, \text{ where } \zeta = \exp(2\pi i/3) = \frac{-1+\sqrt{-3}}{2} \\ \{\pm 1\} & \text{if } d > 3 \text{ or } d = 2. \end{cases}$$

Proof. If $\alpha \in \mathcal{O}_K^\times$, then $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \pm 1$. So writing

$$\alpha = x + y\sqrt{-d}$$

where x, y are integers (if $-d \equiv 2, 3 \bmod 4$) or half-integers (if $-d \equiv 1 \bmod 4$), we get

$$x^2 + dy^2 = \pm 1.$$

Using the fact that squares are non-negative, we can solve this and get just the solutions listed in the statement of the lemma. □

Units of a real quadratic field.

Let K be a real quadratic field. The signature is $(2,0)$ so $r+s-1=1$. Hence

$$\mathcal{O}_K^\times \cong \mu_K \times \mathbb{Z} = \{\pm 1\} \times \mathbb{Z}$$

by Dirichlet's Unit Theorem and Lemma 71.

Thus $\mathcal{O}_K^\times$ is infinite. This matches what we saw for $\mathbb{Q}(\sqrt{2})$: $\{\pm(1+\sqrt{2})^n\}$ is a copy of $\{\pm 1\} \times \mathbb{Z}$ in $\mathcal{O}_K^\times$, though we do not yet know that this is all of $\mathcal{O}_K^\times$.

Another way to state the isomorphism above is: there exists $\varepsilon \in \mathcal{O}_K^\times$ such that

$$\mathcal{O}_K^\times = \{\pm\varepsilon^n : n \in \mathbb{Z}\}.$$

There may be more than one possible choice of ε such that $\mathcal{O}_K^\times$ has this form: indeed, $-\varepsilon$, ε^{-1} or $-\varepsilon^{-1}$ will also work. In fact, these are the only possibilities. This is basically a fact about the structure of the group $\{\pm 1\} \times \mathbb{Z}$.

Lemma 75. *Let K be a real quadratic field and suppose that*

$$\mathcal{O}_K^\times = \{\pm\varepsilon^n : n \in \mathbb{Z}\} = \{\pm\eta^n : n \in \mathbb{Z}\}.$$

Then

$$\eta \in \{\pm\varepsilon, \pm\varepsilon^{-1}\}.$$

Proof. From the defining property of ε and since $\eta \in \mathcal{O}_K^\times$, we have

$$\eta = \alpha\varepsilon^n$$

for some $n \in \mathbb{Z}$ and $\alpha \in \{\pm 1\}$.

Similarly from the defining property of η and since $\varepsilon \in \mathcal{O}_K^\times$, we have

$$\varepsilon = \beta\eta^m$$

for some $m \in \mathbb{Z}$ and $\beta \in \{\pm 1\}$.

Now

$$\varepsilon = \beta\alpha^m\varepsilon^{mn}$$

so

$$\eta^{mn-1} = \beta\alpha^m = \pm 1.$$

Since η is not a root of unity, this implies that $mn - 1 = 0$ and so $m = n = 1$ or $m = n = -1$. Thus $\eta = \pm\varepsilon$ or $\pm\varepsilon^{-1}$. □

We want to make a canonical choice among these four possible ε. To do this, we need to use the order on $\mathbb{R}$, and therefore we need to consider K as a subfield of $\mathbb{R}$. (We do this via the embedding for which $\sqrt{d} > 0$.)

Replacing ε by $-\varepsilon$ if necessary, we may assume that ε is positive. Then replacing ε by ε^{-1} if necessary, we may assume that $\varepsilon > 1$. We then have

$$-\varepsilon < -1 < -\varepsilon^{-1} < 0 < \varepsilon^{-1} < 1 < \varepsilon.$$

In conclusion we see that there is a *unique* ε such that

$$\mathcal{O}_K^\times = \{\pm\varepsilon^n : n \in \mathbb{Z}\} \text{ and } \varepsilon > 1.$$

We call this ε the **fundamental unit** of K.

28. Fundamental units and Pell's equation

Finding the fundamental unit.

The following proposition tells us how to find the fundamental unit.

Proposition 76. *Let $K = \mathbb{Q}(\sqrt{d})$ where $d > 1$ is a square-free integer. The fundamental unit of K is given by $x + y\sqrt{d}$ where (x, y) is the solution to*

$$x^2 - dy^2 = \pm 1$$

*with the **smallest** possible value of x, where*

- *x, y are positive integers if $d \equiv 2, 3 \bmod 4$,*
- *x, y are positive half-integers if $d \equiv 1 \bmod 4$.*

Proof. Let $\varepsilon = x + y\sqrt{d}$ be the fundamental unit.

Every unit $\eta > 1$ has the form ε^n where $n > 0$. Hence $\eta \geq \varepsilon$. So ε is the *smallest unit of K* greater than 1.

We have

$$\pm 1 = \mathrm{Nm}_{K/\mathbb{Q}}(\varepsilon) = (x + y\sqrt{d})(x - y\sqrt{d}).$$

Hence $x - y\sqrt{d} = \pm\varepsilon^{-1}$. Therefore the for elements $\pm x \pm y\sqrt{d}$ are in fact $\pm\varepsilon, \pm\varepsilon^{-1}$ in some order. Since $x + y\sqrt{d} = \varepsilon > 1$, it is the largest of these four elements. Consequently $x, y > 0$.

It remains to show that "smallest $\varepsilon > 1$" is equivalent to "smallest $x > 0$."

Suppose we had $\eta = u + v\sqrt{d} \in \mathcal{O}_K^\times$ such that $\eta > 1$ and $\eta \neq \varepsilon$. We want to show that $u > x$.

Now $\eta = \varepsilon^n$ for some $n > 1$. Expanding out

$$u + v\sqrt{d} = (x + y\sqrt{d})^n,$$

we get

$$u = x^n + \text{some positive terms}$$

(since x, y, d are all positive).

If $x > 1$, then $u \geq x^n > x$.

If $x = 1$, then $u > x^n = x$ because "some positive terms" is always a non-empty list of terms (since $n > 1$).

If $x = \frac{1}{2}$, then $u \geq x$ simply because u is a positive half-integer. So we only have to show that $u \neq \frac{1}{2}$. But $u^2 - dv^2 = \pm 1$. Now $(\frac{1}{2})^2 - dv^2 = +1$ has no solutions (since $(\frac{1}{2})^2 < 1$) and $(\frac{1}{2})^2 - dv^2 = -1$ has at most one positive solution, which just gives back the ε we already had. So $\eta > \varepsilon$ forces $u > x$. □

e.g. $1 + \sqrt{2}$ is a unit of $\mathbb{Q}(\sqrt{2})$ which is bigger than 1. It has $x = 1$, which is certainly the smallest possible value for a positive integer! So $1 + \sqrt{2}$ is the fundamental unit of $\mathbb{Q}(\sqrt{2})$, proving that

$$\mathbb{Z}[\sqrt{2}]^\times = \{\pm(1 + \sqrt{2})^n : n \in \mathbb{Z}\}.$$

e.g. We find the fundamental unit of $K = \mathbb{Q}(\sqrt{6})$.

Here $6 \equiv 2 \bmod 4$ so we look for solutions to $x^2 - 6y^2 = \pm 1$ with x, y positive integers.

If $x = 1$, then $6y^2 = 1 \pm 1 = 0$ or 2, leading to the solution $x = 1$, $y = 0$. But we are looking for y to be a positive integer, so this doesn't count. (We have just rediscovered that $1 \in \mathcal{O}_K^\times$, but it is a root of unity!)

If $x = 2, 3$ or 4, then there are no integer solutions for y because x^2 and $6y^2$ have a common factor.

If $x = 5$, then $6y^2 = 25 \pm 1 = 24$ or 26. This has a solution $y = 2$. Thus the fundamental unit of $\mathbb{Q}(\sqrt{6})$ is

$$5 + 2\sqrt{6}.$$

Pell's equation.

Pell's equation is the equation

$$x^2 - dy^2 = 1 \quad (x, y \in \mathbb{Z}).$$

A classical theorem tells us that there are infinitely many solutions of this equation (for any square-free integer $d > 1$).

We can use Dirichlet's unit theorem and the idea of fundamental units to prove this theorem, and to find the solutions for any given d.

Every solution to Pell's equation gives us a unit $x + y\sqrt{d}$ in $\mathbb{Q}(\sqrt{d})$, but not every unit gives a solution to Pell's equation for two reasons:

(1) the unit might have norm -1 instead of $+1$;
(2) the unit might have half-integer x and y (when $d \equiv 1 \bmod 4$.

In particular, knowing that $\mathcal{O}_K^\times$ is infinite does not immediately tell us that Pell's equation has infinitely many solutions.

Let's work out some examples.

$d = 6$

The fundamental unit $\varepsilon = 5 + 2\sqrt{6}$ in $\mathbb{Q}(\sqrt{6})$ has norm $+1$. Consequently

$$\mathrm{Nm}_{K/\mathbb{Q}}(\pm\varepsilon^n) = \mathrm{Nm}_{K/\mathbb{Q}}(\pm 1)\,\mathrm{Nm}_{K/\mathbb{Q}}(\varepsilon)^n = 1$$

for all $n \in \mathbb{Z}$. Also every power of ε lies in $\mathbb{Z} + \mathbb{Z}.\sqrt{6}$ (either by observing that $\varepsilon \in \mathbb{Z}[\sqrt{6}]$, or because we know that $\mathcal{O}_K = \mathbb{Z}[\sqrt{6}]$).

So all powers of ε give integer solutions of Pell's equation. In other words, the integer solutions of $x^2 - 6y^2 = 1$ are exactly

$$\{(x, y) : x + y\sqrt{6} = \pm(5 + 2\sqrt{6})^n \text{ for some } n \in \mathbb{Z}\}.$$

$d = 2$

The fundamental unit $\varepsilon = 1 + \sqrt{2}$ in $K = \mathbb{Q}(\sqrt{2})$ has norm -1. Consequently

$$\mathrm{Nm}_{K/\mathbb{Q}}(\pm\varepsilon^n) = \mathrm{Nm}_{K/\mathbb{Q}}(\pm 1)\,\mathrm{Nm}_{K/\mathbb{Q}}(\varepsilon)^n = (-1)^n.$$

In other words, $\mathrm{Nm}_{K/\mathbb{Q}}(\pm\varepsilon^n) = 1$ if and only if n is even. Meanwhile every power of ε lies in $\mathbb{Z}[\sqrt{2}]$.

Thus we conclude that the integer solutions of $x^2 - 2y^2 = 1$ are exactly

$$\{(x, y) : x + y\sqrt{2} = \pm(1 + \sqrt{2})^{2n} \text{ for some } n \in \mathbb{Z}\}.$$

We could write this a bit more neatly as

$$\{(x,y) : x + y\sqrt{2} = \pm(3+2\sqrt{2})^n \text{ for some } n \in \mathbb{Z}\}$$

(since $3 + 2\sqrt{2} = (1+\sqrt{2})^n$).

$d = 5$

Since $5 \equiv 1 \bmod 4$, when we look for the fundamental unit we have to allow half-integer solutions to $x^2 - 5y^2 = \pm 1$. Indeed, $x = y = \frac{1}{2}$ is a solution. Since $\frac{1}{2}$ is the smallest possible value of x, the fundamental unit of $K = \mathbb{Q}(\sqrt{5})$ is

$$\varepsilon = \frac{1+\sqrt{5}}{2}.$$

Let's calculate ε^3 (Why? I will say more about that below.)

$$\varepsilon^3 = \frac{1 + 3\sqrt{5} + 3 \times 5 + 5\sqrt{5}}{8} = \frac{16 + 8\sqrt{5}}{8} = 2 + \sqrt{5}.$$

Thus $\varepsilon^3 \in \mathbb{Z}[\sqrt{5}]$ We deduce that $(2+\sqrt{5})^n$ is also in $\mathbb{Z}[\sqrt{5}]$ for all $n \in \mathbb{Z}$ (for positive n this is obvious; for negative n, use that $(2+\sqrt{5})^{-1} = -2 + \sqrt{5}$).

So $x + y\sqrt{5} = \varepsilon^n$ gives an integer solution to Pell's equation whenever n is both a multiple of 3 (so that $x, y \in \mathbb{Z}$) and even (so that $\mathrm{Nm}(\varepsilon^n) = +1$). Thus the set of integer solutions to $x^2 - 5y^2 = 1$ is

$$\{(x,y) : x + y\sqrt{5} = \pm\varepsilon^{6n}, n \in \mathbb{Z}\}$$
$$= \{(x,y) : x + y\sqrt{5} = \pm(9+4\sqrt{5})^n, n \in \mathbb{Z}\}.$$

(using $\varepsilon^6 = (2+\sqrt{5})^2 = 9 + 4\sqrt{5}$)

This proof was not quite complete: we proved that if n is a multiple of 6, then $x + y\sqrt{5} = \pm\varepsilon^n$ is an integer solution to Pell's equation, but we did not prove that if $x + y\sqrt{5} = \pm\varepsilon^n$ gives an integer solution, then n has to be a multiple of 6. It is clear that n has to be even, so that $\mathrm{Nm}_{K/\mathbb{Q}}(\varepsilon^n) = +1$, but we have not shown that if $x, y \in \mathbb{Z}$ then n is a multiple of 3. To prove this: suppose that $\varepsilon^n \in \mathbb{Z}[\sqrt{5}]$ and write $n = 3a + b$ where $a \in \mathbb{Z}$, $b \in \{0,1,2\}$. Now $\varepsilon^{-3a} \in \mathbb{Z}[\sqrt{5}]$ so $\varepsilon^b = \varepsilon^n\varepsilon^{-3a} \in \mathbb{Z}[\sqrt{5}]$. But $\varepsilon^1 = \frac{1+\sqrt{5}}{2} \notin \mathbb{Z}[\sqrt{5}]$ and $\varepsilon^2 = \frac{3+\sqrt{5}}{2} \notin \mathbb{Z}[\sqrt{5}]$.

Why did look at ε^3 in the case $d = 5$? Well, it turns out that ε^3 is always in $\mathbb{Z}[\sqrt{d}]$ (where ε is the fundamental unit). For any given example, you can verify this by calculation. The general proof that it works for all d is not part of the course.

(Non-examinable) One way to prove this is to write $\varepsilon = \frac{u+v\sqrt{d}}{2}$, expand out ε^3 and do some calculations mod 8. This is a bit long: you have to start by showing that if u, v odd (which is the only case that matters!) and satisfy $\frac{u^2-dv^2}{4} = \pm 1$, then $d \equiv 5 \bmod 8$.

Another way is to use "arithmetic modulo $\langle 2\rangle$" in $\mathcal{O}_K$. This uses a generalisation of Fermat's Little Theorem to $\mathcal{O}_K$: $x^{\mathrm{Nm}(\mathfrak{p})-1} \equiv 1 \bmod \mathfrak{p}$ for every $x \in \mathcal{O}_K \setminus \mathfrak{p}$ and every prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$. When $\langle 2\rangle$ is prime in $\mathcal{O}_K$, this tells us that $x^3 \equiv 1 \bmod 2$ for all $x \notin \langle 2\rangle$ and hence $\varepsilon^3 \in \mathbb{Z}[\sqrt{d}]$. When $\langle 2\rangle = \mathfrak{pq}$, then you show that the fact that $\varepsilon \notin \mathfrak{p}$ or $\mathfrak{q}$ forces $\varepsilon \equiv 1 \bmod \langle 2\rangle$.