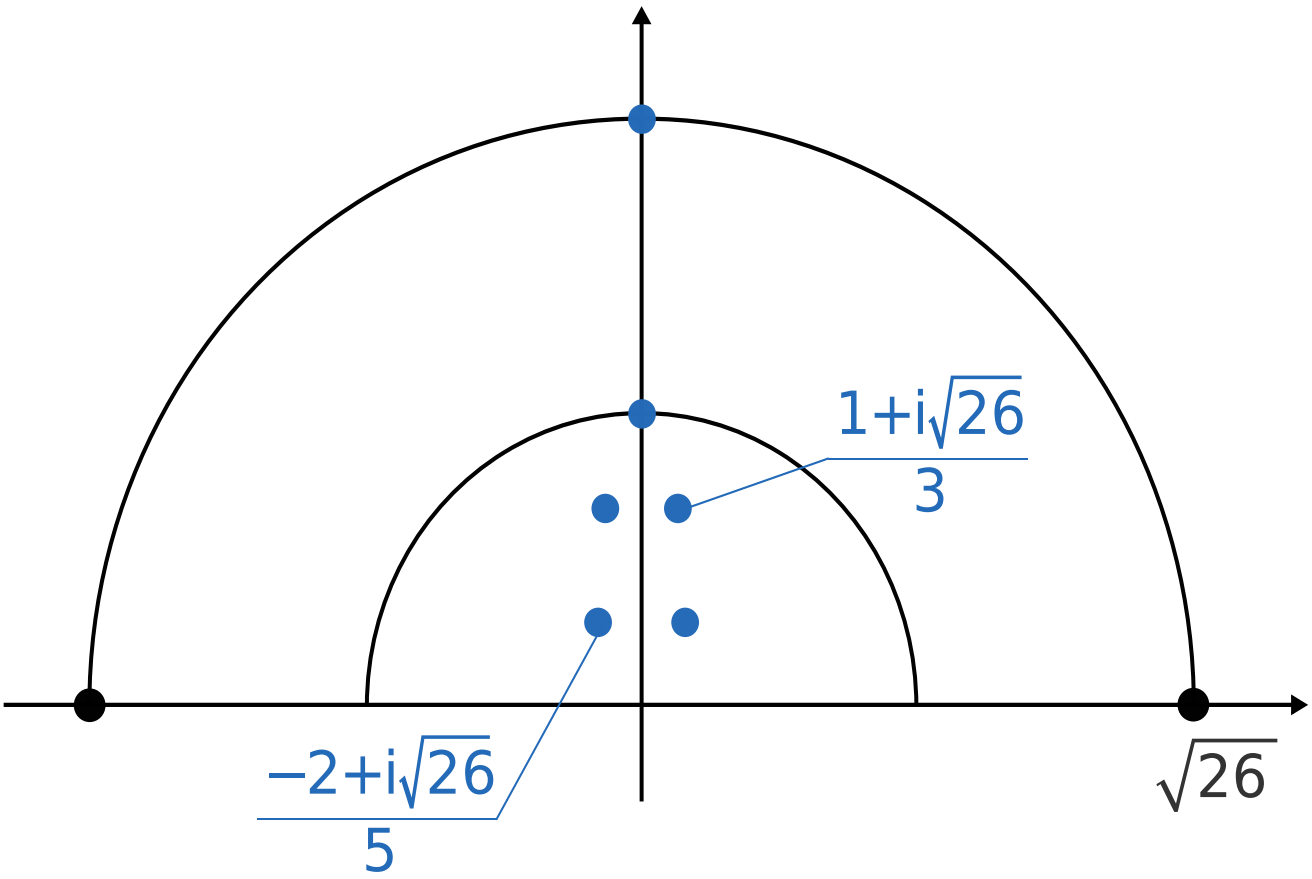
LECTURE NOTES - ALGEBRAIC NUMBER THEORY

MATTEO TAMIOZZO



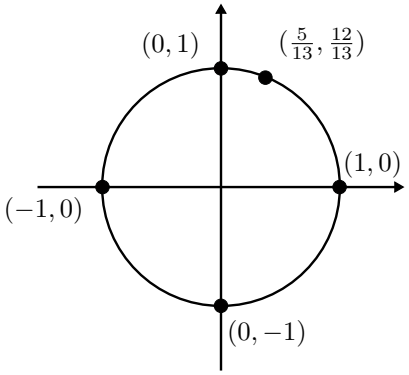
These notes benefitted from lecture notes for previous versions of the Algebraic Number Theory course at Imperial College by Ambrus Pal, Toby Gee, Kevin Buzzard, Ana Caraiani, as well as from the notes [26].

This text gives an introduction to Algebraic Number Theory with emphasis on quadratic fields, using Diophantine equations - especially Mordell equations and representation of numbers by quadratic forms - as a motivation throughout.

Prologue

One of the main aims of Number Theory is to study integral or rational solutions of Diophantine equations. Here is a visual introduction to some of those which we will examine in this course.

Circles. Let us start with the equation $X^2 + Y^2 = 1$. Its integral solutions are $(\pm 1, 0), (0, \pm 1)$. What about rational solutions? We can think of them geometrically as being points with rational coordinates on the unit circle centred at the origin; five of them are depicted in the following picture:



We can also enlarge the radius of the circle; for example, we can look for integral solutions of equations of the form $X^2 + Y^2 = p$ for an odd prime p. Here is a picture for $p = 3, 5, 7, 11, 13$: green (resp. red) circles are those having (resp. not having) integral points.

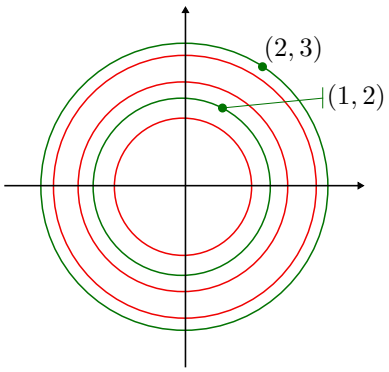
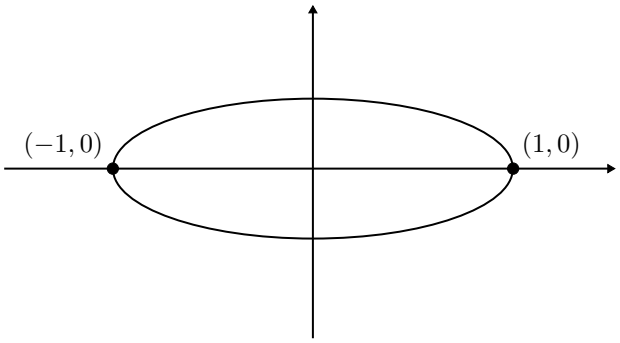


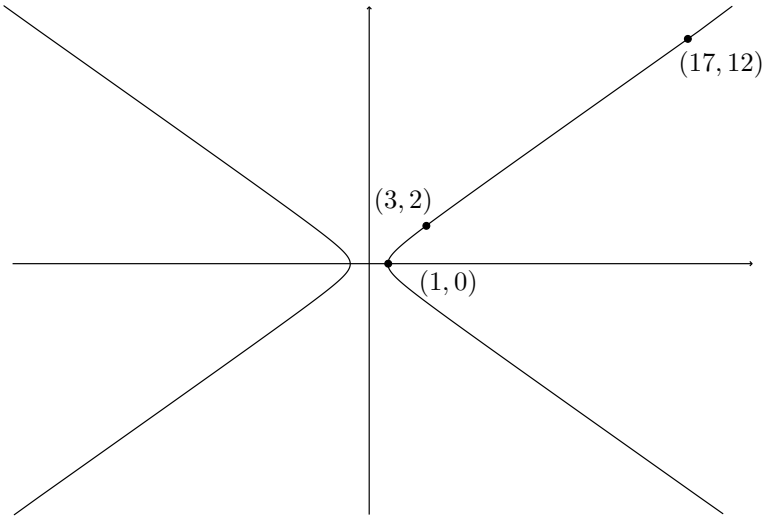
Figure 1. $X^2 + Y^2 = 3, 5, 7, 11, 13$.

Is there a general rule telling us which colour should the circle $X^2+Y^2 = p$ have for an arbitrary prime p? And what about rational points on these circles?



Ellipses. Let us now look at ellipses. For example, the integral points on the ellipse $X^2+2Y^2=1$ are $(\pm 1, 0)$; what about rational points? More generally, given two positive integers n, r, there are *finitely many* integral points on the ellipse $X^2+nY^2=r$, and one can list all of them. For example for $r=1$ and $n>1$ the only integral points are $(\pm 1, 0)$; a more interesting problem, generalising our previous question on circles, is what happens for fixed n as r varies.

Hyperbolas. The situation is quite different for hyperbolas: for example, let us consider the equation $X^2-2Y^2=1$.



In the picture you see three integral solutions; notice that unlike the case of the ellipse, the hyperbola is unbounded, hence just staring at the picture it is unclear whether or not there are finitely many integral points. In fact, we will see that there are *infinitely many* integral solutions of the above equation, and we will be able to describe them.

Mordell curves. Let us now consider equations of degree 3, starting from $Y^2+2=X^3$. Fermat asked as a challenge to contemporary mathematicians to prove that the only integral solutions are those in the left picture below. We will show that this is indeed the case. On the other hand, if we switch sign and consider the equation $Y^2-2=X^3$ then we can easily find the integral solutions $(-1, \pm 1)$; but are there other integral solutions? Are there finitely many of them? This question was raised in 1860, and answered affirmatively by Landau at the beginning of the 20^{th} century. We will study *Mordell equations* $Y^2+k=X^3$ throughout this course, and eventually prove a general *finitess result* about their integral solutions, which will be the hardest theorem in this course. *Rational points* on Mordell curves also have an interesting structure, which we will briefly mention.

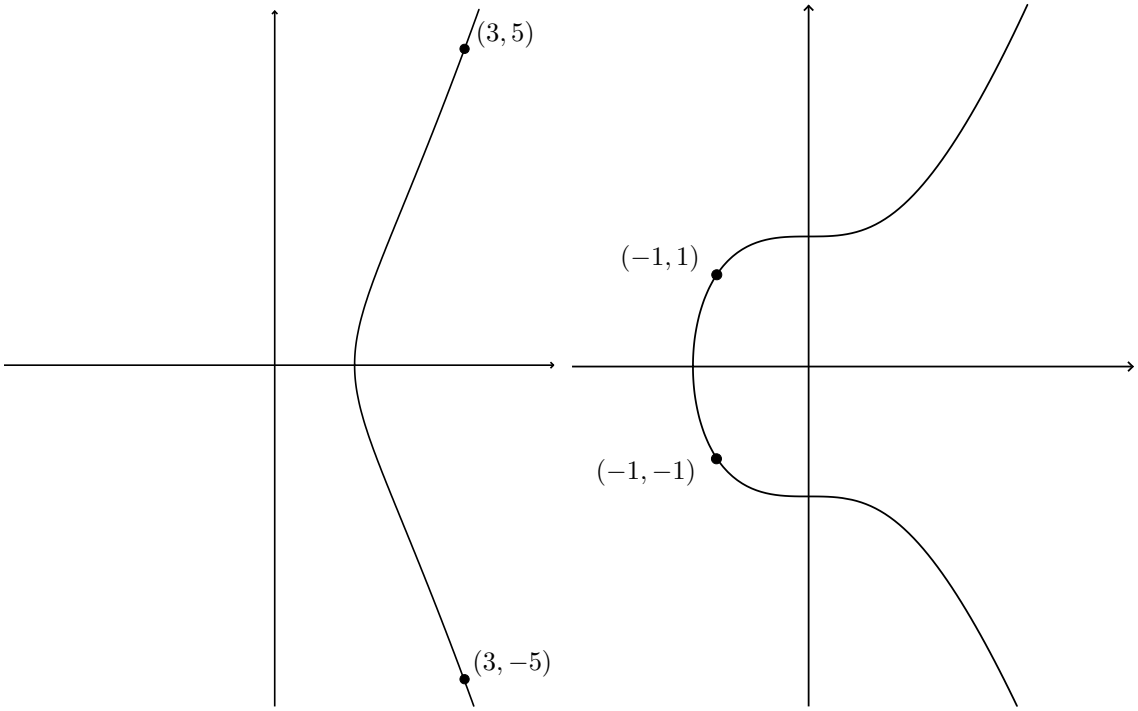


FIGURE 2. $Y^2 + 2 = X^3$ and $Y^2 - 2 = X^3$.

The main themes. The basic method we will use to solve equations like $Y^2 \pm 2 = X^3$ is to factor the left hand side and write the equation as $(Y + \sqrt{\pm 2})(Y - \sqrt{\pm 2}) = X^3$; this leads us to work with the ring $\mathbb{Z}[\sqrt{\pm 2}]$. More generally, in order to solve Mordell equations $Y^2 + k = X^3$ we will need to study *quadratic rings*; most of this course will be devoted to understanding their arithmetic properties. We will see that quadratic rings are closely related to *quadratic forms*, such as $X^2 \pm 2Y^2$, hence their study will also give us information about integral points on ellipses and hyperbolas. The latter also give rise to rational approximations of *quadratic irrational numbers*: for example, integral solutions (x, y) of the equation $X^2 - 2Y^2 = 1$ approach, as y gets larger and larger, one of the lines $X = \pm\sqrt{2}Y$. This will lead us to study *Diophantine approximation*, which will also be the key tool to study integral points on Mordell curves such as $Y^2 - 2 = X^3$.

Contents

1. Lecture 1: Right-angle triangles

In this lecture we will recall the main arithmetic properties of the integers and use them to study Pythagorean triples. We will also give an overview of the aims of this course.

1.1. **Introduction: two questions about right-angle triangles.** Number Theory studies the properties of the integers

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

with the aim of solving *Diophantine equations*: given a polynomial

$$P(X_1, \dots, X_n) \in \mathbb{Z}[X_1, \dots, X_n]$$

we would like to know:

(1) Is the set of solutions $\{(a_1, \dots, a_n) \in \mathbb{Z}^n \mid P(a_1, \dots, a_n) = 0\}$ non empty?
(2) Is this set finite or infinite?
(3) If it is finite, how many elements does it contain?

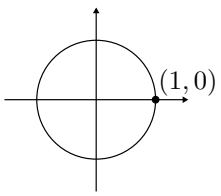
Similar questions may be asked about the set $\{(a_1, \dots, a_n) \in \mathbb{Q}^n \mid P(a_1, \dots, a_n) = 0\}$ of *rational solutions* of the equation $P = 0$.

The name "Diophantine equations" comes from *Diophantus* (3rd century CE), who in his book *Arithmetica* listed 130 problems asking to find integral solutions to several equations. For example, he asked:

Question 1.1.1. (Diophantus, Arithmetica, Book II, Problem 8) Split a given square in two squares. In other words, find the solutions in $\mathbb{Z}$ of the equation

$$X^2 + Y^2 = Z^2.$$

In view of Pythagoras' theorem, we are asking to find all the right-angle triangles whose sides have integral length. Equivalently, we have to find the *rational points* on the unit circle centred at the origin:



Another ancient question about right-angle triangles is the following

Question 1.1.2. (Arab manuscripts, 10th century) Which integers $n > 0$ are areas of a right-angle triangle whose sides have *rational* length? In other words, for which values of n does the system

$$\begin{cases} X^2 + Y^2 = Z^2 \\ \frac{XY}{2} = n \end{cases}$$

have rational solutions?

One of the aims of this course is to study a technique to solve certain Diophantine equations; the starting point of our investigations is the property of *unique factorisation* of integers as a product of primes. After proving it, we will use this property to answer question 1.1.1.

1.2. **Basic properties of $\mathbb{Z}$.** In this section we will recall the key arithmetic properties of the integers. We denote by $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$ the set of non negative integers; if $a, b \in \mathbb{Z}$ we say that a *divides* b, and we write $a \mid b$, if there is an integer $c \in \mathbb{Z}$ such that $b = ac$. If a does not divide b we write $a \nmid b$. A *prime number* is an integer $p > 1$ whose only divisors are $\pm 1, \pm p$.

The first crucial property of $\mathbb{Z}$ is the following

Key fact 1.2.1. *The set $\mathbb{Z}_{\geq 0}$ does not contain any infinite descending sequence*

$$a_1 > a_2 > \ldots > a_k > \ldots;$$

equivalently, every non empty subset of $\mathbb{Z}_{\geq 0}$ contains a minimum.

Let us deduce from this the main properties of $\mathbb{Z}$:

Existence of prime factorisation: Let $n \in \mathbb{Z} \smallsetminus \{0\}$. Then there exist prime numbers $p_1, \ldots, p_r$ (not necessarily distinct) such that

$$n = \pm p_1 p_2 \cdots p_r.$$

Proof. We may assume that $n > 0$. If n is prime or $n = 1$ we are done. Otherwise we can write $n = a_1 a_2$ with $1 < a_1, a_2 < n$ and repeat the argument with a_1 and a_2. By 1.2.1 the process must stop, yielding the desired factorisation. □

Euclid algorithm: Let $a, b \in \mathbb{Z}$ with $b \neq 0$. There exist unique $q, r \in \mathbb{Z}$ such that

$$a = bq + r \text{ with } 0 \leq r < |b|.$$

Proof. Consider the set

$$\mathscr{S} = \{n \in \mathbb{Z}_{\geq 0} \mid n = a - sb, s \in \mathbb{Z}\}.$$

By 1.2.1 we know that $\mathscr{S}$ has a minimum r. Then r is of the form $a - qb$ for some $q \in \mathbb{Z}$. Since $r \in \mathscr{S}$ we have $r \geq 0$. To show that $r < |b|$ we argue by contradiction: if this is not true, then $r - |b| \geq 0$ hence $r - |b| \in \mathscr{S}$. As $r - |b| < r$ we find a contradiction. This proves existence of q, r with the required properties. Uniqueness is left as an exercise. □

Bézout property: Let $a, b \in \mathbb{Z}$ which are not both equal to 0. There exists a unique *positive* integer d such that

$$\{ax + by \mid x, y \in \mathbb{Z}\} = d\mathbb{Z} = \{dx, x \in \mathbb{Z}\}. \tag{1.2.1.1}$$

Such a d divides both a and b. Furthermore, if c is any integer dividing a and b then c divides d. We call d the *greatest common divisor* of a and b and write

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

Proof. The set

$$I = \{ax + by \mid x, y \in \mathbb{Z}\}$$

has a smallest positive element $d = ra + sb$. We claim that such a d satisfies (1.2.1.1). By construction $d\mathbb{Z} \subset I$. To prove the other inclusion let $c \in I$; then we can write $c = dq + r$ with $0 \leq r < d$. In particular r belongs to I. By minimality of d we must have $r = 0$, hence $c \in d\mathbb{Z}$. This proves the existence of d as in (1.2.1.1). Such a d divides a, b; moreover if $c \mid a, b$ then c divides every element in I, hence $c \mid d$.

Finally, if $d' \in \mathbb{Z}$ is another element satisfying (1.2.1.1) we deduce that $d \mid d'$ and $d' \mid d$, hence $d = \pm d'$. Hence d is uniquely determined by the requirement that it is positive. □

Euclid's Lemma: Let p be a prime and $a, b \in \mathbb{Z}$. Then

$$p \mid ab \Rightarrow p \mid a \text{ or } p \mid b.$$

Proof. Assume that $p \mid ab$ and $p \nmid a$. Then $\gcd(a, p) = 1$ hence by the Bézout property we can write $1 = ax + py$ for some $x, y \in \mathbb{Z}$. It follows that $b = abx + bpy$, hence $p \mid b$. □

Uniqueness of factorisation: Let $n \in \mathbb{Z} \smallsetminus \{0\}$ and $p_1, \ldots, p_r$, $q_1, \ldots, q_s$ be primes such that

$$p_1 \cdot p_2 \cdot \cdots \cdot p_r = \pm n = q_1 \cdot q_2 \cdot \cdots \cdot q_s;$$

then $r = s$ and, up to reordering the q_i's, we have $q_i = p_i$ for $i = 1, \ldots, r$.

Proof. Since $p_1 \mid \pm n = q_1 \cdot q_2 \cdot \cdots \cdot q_s$ Euclid's lemma implies that p_1 divides one of the q_i's. Up to reordering we may assume $p_1 \mid q_1$, hence $p_1 = q_1$ and $p_2 \cdot \cdots \cdot p_r = q_2 \cdot \cdots \cdot q_s$. Iterating the argument proves the claim. □

Separating powers: Let $a, b, c \in \mathbb{Z} \smallsetminus \{0\}$ and $n > 0$ such that

$$a^n = bc.$$

If $\gcd(b, c) = 1$ then there exist $b_1, c_1 \in \mathbb{Z}$ such that

$$b = \pm b_1^n \text{ and } c = \pm c_1^n.$$

Remark 1.2.2. The last of the above properties, which is a consequence of unique factorisation, is one of the main elementary tricks to solve Diophantine equations. We will use it repeatedly.

Notation 1.2.3. Given a positive integer n and two integers $a, b \in \mathbb{Z}$ we say that a is *congruent* to b *modulo* n, and we write

$$a \equiv b \pmod{n}$$

if $n \mid (a - b)$ (in other words, if a and b give the same remainder when divided by n). We denote by

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], [2], \ldots, [n-1]\}$$

the set of residue classes modulo n. Defining addition as $[a] + [b] = [a + b]$ and multiplication as $[a] \cdot [b] = [ab]$ makes $\mathbb{Z}/n\mathbb{Z}$ into a ring.

Beware: you are strongly encouraged to avoid using the symbol $\mathbb{Z}_n$ to denote the ring $\mathbb{Z}/n\mathbb{Z}$. While such a notation can sometimes be found in books, nowadays the symbol $\mathbb{Z}_n$ almost always refers to a very different object (which is extremely important in number theory, but will not be discussed in this course). An example for the curious reader: the ring $\mathbb{Z}_{10}$ consists of numbers having *infinite* decimal expansion *on the left*. For example, letting $a = \ldots 111111 \in \mathbb{Z}_{10}$, we have

$$\begin{aligned} 9a + 1 &= \ldots 999999 + \\ &\quad \ldots 000001 = \\ &\quad \ldots 000000, \end{aligned}$$

hence in $\mathbb{Z}_{10}$ the following equality holds

$$-\frac{1}{9} = \frac{1}{1 - 10} = \ldots 111111 = \sum_{i=0}^{\infty} 10^i.$$

1.3. **Pythagorean triples.** We can now answer question 1.1.1. We wish to find all the numbers $a, b, c \in \mathbb{Z} \smallsetminus \{0\}$ such that

$$a^2 + b^2 = c^2 \text{ (e. g. } (3, 4, 5), (33, 56, 65), (132, 1085, 1093)).$$

Notice that, if $\lambda \in \mathbb{Z}$ and $a^2 + b^2 = c^2$ then $(\lambda a)^2 + (\lambda b)^2 = (\lambda c)^2$. Hence we may (and will) restrict to $\gcd(a, b) = 1$ (which implies $\gcd(a, b) = \gcd(b, c) = \gcd(a, c) = 1$). Up to changing signs to a, b, c we may also assume $a, b, c > 0$. One of a, b is odd, and we may assume $2 \nmid a$. We will call such a triple a *primitive Pythagorean triple*. Then

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

Let $d = \gcd(c - b, c + b)$. Then $d \mid \gcd(2c, 2b) = 2\gcd(c, b) = 2$. On the other hand $d \mid a$ and a is odd. Hence we must have $d = 1$. By the "separating powers trick" we know that there exist positive integers a_1, a_2 such that $c + b = a_1^2$ and $c - b = a_2^2$. Hence $a = a_1 a_2$; in particular a_1 and a_2 are odd. Let

$$u = \frac{a_1 + a_2}{2},\ v = \frac{a_1 - a_2}{2} \in \mathbb{Z};$$

then we find

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

and (a, b, c) is primitive if (and only if) $u > v$, $\gcd(u, v) = 1$ and $u \not\equiv v \pmod{2}$. We have proved

Theorem 1.3.1. *Every primitive Pythagorean triple is of the form*

$$(u^2 - v^2, 2uv, u^2 + v^2)$$

for some $u, v > 0$ *such that* $u > v$, $\gcd(u, v) = 1$ *and* $u \not\equiv v \pmod{2}$.

1.4. **Congruent numbers.** Let us now consider question 1.1.2. An integer n which is the area of a right-angle triangle whose sides have rational length is called a *congruent number*. If n is such a congruent number and $\lambda \in \mathbb{Z}$ then $\lambda^2 n$ is also a congruent number; hence we may restrict to square-free values of n. Let Δ be a right-angle triangle with sides of length $a, b, c \in \mathbb{Q}$ and area n. Then we have

$$n = \frac{ab}{2}$$
$$(a, b, c) = \lambda(u^2 - v^2, 2uv, u^2 + v^2)$$

for some $\lambda \in \mathbb{Q} \smallsetminus \{0\}$ and $u, v \in \mathbb{Z}$ as in Theorem 1.3.1. Hence

$$n = \lambda^2 uv(u^2 - v^2) = \lambda^2 v^4 \frac{u}{v}\left(\frac{u^2}{v^2} - 1\right).$$

Letting $t = \frac{u}{v}$ and $s = \lambda v^2$ we see that the numbers t, s satisfy $\frac{1}{s^2} n = t^3 - t$. From this one deduces the following lemma, whose proof is left to the reader.

Lemma 1.4.1. *A square-free positive integer n is a congruent number if and only if the equation*

$$nY^2 = X^3 - X$$

has a solution $(x, y) \in \mathbb{Q}^2$ with $y \neq 0$.

Therefore our original question 1.1.2 has been translated into the following

Question 1.4.2. For which positive square-free numbers n does the equation $nY^2 = X^3 - X$ have a solution $(x, y) \in \mathbb{Q}^2$ with $y \neq 0$?

In fact, this is currently an *open question* - arguably the oldest open question in Mathematics - and object of active research! For example, the following conjecture is not known in general:

Conjecture 1.4.3. *Every square-free positive integer n which is congruent to $5, 6$ or 7 modulo 8 is a congruent number.*

1.5. **Fermat's margin note, and the birth of Algebraic Number Theory.** The above discussion should warn the reader that innocent-looking questions on Diophantine equations may actually turn out to be very hard. At a first glance one may think that there is nothing special about question 1.1.2 making it harder that 1.1.1; however the latter question was answered by the Greeks, whereas the former survived more than one millennium of research.

Similarly, when reading question 1.1.1 Fermat wrote on the margin of his copy of Diophantus' Arithmetics that "one cannot split a cube into two cubes, a fourth power into two fourth powers, nor any power of exponent larger than two into two powers with the same exponent". In other words, if $n > 3$ then the equation

$$X^n + Y^n = Z^n \tag{1.5.0.1}$$

has no solution $(a, b, c) \in \mathbb{Z}$ with $abc \neq 0$. Fermat claimed, but did not write down, a proof of the statement. A proof was eventually completed by Wiles in 1995 - more than 300 years after Fermat's claim - building on the work of several mathematicians.

The first attempts to prove Fermat's statement followed a strategy roughly similar to the one we used to find Pythagorean triples. Fermat himself dealt with the case $n = 4$, hence it suffices to consider $n = p$ an odd prime. Letting $\zeta_p = e^{\frac{2\pi i}{p}} \in \mathbb{C}$, the equation (1.5.0.1) can be written as

$$(X + Y)(X + \zeta_p Y)(X + \zeta_p^2 Y) \cdots (X + \zeta_p^{p-1} Y) = Z^p.$$

Kummer had the idea to enlarge the ring $\mathbb{Z}$ and work in the ring

$$\mathbb{Z}[\zeta_p] = \{a_0 + a_1\zeta_p + \ldots + a_{p-2}\zeta_p^{p-2}, a_0, \ldots, a_{p-2} \in \mathbb{Z}\},$$

and use the "separating powers trick" to study the solutions to (1.5.0.1). However, he soon realised that one is *not always allowed* to use the "separating powers trick". The issue is that *unique factorisation in* $\mathbb{Z}[\zeta_p]$ fails in general, the first counterexample occurring for $p = 23$. It was from the attempt to understand this new phenomenon, and possibly circumvent the problem, that Algebraic Number Theory was born - see Kummer's letter [17]. In this course we will try to understand Kummer's brilliant idea to overcome the failure of unique factorisation, and apply it to solve several Diophantine equations.

1.6. ***The MRDP Theorem, and a general warning.** If one is ambitious (or prefers general questions to specific examples) one may wonder about the following question:

Question 1.6.1. Is there a general method to determine if a given Diophantine equation has an integral solution?

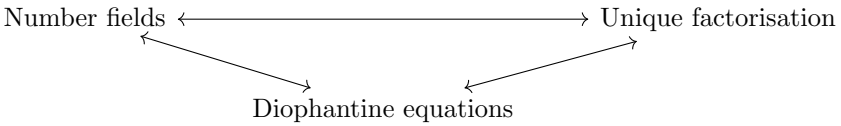
In fact, Hilbert did ask this question in 1900: it was the 10th of a list of 23 problems he presented at the International Congress of Mathematicians. In view of the previous discussion this may seem a hopeless task. Indeed it is so, in the following precise sense:

Theorem 1.6.2. *(Matiyasevich, Robinson, Davis, Putnam) There is* no algorithm *which can determine, given any Diophantine equation, whether it admits or not an integral solution.*

Remark 1.6.3. (for algorithmically inclined readers; see [22], [28]) The proof of this theorem was completed in 1970 by Y. Matiyasevich. Ultimately this result is intimately connected with Hilbert's *Entscheidungsproblem*, which was shown to be *not* resolvable in general by Church and Turing in 1936. This is in turn a consequence of the fact that Turing's *halting problem* cannot be solved.

One should not be too discouraged by the previous theorem. In a sense, it confirms the common experience that there seems to be no "systematic way" to solve Diophantine equations. Each time we are given one, we have to discover some *hidden structure* which allows us to find its solutions, or to show that there are none. A variety of different techniques has been developed so far to approach several kinds of Diophantine equations; nonetheless many questions remain open, and *producing solutions* to equations is still very hard.

In this course we will mainly study one possible technique to solve Diophantine equations: we will work with suitable rings *larger than* $\mathbb{Z}$ (e. g. $\mathbb{Z}[\zeta_p]$) and try to understand to what extent *unique factorisation* and the "separating powers trick" hold. Hence this course will be about contemplating the following picture



Exercise 1.6.4. **The divisor game:** Let n be a positive integer and let $\mathcal{D}_n = \{d > 0, d \mid n\}$ be the set of all positive divisors of n. We play the following game: *you* start, and you have to pick an element $d \in \mathcal{D}_n$. Once you have done so, *d as well as all its divisors* are taken away from $\mathcal{D}_n$. It is now my turn to choose an element among the remaining ones in $\mathcal{D}_n$. As before, this element and all its divisors disappear from $\mathcal{D}_n$. The game continues, and the first player who picks n itself loses the game.

(1) Say $n = p^k$ where p is prime. Can you find a winning strategy?
(2) If $n = pq$ where p, q are distinct primes, can you find a winning strategy?
(3) Do you always have a winning strategy, provided that you start? (N. B. you do not have to exhibit one...)

Irrationality of π: There are several reformulations of the key fact 1.2.1; they all are very basic but have profound consequences. For example, 1.2.1 tells us that there cannot be *arbitrarily small positive integers*. In this exercise we will use this to prove - following [27] - that π is *irrational*.

(1) Assume by contradiction that $\pi = \frac{a}{b}$ with $a \in \mathbb{Z}, b \in \mathbb{Z} \smallsetminus \{0\}$. Set

$$P_n(X) = \frac{X^n(a - bX)^n}{n!} \in \mathbb{Q}[X]$$

for $n \geq 1$. Check that $P_n(X) = P_n(\frac{a}{b} - X)$. Deduce that $P_n(X)$ and all its derivatives $P_n^{(i)}(X), i \geq 1$ take *integer values* at $X = 0$ and $X = \pi$.

(2) Let $Q_n(X) = P_n(X) - P_n^{(2)}(X) + P_n^{(4)}(X) - \cdots + (-1)^n P_n^{(2n)}(X)$. Show that

$$\frac{\mathrm{d}}{\mathrm{d}X}(Q_n'(X)\sin X - Q_n(X)\cos X) = P_n(X)\sin(X)$$

hence

$$\int_0^\pi P_n(X)\sin X \mathrm{d}X = Q_n(\pi) + Q_n(0);$$

in particular, deduce from (1) that the above integral is always an integer.

(3) Show that, for $0 < x < \pi$, the inequalities $0 < P_n(X)\sin X \leq \frac{\pi^n a^n}{n!}$ hold. Deduce that, when n goes to infinity, the integral $\int_0^\pi P_n(X)\sin X \mathrm{d}X$ gets arbitrarily small (without being 0).

(4) Conclude.

Variations of the above steps can be used both to prove that a given number α is irrational and to study its *approximations by rational numbers*. In the latter case, assuming that α has a "very good" rational approximation $\frac{a}{b}$, one often looks at a suitable polynomial $P \in \mathbb{Q}[X]$ such that $P(\frac{a}{b}) \in \mathbb{Q}$ is non-zero and has bounded denominator. Then one tries to estimate P and show that it can be chosen such that $P(\frac{a}{b})$ is arbitrarily small; at this point 1.2.1 will give a contradiction, showing that the good approximation $\frac{a}{b}$ could not exist. We will see examples of this technique later on.

2. Lecture 2: Integral points on affine spaces

In this lecture we study integral solutions of systems of linear equations with integral coefficients. Algebraically, we will be doing linear algebra over $\mathbb{Z}$; geometrically, we we will be looking for integral points on affine spaces.

2.1. **Integral points on affine hyperplanes.** Let $(a_1, \ldots, a_n) \in \mathbb{Z}^n \smallsetminus \{0\}$ and $b \in \mathbb{Z}$. We want to study integral solutions of the equation

$$a_1X_1 + \ldots + a_nX_n = b. \tag{2.1.0.1}$$

As a warm up, let us make the following observations:

(1) if $n \geq 2$ and equation (2.1.0.1) has an integral solution $(x_1, \ldots, x_n)$ then it has infinitely many integral solutions. Indeed any other integral solution is of the form $(x_1, \ldots, x_n) + (z_1, \ldots, z_n)$ with
$$a_1z_1 + \ldots + a_nz_n = 0.$$
This equation is homogeneous and it has a non-zero rational solution - better, the set of rational solutions is a $\mathbb{Q}$-vector space of dimension $n-1$ - hence it has infinitely many integral solutions.

(2) If $n = 2$ then (2.1.0.1) has an integral solution if and only if b belongs to the set $\{a_1x+a_2y \mid x, y \in \mathbb{Z}\}$. By the Bézout property, this is equivalent to the fact that $\gcd(a_1, a_2) \mid b$.

Let us restate the case $n = 2$ slightly differently: let $\mathbf{A} = (a_1, a_2)$; we have to determine whether there is $\mathbf{x} \in \mathbb{Z}^2$ such that

$$\mathbf{A}\mathbf{x}^t = b.$$

To do this we may replace $\mathbf{A}$ by $\mathbf{B} = \mathbf{A}\mathbf{T}$ for some matrix $\mathbf{T} \in GL_2(\mathbb{Z})$: indeed, recall that $GL_2(\mathbb{Z})$ denotes the set of matrices with integral coefficients and determinant ± 1.[1] In particular every $\mathbf{T} \in GL_2(\mathbb{Z})$ has an inverse $\mathbf{T}^{-1} \in GL_2(\mathbb{Z})$ such that $\mathbf{T}\mathbf{T}^{-1} = Id$. Hence, if there is $\mathbf{y} \in \mathbb{Z}^2$ such that $\mathbf{B}\mathbf{y}^t = b$ then, setting $\mathbf{x} = \mathbf{y}\mathbf{T}^t$ we find

$$b = \mathbf{B}\mathbf{y}^t = \mathbf{A}\mathbf{T}\mathbf{y}^t = \mathbf{A}\mathbf{T}\mathbf{T}^{-1}\mathbf{x}^t = \mathbf{A}\mathbf{x}^t;$$

conversely, a solution of the equation $\mathbf{A}\mathbf{x}^t = b$ gives rise to a solution of $\mathbf{B}\mathbf{y}^t = b$.

We now want to find $\mathbf{T}$ such that $\mathbf{B} = \mathbf{A}\mathbf{T} = (d, 0)$ for some $d > 0$. If we can do this then the equation $\mathbf{B}\mathbf{y}^t = b$ will manifestly have an integral solution if and only if $d \mid b$.

The idea is to apply repeatedly Euclid's algorithm in order to make the entries in the given matrix $\mathbf{A}$ smaller. Let us introduce the following notation: for $\lambda \in \mathbb{Z}$,

$$\mathbf{E}^{(\lambda)} = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \quad \mathbf{E}_{(\lambda)} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}.$$

After possibly switching a_1 and a_2, which can be achieved via right multiplication by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, we may assume $a_1 \neq 0$. If $a_2 = ka_1$ then $\mathbf{A}\mathbf{E}^{(-k)} = (a_1, 0)$; after possibly changing sign to a_0 - multiplying by $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ - we obtain a vector in the desired form. If $a_1 \nmid a_2$ then we can write $a_2 = qa_1 + r$ with $0 < r < |a_1|$; hence $\mathbf{A}\mathbf{E}^{(-q)} = (a_1, r)$. Switching a_1 and r we obtain the vector $(b_1, b_2) = (r, a_1)$, with $b_1 < |a_1|$, and we repeat the process. By the key fact 1.2.1 this algorithm terminates after a finite number of steps. In fact, we see that in this case we are exactly running Euclid's algorithm written in matrix form: the outcome will be the vector $(\gcd(a_1, a_2), 0)$, and we recover the criterion for the solubility of $\mathbf{A}\mathbf{x}^t = b$ recalled in (2) above.

The advantage of the approach we just discussed is that it can be generalised to arbitrary systems of linear equations with integral coefficients. The next exercise deals with the case of one equation in several variables (2.1.0.1).

Notation 2.1.1. For an integer $n \geq 1$, we denote by $GL_n(\mathbb{Z})$ the group of $n \times n$ matrices with integral coefficients and determinant ± 1. Equivalently, $\mathbf{A}$ belongs to $GL_n(\mathbb{Z})$ if it has an inverse with *integral* coefficients. We denote by $SL_n(\mathbb{Z}) \subset GL_n(\mathbb{Z})$ the subgroup of matrices with determinant one.

[1] Beware: this is much stronger than requiring $\det(\mathbf{A}) \neq 0$.

Exercise 2.1.2. Let $\mathbf{A} = (a_1, \dots a_n) \in \mathbb{Z}^n \smallsetminus \{0\}$ and $b \in \mathbb{Z}$. Prove that there exists $\mathbf{T} \in GL_n(\mathbb{Z})$ such that $\mathbf{AT} = (d, 0, \dots, 0)$ for a unique $d > 0$, the greatest common divisor of $a_1, \dots, a_n$. Deduce that the equation

$$a_1 X_1 + \dots + a_n X_n = b$$

has an integral solution if and only if every integer dividing all the numbers $a_1, \dots, a_n$ also divides b.

2.2. **General linear systems.** Let us now consider arbitrary linear systems of the form

$$\mathbf{A}\mathbf{x}^t = \mathbf{b}^t \tag{2.2.0.1}$$

where $\mathbf{A}$ is a non-zero $m \times n$ matrix with coefficients in $\mathbb{Z}$ and $\mathbf{b} \in \mathbb{Z}^m$. Given two matrices $\mathbf{S} \in GL_m(\mathbb{Z}), \mathbf{T} \in GL_n(\mathbb{Z})$, the above system has an integral solution if and only if the same is true for the system

$$\mathbf{SATx}^t = \mathbf{Sb}^t.$$

The procedure described in the previous section can be generalised: multiplying $\mathbf{A}$ both on the left and on the right by suitable invertible matrices with integral coefficients we can transform it in a particularly simple diagonal form, called the *Smith normal form*, allowing to check if (2.2.0.1) has integral solutions.

Theorem 2.2.1 (Smith normal form of a matrix)**.** *Let $\mathbf{A}$ be a non zero $m \times n$ matrix with coefficients in $\mathbb{Z}$. Then there exist matrices $\mathbf{S} \in GL_m(\mathbb{Z}), \mathbf{T} \in GL_n(\mathbb{Z})$, an integer $r > 0$ and positive integers $d_1 \mid d_2 \mid \dots \mid d_r$ such that $\mathbf{SAT}$ is of the form*

$$\mathbf{SAT} = \begin{pmatrix} \operatorname{diag}(d_1, \dots, d_r) & \dots & \dots & \vdots \\ \vdots & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \dots & \dots & 0 \end{pmatrix}$$

where $\operatorname{diag}(d_1, \dots, d_r)$ *is the diagonal matrix with entries $d_1, \dots, d_r$ and all the other entries are zero.*

Remark 2.2.2. In fact, the integers $d_1, \dots, d_r$ are unique; furthermore the theorem holds true for matrices over an arbitrary PID (=principal ideal domain; we will recall the definition later); in this case $d_1, \dots, d_r$ are unique up to multiplication by units.

Exercise 2.2.3. Let $\mathbf{A}$ be a non-zero matrix with coefficients in $\mathbb{Z}$. Prove the following statements:

(1) The upper left entry of the Smith normal form of $\mathbf{A}$ equals the greatest common divisor of the coefficients of $\mathbf{A}$.
(2) The determinant of $\mathbf{A}$ equals that of its Smith normal form up to sign.

Deduce that the Smith normal form of $\mathbf{A}$ is unique if $\mathbf{A}$ is a 2×2 matrix.

2.2.4. *Smith normal form in the 2×2 case.* We will explain how to obtain the Smith normal form in the simplest case of a 2×2 matrix, which is the one we will mostly need in this course. The general algorithm is similar to the one we will outline but more involved; we refer the reader to [8, pp. 195-197].

Notice that right multiplication by $\mathbf{E}^{(\lambda)}$ (resp. $\mathbf{E}_{(\lambda)}$) leaves the first column c_1 of a 2×2 matrix unchanged, and replaces the second one c_2 by $\lambda c_1 + c_2$ (resp. leaves the second column unchanged, and replaces the first one $\lambda c_2 + c_1$). Multiplication on the left by $\mathbf{E}_{(\lambda)}, \mathbf{E}^{(\lambda)}$ has a similar effect on the rows.

Let us start with a non-zero matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Up to exchanging rows and columns, we may assume that a is the non-zero element with smallest absolute value in $\mathbf{A}$. We will produce a matrix which

a: either is of the form $\begin{pmatrix} a_1 & 0 \\ 0 & d_1 \end{pmatrix}$ with $a_1, d_1 \geq 0$ and $a_1 \mid d_1$;

b: or contains a non-zero entry whose absolute value is strictly smaller than $|a|$.

We then repeat the process; this algorithm must terminate, yielding the Smith normal form of $\mathbf{A}$. Here are the steps:

(1) Apply the procedure in the previous section to the first row and obtain a matrix of the form
$$\mathbf{A}_1 = \begin{pmatrix} a_1 & 0 \\ c_1 & d_1 \end{pmatrix}$$
with $a_1 = \gcd(a, b)$. If $a_1 < |a|$ then we are in case **b**.
(2) If $a_1 = |a|$ and $a_1 \nmid c_1$ then applying the procedure of the previous section to the first column we will be again in case **b**.
(3) If $a_1 = |a|$ and $c_1 = ka_1$ then $\mathbf{E}_{(-k)}\mathbf{A}_1 = \begin{pmatrix} a_1 & 0 \\ 0 & d_1 \end{pmatrix}$.
(4) If $a_1 \mid d_1$ we are in case **a** (after possibly changing sign to d_1).
(5) If not right multiplication by $\mathbf{E}_{(1)}$ yields the matrix $\begin{pmatrix} a_1 & 0 \\ d_1 & d_1 \end{pmatrix}$ and applying the procedure of the previous section to the first column brings us to case **b**.

2.2.5. **Consequences.*

(1) For every integer $n \geq 1$ the group $SL_n(\mathbb{Z})$ of $n \times n$ matrices with integer coefficients and determinant one is generated by *elementary matrices*, which are generalisations of the matrices $\mathbf{E}^{(\lambda)}, \mathbf{E}_{(\lambda)}$ introduced above. The algorithm we described reduces a given matrix in $GL_2(\mathbb{Z})$ to the identity matrix, but we made use of matrices with negative detarminant in our argument, so one needs to refine it and show that these matrices are not needed. Beware that it is still true that $SL_n(A)$ is generated by elementary matrices if A is a Euclidean domain, but the statement *fails* for general principal ideal domains.
(2) For every integer $n \geq 1$ the group $SL_n(\mathbb{Z})$ is *finitely generated*. This follows from the previous point plus the identities $\mathbf{E}^{(\lambda)} = (\mathbf{E}^{(1)})^\lambda, \mathbf{E}_{(\lambda)} = (\mathbf{E}_{(1)})^\lambda$ - which hold true for general elementary matrices. For example, the group $SL_2(\mathbb{Z})$ is generated by the two matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Finally, $SL_n(\mathbb{Z})$ is *finitely presented*, i.e. there are finitely many relations among its finitely many generators [33, Chapter II].

2.3. **Seigel's lemma.** Let us now consider a linear system of the form
$$\mathbf{A}\mathbf{x}^t = 0$$
where $\mathbf{A}$ is an $m \times n$ matrix with integral coefficients, and $n > m$. This system has the trivial solution $\mathbf{x} = (0, \dots, 0)$; however, as $n > m$ there are also infinitely many non-trivial rational, and even integral, solutions. One may ask if one can bound above the minimal "size" of a non-zero integral solution in terms of the "size" of the matrix $\mathbf{A}$. This is indeed possible, and is the content of the following lemma due to Siegel. Before stating it, let us fix a precise notion of *size*.

Notation 2.3.1. For $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n$ we set $\|\mathbf{x}\| = \max_{1 \leq i \leq n} |x_i|$. If $\mathbf{A} = (a_{ij})$ is an $m \times n$ matrix with integral coefficients we denote $\|\mathbf{A}\| = \max |a_{ij}|$.

Lemma 2.3.2 (Siegel)**.** *Let $n > m$ be two integers and $\mathbf{A}$ a non-zero $m \times n$ matrix with integral coefficients. There exists $\mathbf{x} \in \mathbb{Z}^n$ such that $\mathbf{A}\mathbf{x}^t = 0$ and*
$$0 < \|\mathbf{x}\| \leq (n\|\mathbf{A}\|)^{\frac{m}{n-m}}.$$

Proof. In this proof we will use the following notation: for a real number a, we denote by $\lfloor a \rfloor$ the *floor* of a, i. e. the largest integer not exceeding a. We will also let $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$, so that $|a| = a^+ + a^-$.

The matrix $\mathbf{A}$ gives rise to a linear map (denoted with the same symbol) $\mathbb{Z}^n \to \mathbb{Z}^m$. We want to produce distinct elements $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{Z}^n$ such that $\mathbf{A}\mathbf{x}_1^t = \mathbf{A}\mathbf{x}_2^t$ and the coordinates of $\mathbf{x}_1, \mathbf{x}_2$ belong to a small enough interval $[0, T]$. More precisely, if we can show the existence of such $\mathbf{x}_1$ and $\mathbf{x}_2$ for some $T \leq (n\|\mathbf{A}\|)^{\frac{m}{n-m}}$ then $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ will satisfy the requirements of the lemma.

For $1 \leq i \leq m$ let $L_i = \sum_{j=1}^{n} |a_{ij}|$. We may assume that $L_i \neq 0$: otherwise every coefficient in the i-th row of $\mathbf{A}$ equals zero and we can forget about this row. Set $T = \left\lfloor (L_1 \cdots L_m)^{\frac{1}{n-m}} \right\rfloor$. Notice that

$$L_i \leq n\|\mathbf{A}\| \Rightarrow T \leq (n\|\mathbf{A}\|)^{\frac{m}{n-m}} .$$

There are $(T+1)^n$ integers in $[0,T]^n$, hence, by the pigeonhole principle, in order to complete the proof it suffices to show that $\mathbf{A}([0,T]^n)$ contains strictly less than $(T+1)^n$ integers. For every $\mathbf{y} \in [0,T]^n$, writing $\mathbf{A}\mathbf{y}^t = (z_1, \ldots, z_m)^t$ we have

$$-T\sum_{j=1}^{n} a_{ij}^- \leq z_i \leq T\sum_{j=1}^{n} a_{ij}^+, \quad 1 \leq i \leq m.$$

It follows that the image of $[0,T]^n$ via $\mathbf{A}$ contains at most

$$\prod_{i=1}^{m}(T\sum_{j=1}^{n} a_{ij}^- + T\sum_{j=1}^{n} a_{ij}^+ + 1) = \prod_{i=1}^{m}(1 + T\sum_{j=1}^{n} |a_{ij}|) = \prod_{i=1}^{m}(TL_i + 1)$$

integers. Finally, we have

$$\prod_{i=1}^{m}(TL_i + 1) \leq \prod_{i=1}^{m}(TL_i + L_i) = (L_1 \cdots L_m)(T+1)^m < (T+1)^{n-m}(T+1)^m = (T+1)^n$$

where the last inequality holds true by definition of T. □

Remark 2.3.3. The proof of Siegel's lemma is elementary, only relying on the pigeonhole principle. This lemma plays however a key role in several deep diophantine results, particularly those having to do with diophantine approximation and transcendence. We will see the lemma in action towards the end of this course.

3. Lecture 3: Integral points on circles and the ring $\mathbb{Z}[i]$

In this lecture we study the arithmetic properties of the ring of Gaussian integers $\mathbb{Z}[i]$, and use them to find out which numbers are sum of two squares of integers. Geometrically, we want to understand which circles with integer radius and centre at the origin have integral points.

3.1. **Sums of two squares.** Let us start with the following variation of question 1.1.1:

Question 3.1.1. Split a given integer in two squares. In other words, for which (positive) integers n does the equation

$$X^2 + Y^2 = n$$

have integral solutions?

Diophantus first observed that, if $n = a^2 + b^2$ and $m = c^2 + d^2$, then

$$nm = (a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2 \tag{3.1.1.1}$$

so the product of two sums of two squares is a sum of two squares; let us first of all try to understand which prime numbers p can be written in the form

$$p = a^2 + b^2,\ a, b \in \mathbb{Z}.$$

Here are some experiments (the reader is encouraged to try more)

2	$(\pm 1)^2 + (\pm 1)^2$
3	-
5	$(\pm 1)^2 + (\pm 2)^2$
7	-
11	-
13	$(\pm 2)^2 + (\pm 3)^2$
17	$(\pm 1)^2 + (\pm 4)^2$
...	...
1093	$(\pm 2)^2 + (\pm 33)^2$
...	...

It seems that

(1) $p \equiv 3 \pmod 4 \Rightarrow p$ is *not* of the form $a^2 + b^2$;
(2) $p \equiv 1 \pmod 4 \Rightarrow$ there are $a, b \in \mathbb{Z}$ such that $p = a^2 + b^2$. Furthermore, in this case all the integral solutions of the equation $p = X^2 + Y^2$ are $(\pm a, \pm b), (\pm b, \pm a)$.

What is going on? Where does (3.1.1.1) come from?

Lemma 3.1.2. *Let $p \equiv 3 \pmod 4$ be a prime. Then there are no integers $a, b \in \mathbb{Z}$ such that $a^2 + b^2 = p$.*

Proof. Suppose by contradiction that there exist $a, b \in \mathbb{Z}$ such that $p = a^2 + b^2$. Reducing modulo 4 we obtain

$$3 \equiv a^2 + b^2 \pmod 4$$

On the other hand we have $\{[0]^2, [1]^2, [2]^2, [3]^2\} = \{[0], [1]\} \subset \mathbb{Z}/4\mathbb{Z}$. It follows that the sum of two squares is never congruent to 3 modulo 4. □

Remark 3.1.3. The technique employed in the above proof is perhaps the simplest way to study Diophantine equations: if such an equation has a solution in $\mathbb{Z}$, then it has a solution in $\mathbb{Z}/n\mathbb{Z}$ for every $n \geq 1$. The latter fact is much easier to check, hence this idea can often be used to show that certain equations have *no integral solutions*. Conversely, one may wonder if a Diophantine equation having a solution in $\mathbb{Z}/n\mathbb{Z}$ for every $n \geq 1$ also has an integral solution. This is false in general, but (a slightly improved version of) this property holds for certain classes of equations, which are said to satisfy the *Hasse-Minkowski local-global principle*.

We still have to understand whether every prime $p \equiv 1 \pmod 4$ is a sum of two squares. This turns out to be true; it was proven by Fermat in a letter to Mersenne dated December 25th, 1640. Zagier found a one-sentence proof in 1990 [39]; an analytic approach to the question is explained in [34]. In the next section we will give a scientific proof of this fact, based on the arithmetic properties of the ring $\mathbb{Z}[i]$. Before that, following Remark 3.1.3, let us consider the analogous question over $\mathbb{Z}/l\mathbb{Z}$ for every prime l.

3.1.4. *The field $\mathbb{F}_l$.* Let l be a prime. The Bézout property of $\mathbb{Z}$ implies that the ring $\mathbb{Z}/l\mathbb{Z}$ is a *field*: for every element $a \in \mathbb{Z}/l\mathbb{Z} \smallsetminus \{0\}$ there exists $b \in \mathbb{Z}/l\mathbb{Z} \smallsetminus \{0\}$ such that $ab = 1 \in \mathbb{Z}/l\mathbb{Z}$. The field $\mathbb{Z}/l\mathbb{Z}$ is called the field with l elements, and it is denoted by $\mathbb{F}_l$. We will also use the notation $\mathbb{F}_l^\times = (\mathbb{Z}/l\mathbb{Z})^\times = \mathbb{Z}/l\mathbb{Z} \smallsetminus \{0\}$.

Beware: if n is not a prime number, you should *not* use the notation $\mathbb{F}_n$ to denote the ring $\mathbb{Z}/n\mathbb{Z}$. For example, the symbol $\mathbb{F}_4$ is commonly used to denote the *field with four elements*: such a field exists, and you can try to construct it as an exercise[2], but it is not isomorphic to the ring $\mathbb{Z}/4\mathbb{Z}$.

3.1.5. *Squares in $\mathbb{Z}/l\mathbb{Z}$.* Let l be an *odd* prime. The squares in $\mathbb{F}_l^\times$ are

$$(\mathbb{F}_l^\times)^2 = \{[1]^2, [2]^2, \cdots, [\frac{l-1}{2}]^2\}.$$

In particular, there are $\frac{l-1}{2}$ squares in $\mathbb{F}_l^\times$, hence $\frac{l+1}{2}$ squares in $\mathbb{F}_l$. The multiplicative group $\mathbb{F}_l^\times$ has cardinality $l-1$, hence we have, for every $a \in \mathbb{Z}$ such that $\gcd(a,l) = 1$:

$$a^{l-1} \equiv 1 \pmod{l}.$$

This is known as Fermat little theorem (a special case of Lagrange theorem).

Exercise 3.1.6. Comparing the products $P = \prod_{b \in \mathbb{F}_l^\times} b$ and $\prod_{b \in \mathbb{F}_l^\times} (ab)$, prove Fermat's little theorem. What is the value of P?

It follows from Fermat little theorem that, if $\gcd(a,l) = 1$ and a is a square modulo l, i.e. $a \equiv b^2 \pmod{l}$, then $a^{\frac{l-1}{2}} \equiv b^{l-1} \equiv 1 \pmod{l}$. Hence every square in $\mathbb{F}_l^\times$ is a root of the polynomial $X^{\frac{l-1}{2}} - 1 \in \mathbb{F}_l[X]$. As $\mathbb{F}_l$ is a field, this polynomial cannot have more than $\frac{l-1}{2}$ roots. It follows that

$$a \text{ is a square modulo } l \Leftrightarrow a^{\frac{l-1}{2}} \equiv 1 \pmod{l}.$$

In other words, denoting by

$$\left(\frac{a}{l}\right) = \begin{cases} 0 & \text{if } l \mid a \\ 1 & \text{if } [a] \in (\mathbb{F}_l^\times)^2 \\ -1 & \text{if } [a] \notin (\mathbb{F}_l^\times)^2 \end{cases}$$

the *Legendre symbol*, we have, for every $a \in \mathbb{Z}$:

$$\left(\frac{a}{l}\right) \equiv a^{\frac{l-1}{2}} \pmod{l} \text{ (Euler's criterion)}.$$

Lemma 3.1.7. *Let $p \equiv 1 \pmod 4$ be a prime and l a prime. Then the equation*

$$X^2 + Y^2 = p \tag{3.1.7.1}$$

has a solution $(a,b) \neq (0,0)$ in $\mathbb{F}_l$.

Proof. If $l = 2$ we can take $(a,b) = ([1],[0])$. Let l be an odd prime different from p; we know that there are $\frac{l+1}{2}$ squares in $\mathbb{F}_l$. Hence the sets

$$Q = \{x^2, x \in \mathbb{F}_l\}$$
$$Q' = \{[p] - y^2, y \in \mathbb{F}_l\}$$

have non empty intersection. If $c \in Q \cap Q'$ we have $c = a^2 = [p] - b^2 \Rightarrow [p] = a^2 + b^2$ for some $a, b \in \mathbb{F}_l$. As $[p] \neq 0 \in \mathbb{F}_l$ one of a, b must be non zero.

Finally, let us consider the case $l = p$. As $p \equiv 1 \pmod 4$ we have $\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. Hence there is $a \in \mathbb{F}_p$ such that $a^2 = -1 \pmod{p}$, and $(a,1) \neq (0,0)$ is a solution of (3.1.7.1). □

Remark 3.1.8. The reader may wonder why we discarded the obvious solution $(0,0)$ of (3.1.7.1) modulo p. The reason is that we are ultimately interested in finding a solution in $\mathbb{Z}$, and the mod p solution $(0,0)$ does *not* lift to a solution in $\mathbb{Z}$. Indeed, it does not even lift to a solution in $\mathbb{Z}/p^2\mathbb{Z}$: if $a, b \in \mathbb{Z}/p^2\mathbb{Z}$ are of the form $a = pa', b = pb'$, then $a^2 + b^2 = 0 \neq p \in \mathbb{Z}/p^2\mathbb{Z}$.

[2]You can also learn several interesting facts and conjectures about *finite fields* in [1].

3.2. **Properties of $\mathbb{Z}[i]$.** We wish to find the integral solutions of (3.1.7.1) when $p \equiv 1 \pmod 4$. The starting point is the observation that the equation $p = X^2 + Y^2$ can be written as

$$p = (X + iY)(X - iY).$$

Let us consider the ring of *Gaussian integers*

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

For $\alpha = a + ib \in \mathbb{Z}[i]$ we will denote by $\bar{\alpha} = a - ib$ its complex conjugate. Then the *norm map*

$$N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$$
$$\alpha = a + ib \mapsto |\alpha|^2 = \alpha\bar{\alpha} = a^2 + b^2$$

satisfies

(1) $N(\alpha) = 0 \Leftrightarrow \alpha = 0$.
(2) $N(\alpha\beta) = N(\alpha)N(\beta)$ for every $\alpha, \beta \in \mathbb{Z}[i]$.
(3) $N(\alpha) = 1 \Leftrightarrow \alpha \in \mathbb{Z}[i]^\times = \{\alpha \in \mathbb{Z}[i] \mid \exists \beta \in \mathbb{Z}[i], \alpha\beta = 1\}$ ($\Leftarrow$ follows from (2); conversely, if $N(\alpha) = 1$ then $\bar{\alpha}$ is the multiplicative inverse of α).

It follows from (3) that the *group of units* of $\mathbb{Z}[i]$ is

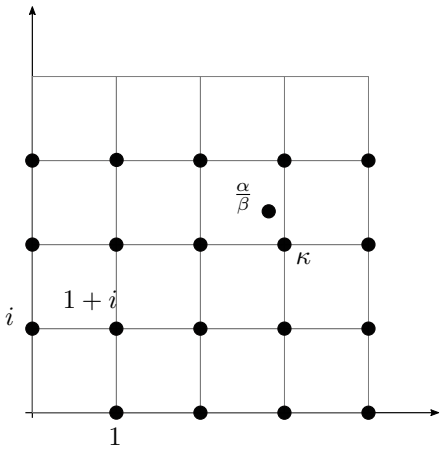
$$\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}.$$

Furthermore, our original question 3.1.1 can be reformulated as: which positive integers n are in the image of the norm map N? Property (2) above explains equation (3.1.1.1).

Crucially for us, the following analogue of Euclid algorithm works in the ring $\mathbb{Z}[i]$.

Lemma 3.2.1. *Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$. Then there exist $\kappa, \lambda \in \mathbb{Z}[i]$ such that $N(\lambda) < N(\beta)$ and $\alpha = \kappa\beta + \lambda$.*

Proof. The set $\{a + ib, a, b \in \mathbb{Z}\} \subset \mathbb{C}$ is a square lattice in the plane, with sides of the squares of length 1. Given $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$ let $\kappa \in \mathbb{Z}[i]$ be an element with minimal distance from the ratio $\frac{\alpha}{\beta} \in \mathbb{C}$.



Then we have

$$\left|\frac{\alpha}{\beta} - \kappa\right|^2 \leq \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 < 1$$
$$\Rightarrow |\alpha - \beta\kappa| < |\beta|.$$

Hence κ and $\lambda = \alpha - \beta\kappa$ satisfy the conclusion of the lemma. □

As a consequence of the lemma, the ring $\mathbb{Z}[i]$ enjoys arithmetic properties which are analogous to the familiar properties of $\mathbb{Z}$. The proofs we gave in lecture 1 go through in this setting, replacing "minimum element" with "element with smallest norm" everywhere. Furthermore, *prime numbers* should be replaced by *irreducible elements*: an element $\pi \in \mathbb{Z}[i]$ is called *irreducible* if it is not a unit and whenever $\pi = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$ we have either $\alpha \in \mathbb{Z}[i]^\times$ or $\beta \in \mathbb{Z}[i]^\times$.

Bézout property: Let $\alpha, \beta \in \mathbb{Z}[i]$ which are not both equal to 0. There exists $\delta \in \mathbb{Z}[i]$ such that

$$\{\alpha x + \beta y \mid x, y \in \mathbb{Z}[i]\} = \delta\mathbb{Z}[i] = \{\delta x, x \in \mathbb{Z}[i]\}.$$

Such a δ divides both α and β. Furthermore, if $\gamma \in \mathbb{Z}[i]$ divides α and β then γ divides δ. The element δ is unique *up to multiplication by an element in* $\mathbb{Z}[i]^\times$. We call δ the *greatest common divisor* of α and β and write

$$\delta = \gcd(\alpha, \beta).$$

Euclid's Lemma: Let $\pi \in \mathbb{Z}[i]$ be irreducible and $\alpha, \beta \in \mathbb{Z}$. Then

$$\pi \mid \alpha\beta \Rightarrow \pi \mid \alpha \text{ or } \pi \mid \beta.$$

Existence and uniqueness of factorisation: Let $\alpha \in \mathbb{Z}[i] \smallsetminus \{0\}$. There exist $\pi_1, \dots, \pi_r$ irreducible and $u \in \mathbb{Z}[i]^\times$ such that

$$\alpha = u\pi_1\pi_2 \cdots \pi_r.$$

Furthermore, if $v \in \mathbb{Z}[i]^\times$ and $\kappa_1, \dots, \kappa_s$ are irreducible elements such that

$$u\pi_1 \cdot \pi_2 \cdot \cdots \cdot \pi_r = \alpha = v\kappa_1 \cdot \kappa_2 \cdot \cdots \cdot \kappa_s$$

then $r = s$ and, up to reordering the κ_i's, there exist units $u_1, \dots, u_r$ such that $\kappa_i = u_i\pi_i$ for $i = 1, \dots, r$.

Separating powers: Let $\alpha, \beta, \gamma \in \mathbb{Z}[i] \smallsetminus \{0\}$ and $n > 0$ such that

$$\alpha^n = \beta\gamma.$$

If $\gcd(\beta, \gamma)$ is a unit then there exist $\beta_1, \gamma_1 \in \mathbb{Z}[i]$ and $u, v \in \mathbb{Z}[i]^\times$ such that

$$\beta = u\beta_1^n \text{ and } \gamma = v\gamma_1^n.$$

3.2.2. *Irreducible elements in* $\mathbb{Z}[i]$. Let $\pi \in \mathbb{Z}[i]$ be an irreducible element. As $\pi \mid N(\pi)$, by Euclid lemma there is a (unique) prime $p \in \mathbb{Z}$ such that $\pi \mid p$. Therefore in order to understand irreducible elements in $\mathbb{Z}[i]$ we have to study how prime numbers decompose into irreducible elements in $\mathbb{Z}[i]$. Observe that, by properties (2), (3) of the norm function, if the norm of an element $\alpha \in \mathbb{Z}[i]$ is a prime number then α is irreducible.

(1) We have $2 = (1+i)(1-i) = (-i)(1+i)^2$, and $(1+i)$ is irreducible, as $N(1+i) = 2$ is prime.
(2) If p is prime and $p \equiv 3 \pmod 4$ then p is irreducible in $\mathbb{Z}[i]$. If this were not the case, we could write $p = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$ such that $N(\alpha)$ and $N(\beta)$ are not equal to one. As $N(\alpha)N(\beta) = N(p) = p^2$ we deduce that $N(\alpha) = N(\beta) = p$. Writing $\alpha = a + ib$ this yields $a^2 + b^2 = p$. This is impossible by Lemma 3.1.2.
(3) If $p \equiv 1 \pmod 4$ is prime then p is *not* irreducible in $\mathbb{Z}[i]$. Indeed, in this case we know that there exists $a \in \mathbb{Z}$ such that $a^2 \equiv -1 \pmod p$, hence $p \mid (a+i)(a-i)$. However $\frac{a+i}{p}$ and $\frac{a-i}{p}$ do not belong to $\mathbb{Z}[i]$, hence $p \nmid a \pm i$. Euclid's lemma implies that p is not irreducible in $\mathbb{Z}[i]$. Therefore we can write $p = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$ such that $N(\alpha) = N(\beta) = p$ (hence α and β are irreducible). Writing $\alpha = a + bi$ we have $a^2 + b^2 = p$ hence $\alpha\bar{\alpha} = p$ and $\beta = \bar{\alpha}$; finally, one checks that the elements $\alpha, \bar{\alpha}$ do not differ by multiplication by a unit.

We have proved

Theorem 3.2.3. *A set $\mathcal{P}_{\mathbb{Z}[i]}$ of representatives of irreducible elements of $\mathbb{Z}[i]$ up to multiplication by $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$ is given by*

$$\mathcal{P}_{\mathbb{Z}[i]} = \{(1+i)\} \cup \{p \mid p \equiv 3 \pmod 4\} \cup \{\pi_p, \bar{\pi}_p \mid p \equiv 1 \pmod 4\}$$

where, for $p \equiv 1 \pmod 4$, $\pi_p \in \mathbb{Z}[i]$ is an element of norm p.

Hence every element $\alpha \in \mathbb{Z}[i] \smallsetminus \{0\}$ can be uniquely written in the form

$$\alpha = i^k(1+i)^a \prod_{p \equiv 1 \pmod 4} \pi_p^{b_p}\bar{\pi}_p^{c_p} \prod_{p \equiv 3 \pmod 4} p^{d_p}, \quad k \in \mathbb{Z}/4\mathbb{Z},\ a, b_p, c_p, d_p \in \mathbb{Z}_{\geq 0}.$$

We are finally able to answer question 3.1.1.

Corollary 3.2.4. *An odd prime p can be written in the form $p = a^2 + b^2$ with $a, b \in \mathbb{Z}$ if and only if $p \equiv 1 \pmod 4$. In this case, the integral solutions of the equation $X^2 + Y^2 = p$ are $(\pm a, \pm b), (\pm b, \pm a)$.*

More generally, a positive integer n is a sum of two squares of integers if and only if for every prime $p \equiv 3 \pmod 4$, writing $n = p^{v_p(n)}m$ with $v_p(n) \geq 0$ and $\gcd(m, p) = 1$, the exponent $v_p(n)$ is even.

Exercise 3.2.5.
(1) Prove that there are infinitely many primes congruent to 3 modulo 4 (Hint: given $p_1, \dots, p_k$, consider $4p_1 \cdots p_k - 1$).
(2) Prove that there are infinitely many primes congruent to 1 modulo 4 (Hint: given $p_1, \dots, p_k$, consider $(2p_1 \cdots p_k)^2 + 1$ and use Euler's criterion).
(3) Prove that a prime p can be written in the form
$$p = a^2 + 16b^2, \ a, b \in \mathbb{Z}$$
if and only if $p \equiv 1 \pmod 8$.
(4) Use unique factorisation in $\mathbb{Z}[i]$ to determine primitive Pythagorean triples. Then determine all the prime numbers p such that there exists a right-angle triangle whose sides have integer length and whose hypotenuse has length p.

Exercise 3.2.6. In this exercise we will prove that the equation
$$Y^2 + 5 = X^3$$
has no integral solution. Suppose that $x, y \in \mathbb{Z}$ satisfy
$$y^2 + 4 = x^3 - 1;$$
(1) Show that $2 \mid y$.
(2) Deduce that $x \equiv 1 \pmod 4$, hence $x^2 + x + 1 \equiv 3 \pmod 4$.
(3) Deduce that there is a prime $p \equiv 3 \pmod 4$ dividing $y^2 + 4$, and obtain a contradiction.
(4) Use the same method to show that the equation $Y^2 + 101 = X^3$ has no integral solution (Hint: write it as $Y^2 + 4 \cdot 5^2 = X^3 - 1$).

3.3. ***Remarks on sums of cubes.** Having solved the problem of determining which integers are sums of two squares, we may wonder

Question 3.3.1. Which integers are sums of three cubes of integers?

Looking at the question in $\mathbb{Z}/n\mathbb{Z}$ we find a non-trivial obstruction for $n = 9$: since the only cubes in $\mathbb{Z}/9\mathbb{Z}$ are $\bar{0}$ and $\pm\bar{1}$ we see that no number congruent to 4 or 5 modulo 9 is a sum of three cubes. It is not known whether the converse is true. It was discovered in 2019 that
$$33 = 8866128975287528^3 - 8778405442862239^3 - 2736111468807040^3$$
$$42 = -80538738812075974^3 + 80435758145817515^3 + 12602123297335631^3.$$
(1) The number 0 has no representation as a sum of three non-zero cubes (=the Fermat equation for $n = 3$ has no non-trivial solution). This was proved by Euler.
(2) There are infinitely many ways of representing 1 (hence every integer cube) as a sum of three cubes of integers, in view of the following formula found by Mahler
$$1 = (9b^4)^3 + (3b - 9b^4)^3 + (1 - 9b^3)^3.$$
(3) At present (=January 2021) the first positive integer for which it is unknown whether it is a sum of three integer cubes is 114.
(4) Let us point out, in connection with the discussion in 1.6, that for the time being there is no known algorithm which, given a positive integer n as input, determines after a finite amount of time whether n is a sum of three integer cubes or not.
(5) Hardy and Littlewood conjectured that there are infinitely many prime numbers p which are sum of three integer cubes. This was proved by Heath-Brown in 2001, using *sieve methods* [13].
(6) On the other hand, it is known that every rational number q is the sum of three cubes of rational numbers: if $r = \frac{q}{9}$ then
$$q = (r-1)^3 + \frac{27(r^2+r)^3}{(r^2+r+1)^3} + \frac{(-r^3+3r+1)^3}{(r^2+r+1)^3}.$$

Exercise 3.3.2. Let $p > 2$ be a prime number.

(1) Show that $p = a^3 + b^3$ for some $a, b \in \mathbb{Z}$ if and only if $p = 3x^2 - 3x + 1$ for some $x \in \mathbb{Z}$.
(2) Show that in fact one can take $x \in \mathbb{Z}_{\geq 0}$ in the previous point.
(3) Deduce that, if $p = a^3 + b^3$ for some $a, b \in \mathbb{Z}$ then $p \equiv 1, 7 \pmod 9$.
(4) Notice that $2 \equiv 1 + 1 \pmod 9$, hence the result you just proved could not be established by merely looking at the equation $X^3 + Y^3 = p$ modulo 9. Which extra ingredient have you used?

4. Lecture 4: Mordell curves and quadratic rings

In this lecture we will study integral points on certain cubic curves called *Mordell curves*; this will lead us to investigate unique factorisation in suitable quadratic rings.

4.1. **The Mordell equation.** In this section we will study several examples of equations of the form

$$Y^2 = X^3 + k, \text{ for } k \in \mathbb{Z},$$

usually called *Mordell equations*. We are interested in finding the integral solutions of the above equation. Let us write it in the form $(Y - \sqrt{k})(Y + \sqrt{k}) = X^3$; we are led to work with the ring

$$\mathbb{Z}[\sqrt{k}] = \{a + b\sqrt{k}, a, b \in \mathbb{Z}\} \subset \mathbb{C}.$$

4.1.1. *The equation $Y^2 = X^3 - 1$.* Clearly the integers $x = 1, y = 0$ satifsfy the equation $y^2 = x^3 - 1$. Are there other integral solutions? Suppose that (x, y) is such a solution. First of all, x is odd. Indeed if it were even we would obtain $y^2 \equiv -1 \pmod 8$, which is not possible.

Write

$$x^3 = (y + i)(y - i);$$

we claim that $\delta = \gcd(y + i, y - i)$ is a unit in $\mathbb{Z}[i]^\times$. Indeed δ divides $2i$, hence $N(\delta) \mid 4$. On the other hand $\delta \mid x^3$ hence $N(\delta) \mid x^6$. As x is odd we find $N(\delta) = 1$, hence δ is a unit.

By the "separating powers trick" in $\mathbb{Z}[i]$, plus the fact that every element in $\mathbb{Z}[i]^\times$ is a cube in $\mathbb{Z}[i]$, there exist $a, b \in \mathbb{Z}$ such that

$$\begin{aligned} y + i &= (a + bi)^3 \\ \implies y &= a^3 - 3ab^2 = a(a^2 - 3b^2) \\ 1 &= 3a^2b - b^3 = b(3a^2 - b^2). \end{aligned}$$

The last equation implies that $b = \pm 1$. If $b = 1$ then we find $3a^2 = 2$ which has no integral solution. If $b = -1$ we obtain $a = 0$ hence $y = 0$.

Exercise 4.1.2. Using unique factorisation in $\mathbb{Z}$ show that the only integral solutions of the equation $Y^2 = X^3 + 16$ are $(x, y) = (0, \pm 4)$. Here are the main steps:

(1) Suppose that $(x, y) \in \mathbb{Z}^2$ satisfy $(y + 4)(y - 4) = x^3$ and $x \neq 0$. If y is odd show that $\gcd(y + 4, y - 4) = 1$ hence that $y \pm 4$ are cubes. Deduce a contradiction.
(2) Deduce from (1) that x, y are even; better, show that $4 \mid y$, hence that $4 \mid x$.
(3) Write $x = 4x_1, y = 4y_1$ with $x_1, y_1 \in \mathbb{Z}$ and check that $y_1^2 = 4x_1^3 + 1$. In particular y_1 is odd.
(4) Write $y_1 = 2m + 1$ and show that $m^2 + m = x_1^3$.
(5) Conclude.

4.2. **The equation $Y^2 = X^3 - 2$ and the ring $\mathbb{Z}[i\sqrt{2}]$.** Let $x, y \in \mathbb{Z}$ such that $y^2 = x^3 - 2$. Reducing modulo 8 we see that x must be odd. Write

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

Let us work in the ring $\mathbb{Z}[i\sqrt{2}] = \{a + i\sqrt{2}b, a, b \in \mathbb{Z}\}$. As in the case of $\mathbb{Z}[i]$ we can define a *norm function*

$$\begin{aligned} N : \mathbb{Z}[i\sqrt{2}] &\to \mathbb{Z}_{\geq 0} \\ \alpha &\mapsto \alpha\bar{\alpha}. \end{aligned}$$

Units in $\mathbb{Z}[i\sqrt{2}]$ are precisely elements of norm one, hence $\mathbb{Z}[i\sqrt{2}]^\times = \{\pm 1\}$. Let us *assume*, for the moment, that the Euclidean algorithm in $\mathbb{Z}[i\sqrt{2}]$ *works*, i.e. that an analogue of Lemma 3.2.1 holds true in our situation. Then the arithmetic properties of $\mathbb{Z}[i]$ explained in the previous lecture hold also for $\mathbb{Z}[i\sqrt{2}]$. In particular, provided that $\delta = \gcd(y + i\sqrt{2}, y - i\sqrt{2}) = 1$, we can apply the "separating powers trick". To show that $\delta = 1$ it suffices to notice that δ divides $2i\sqrt{2}$ and x^3,

hence $N(\delta) \mid \gcd(8, x^6) = 1$ (as x is odd). It follows that there exist $a, b \in \mathbb{Z}$ such that

$$\begin{aligned} y + i\sqrt{2} &= (a + i\sqrt{2}b)^3 \\ \implies y &= a^3 - 6ab^2 = a(a^2 - 6b^2) \\ 1 &= 3a^2b - 2b^3 = b(3a^2 - 2b^2). \end{aligned}$$

The last equation implies that $b = \pm 1$. If $b = -1$ then $3a^2 = 1$, which has no integral solution. Hence $b = 1$, so $a = \pm 1$ and $(x, y) = (3, \pm 5)$.

We still have to justify the fact that Euclid algorithm works in $\mathbb{Z}[i\sqrt{2}]$:

Lemma 4.2.1. *Let $\alpha, \beta \in \mathbb{Z}[i\sqrt{2}]$ with $\beta \neq 0$. Then there exist $\kappa, \lambda \in \mathbb{Z}[i\sqrt{2}]$ such that $N(\lambda) < N(\beta)$ and $\alpha = \kappa\beta + \lambda$.*

Proof. As in the proof of Lemma 3.2.1. This time the points in $\mathbb{Z}[i\sqrt{2}] \subset \mathbb{C}$ give a rectangular tiling of the plane, with sides of length 1, $\sqrt{2}$. As $(1/2)^2 + (\sqrt{2}/2)^2 < 1$ the argument we gave goes through. □

Remark 4.2.2. Notice that the previous argument *fails* for rings of the form $\mathbb{Z}[i\sqrt{k}]$ with $k \geq 3$. In fact, in the ring $\mathbb{Z}[i\sqrt{3}]$ unique factorisation does not hold! For example

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

The elements on the left and right hand side of the above equation are irreducible in $\mathbb{Z}[i\sqrt{3}]$ (as this ring does not contain any element of norm 2). However $2 \nmid (1 \pm i\sqrt{3})$, hence unique factorisation fails.

Similarly, the factorisations

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

show that unique factorisation does *not* hold in $\mathbb{Z}[i\sqrt{5}]$.

Exercise 4.2.3. (cf. Problem session)

(1) Find all the irreducible elements of $\mathbb{Z}[i\sqrt{2}]$.
(2) Using point (1) show that an odd prime p can be written in the form
$$p = a^2 + 2b^2, a, b \in \mathbb{Z}$$
if and only if $p \equiv 1, 3 \pmod 8$.
(3) Characterise the integers n of the form $n = a^2 + 2b^2$.

4.3. The equations $Y^2 = X^3 - 5$ and $Y^2 = X^3 - 26$.

4.3.1. *The equation $Y^2 = X^3 - 5$.* We have seen in Exercise 3.2.6 that there are no integers $x, y \in \mathbb{Z}$ such that $y^2 + 5 = x^3$. Let us try to prove this again using the methods in this lecture. If there was such a couple of integers (x, y), we would have

$$(y + i\sqrt{5})(y - i\sqrt{5}) = x^3.$$

Reducing modulo 8 we see that x must be odd. On the other hand if $\delta \in \mathbb{Z}[i\sqrt{5}]$ divides $y \pm i\sqrt{5}$ then $N(\delta) \mid \gcd(4 \cdot 5, x^6) \mid 5$. Hence $N(\delta) = 1$ or 5. However if $N(\delta) = 5$ then $5 \mid y^2 + 5$ hence $5 \mid x$ and $5 = x^3 - y^2$ is a multiple of 25, contradiction. Hence $N(\delta) = 1$.

Assumption 4.3.2. *Assume that the "separating powers trick"* works in $\mathbb{Z}[i\sqrt{5}]$.

Then we deduce that there exist $a, b \in \mathbb{Z}$ such that

$$\begin{aligned} y + i\sqrt{5} &= (a + i\sqrt{5}b)^3 \\ \implies y &= a^3 - 15ab^2 = a(a^2 - 15b^2) \\ 1 &= 3a^2b - 5b^3 = b(3a^2 - 5b^2) \end{aligned}$$

It follows that $b = \pm 1$. However both $3a^2 = 4$ and $3a^2 = 6$ have no integral solutions. Hence we recover (modulo assumption 4.3.2) the fact that the equation $Y^2 = X^3 - 5$ has no integral solutions.

4.3.3. *The equation $Y^2 = X^3 - 26$.* Let us now try to find the integers $x, y \in \mathbb{Z}$ such that $y^2 + 26 = x^3$. Reducing modulo 8 (as usual) we see that x is odd. Write

$$(y + i\sqrt{26})(y - i\sqrt{26}) = x^3;$$

if $\delta \in \mathbb{Z}[i\sqrt{26}]$ divides $y \pm i\sqrt{26}$, then $N(\delta) \mid \gcd(4 \cdot 26, x^6)$. As x is odd we find that $N(\delta) = 1$ or 13. However there is no element $a + i\sqrt{26}b \in \mathbb{Z}[i\sqrt{26}]$ with norm $a^2 + 26b^2 = 13$. It follows that $\delta \in \mathbb{Z}[i\sqrt{26}]^\times = \{\pm 1\}$.

Assumption 4.3.4. *Assume that the "separating powers trick"* works in $\mathbb{Z}[i\sqrt{26}]$.

The above assumption implies that there are $a, b \in \mathbb{Z}$ such that

$$\begin{aligned} y + i\sqrt{26} &= (a + i\sqrt{26}b)^3 \\ \implies y &= a^3 - 78ab^2 = a(a^2 - 78b^2) \\ 1 &= 3a^2b - 26b^3 = b(3a^2 - 26b^2). \end{aligned}$$

Hence we must have $b = \pm 1$. If $b = -1$ then $3a^2 = 25$, which has no integral solution. If $b = 1$ we find

$$3a^2 = 27 \implies a = \pm 3 \implies y = \pm 207 \implies (x, y) = (35, \pm 207).$$

We have found two non obvious solutions of our equation. However we are *missing* the easier one $(x, y) = (3, \pm 1)$!

4.3.5. *What is going on??* Assuming 4.3.2 we obtained the correct result, but Assumption 4.3.4 leads us to miss a solution! Let us try to understand:

(1) First of all, observe that the argument we used produces *correct solutions* to our equation in the second case. The reader is encouraged to go through the above steps again and check that no assumption was needed to guarantee that the solutions we obtained are indeed correct. The problem is that we are not finding *enough solutions*.
(2) In both rings $\mathbb{Z}[i\sqrt{5}]$ and $\mathbb{Z}[i\sqrt{26}]$ unique factorisation fails, as shown by the counterexamples

$$\begin{aligned} 2 \cdot 3 = 6 &= (1 + i\sqrt{5})(1 - i\sqrt{5}) \\ 3 \cdot 3 \cdot 3 = 27 &= (1 + i\sqrt{26})(1 - i\sqrt{26}). \end{aligned}$$

Hence we *cannot* expect to deduce the validity of the "separating powers trick" from a version of Euclid algorithm.
(3) On the other hand, in our arguments we never used Euclid algorithm, nor unique factorisation. The *only* thing which we need is that the "separating powers trick" can be used.

Question 4.3.6. Is there any reason why the "separating powers trick" for cubes may work in $\mathbb{Z}[i\sqrt{5}]$ but not in $\mathbb{Z}[i\sqrt{26}]$? Is there a deeper property than unique factorisation which guarantees that the trick can be used? If so, when does this property hold?

We will spend half of this course trying to answer this question, which turns out to be very deep. We end this lecture observing another general feature which showed up in all the cases we considered.

Remark 4.3.7. In all the examples studied above, the problem of finding integral solutions of equations of the form $Y^2 = X^3 + k$ was reduced to the problem of solving an equation of the form $P(X, Y) = m$ for a suitable *homogeneous polynomial $P(X, Y)$ of degree 3 in 2 variables* and an integer m (in fact, we had $m = 1$ in all the previous examples). Moreover, we always found a *finite number* of solutions.

Exercise 4.3.8. Solve the Diophantine equation $Y^2 + 1 = X^5$.

5. Lecture 5: Quadratic rings, Euclidean domains, unique factorisation domains

In this lecture we will begin our study of quadratic rings and fields, and formalise some of the phenomena we observed in the previous lectures.

5.1. **Quadratic fields.** Let $k \in \mathbb{Z}$ such that $\sqrt{k} \notin \mathbb{Z}$ (hence $\sqrt{k} \notin \mathbb{Q}$). Then

$$\mathbb{Q}(\sqrt{k}) = \{a + b\sqrt{k} \mid a, b \in \mathbb{Q}\} \subset \mathbb{C}$$

is a field. Indeed, for $\alpha = a + b\sqrt{k} \in \mathbb{Q}(\sqrt{k})$, let $\alpha' = a - b\sqrt{k}$ and define the *norm map*

$$N : \mathbb{Q}(\sqrt{k}) \to \mathbb{Q}$$
$$\alpha = a + b\sqrt{k} \mapsto N(\alpha) = \alpha\alpha' = a^2 - kb^2.$$

If $\alpha \neq 0$ then $N(\alpha) \neq 0$ and $\alpha \cdot \frac{\alpha'}{N(\alpha)} = 1$, hence $\frac{\alpha'}{N(\alpha)}$ is the multiplicative inverse of α. The map sending α to α' is an *involution* of $\mathbb{Q}(\sqrt{k})$, i.e. it is a field morphism which composed with itself gives the identity. [3] Furthermore the norm map satisfies $N(\alpha\beta) = N(\alpha)N(\beta)$ for $\alpha, \beta \in \mathbb{Q}(\sqrt{k})$.

Let us also introduce the *trace map*

$$Tr : \mathbb{Q}(\sqrt{k}) \to \mathbb{Q}$$
$$\alpha = a + b\sqrt{k} \mapsto Tr(\alpha) = \alpha + \alpha' = 2a.$$

It satisfies $Tr(\alpha + \beta) = Tr(\alpha) + Tr(\beta)$ for $\alpha, \beta \in \mathbb{Q}(\sqrt{k})$. Furthermore every $\alpha \in \mathbb{Q}(\sqrt{k})$ is a root of the polynomial

$$P_\alpha(X) = X^2 - Tr(\alpha)X + N(\alpha) \in \mathbb{Q}[X].$$

In order to study the Diophantine equation $Y^2 - k = X^3$ - as well as integers of the form $x^2 - ky^2$ - it seems natural, generalizing the approach in the previous lecture, to work with the rings $\mathbb{Z}[\sqrt{k}] = \{a + b\sqrt{k} \mid a, b \in \mathbb{Z}\} \subset \mathbb{Q}(\sqrt{k})$. However, as pointed out in Remark 4.2.2, unique factorisation fails in $\mathbb{Z}[i\sqrt{3}]$. We will give a conceptual explanation of the example given in Remark 4.2.2, and then offer a different point of view on the issue which will point us to a solution. Let us first introduce some general terminology.

5.1.1. *Algebraic terminology.* Let A be a commutative ring with unit. [4]

(1) The ring A is an *integral domain* if $1 \neq 0$ and if for every $a, b \in A$ such that $ab = 0$ we have $a = 0$ or $b = 0$. The *fraction field* of A is the field K whose elements are fractions of the form $\frac{a}{b}$ with $a \in A$ and $b \in A \setminus \{0\}$, subject to the relation $\frac{a}{b} = \frac{c}{d}$ if $ad = bc$. Sum and multiplication are performed as with usual fractions (check: these are well defined and turn K into a field).
(2) The *group of units* of A is $A^\times = \{a \in A \mid \exists b \in A, ab = 1\}$. It is an abelian group with respect to multiplication, and A is a *field* if and only if $1 \neq 0$ and $A^\times = A \setminus \{0\}$.
(3) For $a, b \in A$ we say that b divides a, and write $b \mid a$, if there exists $c \in A$ such that $a = bc$.
(4) If $a \in A$ and $u \in A^\times$, setting $b = au$ we have $a \mid b$ and $b \mid a$. The converse is true if A is an integral domain. For an integral domain A and $a, b \in A$ we write $a \sim b$, and we say that a and b are *associates*, if $a \mid b$ and $b \mid a$ (check that this is an equivalence relation).
(5) Let $a, b, m \in A$. We say that a is congruent to b modulo m, and write $a \equiv b \pmod{mA}$, if $m \mid (a - b)$.
(6) An element $a \in A \setminus \{0\}$ is *irreducible* if $a \notin A^\times$ and whenever $a = bc$ we have $b \in A^\times$ or $c \in A^\times$.
(7) An element $a \in A \setminus \{0\}$ is *prime* if $a \notin A^\times$ and whenever $a \mid bc$ we have $a \mid b$ or $a \mid c$.

Lemma 5.1.2. *Let A be an integral domain and $a \in A \setminus \{0\}$ a prime element. Then a is irreducible.*

Proof. Let $b, c \in A$ such that $a = bc$. Then $a \mid bc$, so $a \mid b$ or $a \mid c$. Without loss of generality, assume that $a \mid b$, i.e. there exists $d \in A$ such that $ad = b$. Then $a = adc$, so $a(1 - dc) = 0$ and, since A is an integral domain and a is non-zero, we have $dc = 1$. This shows that c is a unit. □

[3] If $k < 0$ then α' is the complex conjugate of α, which we denoted by $\bar{\alpha}$ in the previous lectures; however this is not anymore the case if $k > 0$, in which case complex conjugation restricts to the identity on $\mathbb{Q}(\sqrt{k})$.

[4] In this course, *ring* will *always* mean *commutative ring with unit.*

Remark 5.1.3. In general, it is *not* true that an irreducible element is a prime. For example, in $\mathbb{Z}[i\sqrt{3}]$, we have

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

As observed in Remark 4.2.2, 2 is irreducible in $\mathbb{Z}[i\sqrt{3}]$. However $2 \nmid 1 \pm \sqrt{-3}$, hence 2 is *not* prime.

5.2. **Unique factorisation domains, Euclidean domains, integrally closed domains.** The phenomenon observed in the above remark never occurs in presence of unique factorisation.

Definition 5.2.1. *An integral domain A is a* unique factorization domain *(UFD) if, for every $a \in A \smallsetminus \{0\}$, there is a factorization*

$$a = u \cdot p_1 p_2 \cdots p_r,$$

where $r \geq 0$, p_i is irreducible for $i = 1, \ldots, r$ and $u \in A^\times$. Furthermore, the factorisation is required to be unique in the following sense: if there is another factorization

$$a = v \cdot q_1 q_2 \cdots q_s,$$

with each q_i irreducible and $v \in A^\times$, then $r = s$ and, up to reordering the factors, $p_i \sim q_i$ for $i = 1, \ldots, r$.

Lemma 5.2.2. *If A is a UFD, then every irreducible element in A is prime.*

Proof. Let $a \in A \smallsetminus \{0\}$ be irreducible and $b, c \in A$ such that $a \mid bc$. We may assume b, c are both non-zero. Then there exists $d \in A \smallsetminus \{0\}$ such that $ad = bc$. As a is irreducible and A is a UFD, the factorisation of bc must contain an irreducible element $r \sim a$, which will be a factor of b or c. This implies that $a \mid b$ or $a \mid c$. □

Let us now define the class of rings where Euclid algorithm works.

Definition 5.2.3. *Let A be an integral domain and let $\lambda : A \smallsetminus \{0\} \to \mathbb{Z}_{\geq 0}$ be a function. We say that A is a* Euclidean domain *with respect to λ if for every $a, b \in A$ with $b \neq 0$ there exist $q, r \in A$ such that $a = qb + r$ and $r = 0$ or $\lambda(r) < \lambda(b)$.*

Example 5.2.4.

(1) Let K be a field. Prove that the ring $K[X]$ of polynomials with coefficients in K is Euclidean with respect to the function sending a polynomial to its degree.

(2) (cf. Problem session) Prove that the ring $\mathbb{Z}[\sqrt{2}]$ is Euclidean with respect to the function

$$\mathbb{Z}[\sqrt{2}] \to \mathbb{Z}_{\geq 0}$$
$$\alpha = a + \sqrt{2}b \mapsto |N(\alpha)| = |a^2 - 2b^2|.$$

Notice that $\mathbb{Z}[\sqrt{2}]$ is contained in $\mathbb{R}$, hence the elements of $\mathbb{Z}[\sqrt{2}]$ do not form a tiling of the plane, unlike the case of $\mathbb{Z}[i\sqrt{2}]$. However the heart of the proof of Lemma 4.2.1 can be adapted to this setting (check this!).

Question 5.2.5. Both $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[i\sqrt{2}]$ are Euclidean domains, hence enjoy several similar arithmetic properties (see the following discussion). However there is one *crucial* arithmetic difference between these rings. Can you see it? It may help trying to adapt the method of section 4.2 to solve the Diophantine equation $Y^2 = X^3 + 2$.

5.2.6. *Properties of Euclidean domains.* In Lectures 1 and 2 we saw that the fact $\mathbb{Z}$ and $\mathbb{Z}[i]$ are Euclidean domains implies that they enjoy several pleasant arithmetic properties. We will now prove that these properties hold for general Euclidean domains. Let A be a Euclidean domain (with respect to a fixed function λ).

Bézout property: Let $a, b \in A$ which are not both equal to 0. There exists an element $d \in A$ such that

$$\{ax + by \mid x, y \in A\} = dA = \{dx, x \in A\}.$$

Such a d divides both a and b. Furthermore, if $c \in A$ is an element dividing a and b then c divides d. We call d the *greatest common divisor* of a and b and write

$$d = \gcd(a, b);$$

it is well defined up to multiplication by a unit in A.

Proof. Let d be a non-zero element in

$$I = \{ax + by \mid x, y \in A\}$$

such that $\lambda(d) = \min\{\lambda(i), i \in I \setminus \{0\}\}$. By construction $dA \subset I$. On the other hand, since A is Euclidean, for every $c \in I$ we can write $c = dq + r$ with $r = 0$ or $\lambda(r) < \lambda(d)$. In the first case $d \mid c$, i.e. $c \in dA$. If $r \neq 0$ then $r = c - dq \in I \setminus \{0\}$; but this contradicts minimality of $\lambda(d)$. Hence we have proved that $I = dA$ for some $d \in A$. Such a d divides a, b; moreover if $c \mid a, b$ then c divides every element in I, hence $c \mid d$; furthermore if $d' \in A$ has the same properties as d then $d \mid d'$ and $d' \mid d$, hence $d \sim d'$. □

Euclid's Lemma: Every irreducible element in A is prime.

Proof. Let $a \in A$ be an irreducible element, and $b, c \in A$ such that $a \mid bc$. The greatest common divisor of a and b divides a, hence we have either $\gcd(a, b) \sim 1$ or $\gcd(a, b) \sim a$. If $a \nmid b$ then $\gcd(a, b) \sim 1$, hence by the Bézout property we can write $1 = bx + ay$ for some $x, y \in A$. It follows that $c = cbx + cay$, hence $a \mid c$. □

Uniqueness of factorisation: Let $A \in A \setminus \{0\}$. Suppose that

$$v \cdot q_1 q_2 \cdots q_s = a = u \cdot p_1 p_2 \cdots p_r,$$

where the elements p_i, q_j are irreducible for $i = 1, \ldots, r$, $j = 1, \ldots, s$, and $u, v \in A^\times$. Then $r = s$ and, up to reordering the factors, $p_i \sim q_i$ for $i = 1, \ldots, r$.

Proof. By induction on r, using Euclid's lemma as in the case $A = \mathbb{Z}$. □

Existence of prime factorisation: Let $a \in A \setminus \{0\}$. Then there exist irreducible elements $p_1, \ldots, p_r$ and $u \in A^\times$ such that

$$a = up_1 p_2 \cdots p_r$$

(NB: r may equal 0, in which case $p_1 \cdots p_r$ is meant to be the empty product, which equals 1).

The proof we gave for $\mathbb{Z}$ and $\mathbb{Z}[i]$ relied on the fact that these rings are Euclidean with respect to a function λ satisfying the additional properties $\lambda(ab) = \lambda(a)\lambda(b)$, $\lambda(a) \neq 0$ for $a \neq 0$ and $\lambda(a) = 1 \Leftrightarrow a \in A^\times$. While these properties hold true for several rings of interest to us, they may fail for general Euclidean domains, hence existence of factorisation requires a different proof. It is possible to prove that every Euclidean domain A is Euclidean with respect to a function $\lambda : A \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ satisfying the additional property $\lambda(a) \leq \lambda(ab)$ for every $a, b \in A \setminus \{0\}$, and use this to prove existence of factorisation with an argument close to the one we used for $\mathbb{Z}$ and $\mathbb{Z}[i]$ - see [6].[5] We will however proceed in a different way, more in the spirit of what we will learn later in the course.

Proof. Let $a \in A \setminus \{0\}$. If $a \in A^\times$ or a is irreducible, we are done. Otherwise we can write $a = a_1 b_1$ for some $a_1, b_1 \in A$ none of which is a unit. In particular we have proper inclusions $aA \subsetneq a_1 A$, $aA \subsetneq b_1 A$. If both a_1, b_1 are irreducible we are done; otherwise we can repeat the argument with a_1 and b_1 in place of a. We have to show that the process eventually stops. If it does not, then we obtain an infinite chain of *proper inclusions*

$$aA \subsetneq a_1 A \subsetneq a_2 A \ldots \subsetneq a_k A \subsetneq \ldots$$

We have to show that such a chain cannot exist in a Euclidean domain. By contradiction, suppose it exists and let $I = \cup_k a_k A \subset A$. There exists an element $c \in I \setminus \{0\}$ such that $\lambda(c) = \min\{\lambda(i), i \in I \setminus \{0\}\}$; such a c belongs to $a_{k_0} A$ for some k_0. We claim that $a_k A = a_{k_0} A$ for every $k \geq k_0$. This would give a contradiction and end the proof.

Let us prove our claim: pick any $k \geq k_0$ and $d \in a_k A$. If $c \nmid d$ then writing $d = cq + r$ we must have $r \neq 0$ hence $\lambda(r) < \lambda(c)$. But $r = d - cq$ belongs to $a_k A \subset I$, contradicting the minimality of $\lambda(c)$. Hence we must have $c \mid d$, which yields $d \in cA \subset a_{k_0} A$. □

Proposition 5.2.7 (Two consequences of unique factorisation)**.** *Let A be a UFD. Then:*

[5] One may wonder if for every Euclidean domain A it is possible to choose a function λ such that $\lambda(ab) = \lambda(a)\lambda(b)$ for every $a, b \in A \setminus \{0\}$. This was an open problem for quite some time; a counterexample was found in 2018 [5].

Separating powers: *let $n \in \mathbb{Z}_{>0}$ and $a, b, c \in A \smallsetminus \{0\}$ such that $a^n = bc$. If b and c have no common factor (up to unit) then $b = ub_1^n$ and $c = vc_1^n$ for some $u, v \in A^\times$ and $b_1, c_1 \in A \smallsetminus \{0\}$.*

UFD $\Rightarrow$ integrally closed: *Let $n \in \mathbb{Z}_{>0}$ and $f(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in A[X]$. Let $a, b \in A \smallsetminus \{0\}$ with no common irreducible factor (up to unit). If $f\left(\frac{a}{b}\right) = 0$ then $\frac{a}{b} \in A$ (i. e. $b \in A^\times$).*

Proof. The first property follows as usual from unique factorisation. Let us prove the second one. By hypothesis we have

$$\left(\frac{a}{b}\right)^n + a_{n-1}\left(\frac{a}{b}\right)^{n-1} + \ldots + a_0 = 0$$

hence multiplying by b^n we find $a^n + a_{n-1}ba^{n-1} + \ldots + a_0 b^n = 0$. It follows that

$$a^n = -b(a_{n-1}a^{n-1} + \ldots + a_0 b^{n-1})$$

hence $b \mid a^n$. By assumption there is no irreducible element dividing both a and b, hence $b \in A^\times$. □

Definition 5.2.8. *Let A be an integral domain with fraction field K. We say that A is* integrally closed *if, for every* monic *polynomial $f(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in A[X]$ and every $\frac{a}{b} \in K$, the implication*

$$f\left(\frac{a}{b}\right) = 0 \Rightarrow \frac{a}{b} \in A$$

holds.

With this terminology, Proposition 5.2.7 says that every UFD is integrally closed. This gives another explanation of the fact that $\mathbb{Z}[i\sqrt{3}]$ is not a UFD. Indeed, the element $\zeta_3 = \frac{-1+i\sqrt{3}}{2}$ belongs to the fraction field of $\mathbb{Z}[i\sqrt{3}]$, it is a root of the monic polynomial with integer coefficients $X^2 + X + 1$, but it does not belong to $\mathbb{Z}[i\sqrt{3}]$. More generally, we have

Corollary 5.2.9. *Let $k \in \mathbb{Z}$ such that $\sqrt{k} \notin \mathbb{Z}$ and $k \equiv 1 \pmod 4$. Then the ring $\mathbb{Z}[\sqrt{k}]$ is not a UFD.*

Proof. The element $\frac{1+\sqrt{k}}{2}$ belongs to $\mathbb{Q}(\sqrt{k}) \smallsetminus \mathbb{Z}[\sqrt{k}]$. Moreover it is a root of the polynomial $P_\alpha(X) = X^2 - X + \frac{1-k}{4}$, which is monic with integer coefficients. So the ring $\mathbb{Z}[\sqrt{k}]$ is not integrally closed, hence it is not a UFD by Proposition 5.2.7. □

Hence, if $k \equiv 1 \pmod 4$ the ring $\mathbb{Z}[\frac{1+\sqrt{k}}{2}]$ seems to have better properties than the ring $\mathbb{Z}[\sqrt{k}]$. The following exercise studies some properties of such a ring.

Exercise 5.2.10. (cf. Problem session)

(1) Let $k \equiv 1 \pmod 4$ such that $\sqrt{k} \notin \mathbb{Z}$. Prove that $\mathbb{Z}[\frac{1+\sqrt{k}}{2}]$ (which, by definition, is the smallest subring of $\mathbb{C}$ containing $\frac{1+\sqrt{k}}{2}$) is equal to

$$\mathbb{Z} \cdot 1 \oplus \mathbb{Z} \cdot \frac{1+\sqrt{k}}{2} = \left\{ \frac{a+b\sqrt{k}}{2} \mid a, b \in \mathbb{Z}, a \equiv b \pmod 2 \right\}.$$

(2) Prove that $\mathbb{Z}[\frac{1+\sqrt{k}}{2}]^\times = \left\{\alpha \in \mathbb{Z}[\frac{1+\sqrt{k}}{2}] \mid N(\alpha) = \pm 1\right\}$. Deduce that $\mathbb{Z}[\zeta_3]^\times = \{\zeta_6^r, r \in \mathbb{Z}/6\mathbb{Z}\}$, and, if $k < -3$ and $k \equiv 1 \pmod 4$, then $\mathbb{Z}[\frac{1+\sqrt{k}}{2}]^\times = \{\pm 1\}$.

(3) Prove that, if $\alpha \in \mathbb{Z}[\frac{1+\sqrt{k}}{2}]$ has norm $\pm p$, where p is a prime number, then α is irreducible.

(4) Prove that $\mathbb{Z}[\zeta_3]$ is a Euclidean domain with respect to $\alpha \mapsto N(\alpha)$. Does the proof generalise to other rings?

6. Lecture 6: Problem session I

(1) Study the arithmetic properties of $\mathbb{Z}[i\sqrt{2}]$ and determine numbers of the form x^2+2y^2 (= solve Exercise 4.2.3); you can assume the fact that -2 is a square modulo an odd prime p if and only if $p \equiv 1, 3 \pmod 8$.

(2) Prove that the ring $\mathbb{Z}[\sqrt{2}]$ is Euclidean with respect to the function

$$\mathbb{Z}[\sqrt{2}] \to \mathbb{Z}_{\geq 0}$$
$$\alpha = a + \sqrt{2}b \mapsto |N(\alpha)| = |a^2 - 2b^2|.$$

(3) Solve Exercise 5.2.10.

(4) Show that $\mathcal{O} = \mathbb{Z}\left[\frac{1+i\sqrt{19}}{2}\right]$ is not Euclidean in the following steps.
   (a) Let A be a Euclidean domain with Euclidean norm λ satisfying $\lambda(a) \leq \lambda(ab)$ for every $a, b \in A \smallsetminus \{0\}$ (recall that such a λ always exists). Describe the set of elements in $A \setminus \{0\}$ for which λ is minimal.
   (b) Compute $\mathcal{O}^\times$.
   (c) Show that 2 and 3 are irreducible in $\mathcal{O}$.
   (d) Assume by contradiction that $\mathcal{O}$ is Euclidean. For $a \in \mathcal{O} \setminus \{0\}$ such that $\lambda(a)$ is the second-smallest value, describe the set $\mathcal{O}/a\mathcal{O}$ using the first part of the exercise.
   (e) Deduce that $a \mid 2$ or $a \mid 3$ and get a contradiction.

(5) Reduce the matrix $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 7 & 5 \end{pmatrix}$ in Smith normal form. Then show that, for $\mathbf{b} = (b_1, b_2) \in \mathbb{Z}^2$, the system $\mathbf{A}\mathbf{x}^t = \mathbf{b}^t$ has an integral solution if and only if $b_2 \equiv 5b_1 \pmod 8$.

(6) An n-tuple $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ of vectors in $\mathbb{Z}^n$ is called a *basis* of $\mathbb{Z}^n$ if every element of $\mathbb{Z}^n$ can be written uniquely in the form $\sum_{i=1}^n a_i \mathbf{e}_i, a_i \in \mathbb{Z}$. Show that an element $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{Z}^n$ can be extended to a basis of $\mathbb{Z}^n$ if and only if $\gcd(v_1, \dots, v_n) = 1$.

Bonus exercise: algebraic properties of certain power series rings. For a real number $r > 0$ let E_r be the ring of power series with complex coefficients $\sum_{n\geq 0} a_n z^n$ whose radius of convergence is *strictly* larger than r. Let $E = \cap_{r>0} E_r$ be the ring of power series converging on the whole $\mathbb{C}$. In this exercise we study some algebraic properties of the above rings; this requires some basic results in complex analysis, which can be found in [34]. In particular you can make use of the following properties of holomorphic functions:

- Every holomorphic function on $\mathbb{C}$ can be uniquely written as a power series $\sum_{n\geq 0} a_n z^n \in E$.
- If f is a non-zero holomorphic function on $\mathbb{C}$, then the set of zeros of f has no limit point.
- Given a sequence $(z_k)_{k\geq 0}$ of complex numbers with no limit point, and a sequence $(n_k)_{k\geq 0}$ of positive integers, there exists a holomorphic function on $\mathbb{C}$ having a zero of multiplicity n_k at z_k, and no other zero.
- Given a sequence $(z_k)_{k\geq 0}$ of complex numbers with no limit point, a sequence $(n_k)_{k\geq 0}$ of positive integers, and a sequence $(v_k)_{k\geq 0}$ of complex numbers, there is a function f holomorphic on $\mathbb{C}$ such that $f(z_k) = v_k$ for every $k \geq 0$ and $f - v_k$ has a zero at z_k with multiplicity n_k.

(1) Show that every $f \in E_r \smallsetminus \{0\}$ has finitely many zeros in the closed disc $D_r = \{|z| \leq r\}$. Denote by $\lambda(f)$ the number of zeros of f in D_r, counted with multiplicity.

(2) Show that the map

$$\lambda : E_r \smallsetminus \{0\} \to \mathbb{Z}_{\geq 0}$$
$$f \mapsto \lambda(f)$$

satisfies the following properties:
   (a) $\lambda(fg) = \lambda(f) + \lambda(g)$.
   (b) $\lambda(f) = 0 \Leftrightarrow f \in E_r^\times$.

(3) Show that every $f \in E_r$ can be written as $f = P(X)u$ for a polynomial $P(X) \in \mathbb{C}[X]$ and a unit $u \in E_r^\times$.

(4) Deduce that E_r is Euclidean with respect to λ. Describe irreducible elements in E_r.

(5) Show that E is an integral domain, with fraction field the field of meromorphic functions on $\mathbb{C}$. Prove that E is integrally closed.

(6) Let $f, g \in E \setminus \{0\}$. Assume that f, g have no common zero. Prove that $\{fx + gy \mid x, y \in E\} = E$.
(7) Prove that E is a Bézout domain, i. e. for every $f, g \in E \setminus \{0\}$ there exists $d \in E$ such that $\{fx + gy \mid x, y \in E\} = \{dx, x \in E\}$.
(8) Describe units and irreducible elements in E. Prove that a non-zero element of E factors as a product of irreducible elements if and only if it has finitely many zeros. In particular deduce that E is not a UFD.

7. Lecture 7: Ideals

In this lecture we will start investigating how failure of unique factorisation and the separating powers trick may be overcome by the introduction of ideal numbers (=ideals).

7.1. **Properties of rings: summary.** In the previous lecture we introduced several abstract properties of rings related to unique factorisation, and studied the relations among them. Our discussion can be summarised by the following picture, where *Bézout domains* are integral domains which satisfy the Bézout property.

$$\begin{array}{ccccc} \text{Euclidean domains} & \subset & \text{Unique Factorisation Domains} & \subset & \text{Integrally closed domains} \\ \cap & & & & \\ \text{Bézout domains} & & & & \end{array}$$

(1) The proof of the fact that every UFD is integrally closed can be adapted to show that every Bézout domain is integrally closed.
(2) The ring $\mathbb{Z}\left[\frac{1+i\sqrt{19}}{2}\right]$ is not a Euclidean domain (cf. Problem Session), but it is a UFD.
(3) The ring $\mathbb{Z}[i\sqrt{5}]$ is integrally closed (we will see this later), but it is not a UFD. It is also not a Bézout domain. For example, the subset

$$I = \{2x + (1+i\sqrt{5})y,\ x, y \in \mathbb{Z}[i\sqrt{5}]\} \tag{7.1.0.1}$$

is not of the form $d\mathbb{Z}[i\sqrt{5}]$ for any $d \in \mathbb{Z}[i\sqrt{5}]$. Indeed, if such a d existed it should divide 2; since $\mathbb{Z}[i\sqrt{5}]$ has no element of norm 2, we would have $d \sim 2$ or $d \sim 1$. As $2 \nmid 1 + i\sqrt{5}$ we cannot have $d \sim 2$; on the other hand $I \subsetneq \mathbb{Z}[i\sqrt{5}]$ hence $d \sim 1$ can't happen either.
(4) The ring $\mathbb{Z}[\sqrt{14}]$ is Euclidean, but not with respect to $|N|$.

Question 7.1.1. (for those who like abstract algebra) What is the relation between UFDs and Bézout domains?

7.2. **Experiments in the ring $\mathbb{Z}[i\sqrt{5}]$.** Let us investigate the properties of the ring $\mathbb{Z}[i\sqrt{5}]$: motivated by question 4.3.6, we would like to understand whether a version of the "separating powers trick" holds in this ring. For example, let $\alpha, \beta, \gamma \in \mathbb{Z}[i\sqrt{5}] \smallsetminus \{0\}$. Assume that α, β have *no common factor*, i.e.

$$\forall \delta \in \mathbb{Z}[i\sqrt{5}], \delta \mid \alpha, \beta \Rightarrow \delta \in \mathbb{Z}[i\sqrt{5}]^\times, \tag{7.2.0.1}$$

and that $\alpha\beta = \gamma^3$. Is it true that $\alpha = \alpha_1^3$ and $\beta = \beta_1^3$ for some $\alpha_1, \beta_1 \in \mathbb{Z}[i\sqrt{5}]$?

As $9 = (2+i\sqrt{5})(2-i\sqrt{5})$, we have

$$3^3 = 27 = 3(2+i\sqrt{5})(2-i\sqrt{5})$$

Let $\alpha = 3(2+i\sqrt{5})$ and $\beta = (2-i\sqrt{5})$. Then

(1) if δ divides both α and β then $N(\delta) \mid 9$, hence $N(\delta)$ equals either 1 or 9. The only elements of norm 9 are ± 3 and $\pm(2 \pm i\sqrt{5})$, and none of them divides both α and β. Hence δ must be a unit;
(2) $N(\alpha) = 81$ and $N(\beta) = 9$, hence neither α nor β is a cube; on the other hand $\alpha\beta = 27$ is a cube.

Hence the answer to our question is negative. Notice that this failure appears to be related to the existence of two distinct factorisations of 9 into irreducible elements:

$$3 \cdot 3 = 9 = (2+i\sqrt{5})(2-i\sqrt{5}).$$

7.2.1. *Ideal numbers.* One may dream to solve both problems at once: perhaps the number 3 can be written as a product of two "ideal numbers"

$$3 = \mathfrak{p}\mathfrak{q} \text{ such that } 2+i\sqrt{5} = \mathfrak{p}^2, 2-i\sqrt{5} = \mathfrak{q}^2$$

so that $\alpha = \mathfrak{p}^3\mathfrak{q}$ and $\beta = \mathfrak{q}^2$. This would restore unique factorisation, and also explain why the "separating powers trick" failed: the elements α and β have a common "ideal factor", hence should not be regarded as coprime.

If we take this idea seriously, we should try to imagine what a "common divisor" of α and β could be. In a Bézout domain A, the greatest common divisor of two elements a, b is characterised (up to unit) by the fact that

$$\gcd(a,b)A = \{ax + by, x, y \in A\}. \tag{7.2.1.1}$$

We have seen above that $\mathbb{Z}[i\sqrt{5}]$ is *not* a Bézout domain. Nevertheless, we may regard the right hand side of (7.2.1.1) as a *definition of the greatest common divisor* of two elements. In other words, we may try to define an *ideal number* in $\mathbb{Z}[i\sqrt{5}]$ as a subset of the form

$$(\alpha, \beta) = \{\alpha x + \beta y, x, y \in \mathbb{Z}[i\sqrt{5}]\}.$$

Every number $\alpha \in \mathbb{Z}[i\sqrt{5}]$ gives rise to an ideal number $(\alpha, 0) = (\alpha) = \alpha\mathbb{Z}[i\sqrt{5}]$, but there are ideal numbers which do not come from any $\alpha \in \mathbb{Z}[i\sqrt{5}]$: an example is the ideal number I in (7.1.0.1). In order to talk about factorisation we would like to be able to multiply ideal numbers, extending multiplication of usual numbers. However our definition has the drawback that, while it is natural to set $(\alpha, \beta) \cdot (\gamma, \delta) = \{\alpha\gamma x + \alpha\delta y + \beta\gamma t + \beta\delta u, x, y, t, u \in \mathbb{Z}[i\sqrt{5}]\}$, it is not immediately clear whether this set is of the form (κ, λ) for some $\kappa, \lambda \in \mathbb{Z}[i\sqrt{5}]$.[6] It seems better to modify slightly our definition, declaring that an ideal number is a subset of $\mathbb{Z}[i\sqrt{5}]$ of the form

$$(\alpha_1, \dots \alpha_n) = \{\alpha_1 x_1 + \dots + \alpha_n x_n, x_1, \dots, x_n \in \mathbb{Z}[i\sqrt{5}]\} \tag{7.2.1.2}$$

for some $\alpha_1, \dots, \alpha_n \in \mathbb{Z}[i\sqrt{5}]$; we can then define

$$(\alpha_1, \dots, \alpha_n) \cdot (\beta_1, \dots, \beta_m) = (\alpha_i\beta_j)_{1 \le i \le n, 1 \le j \le m}$$

With this definition, setting as above $\alpha = 3(2 + i\sqrt{5})$ and $\beta = (2 - i\sqrt{5})$, we find

$$(\alpha, \beta) = (3(2 + i\sqrt{5}), 2 - i\sqrt{5}) = (3, 1 + i\sqrt{5})$$

Indeed, $3 - (1 + i\sqrt{5}) = 2 - i\sqrt{5}$ and $3 + 3(1 + i\sqrt{5}) = 3(2 + i\sqrt{5})$, hence $(3, 1 + i\sqrt{5}) \supset (\alpha, \beta)$. Conversely,

$$3 = 12 - 9 = 3(2 + i\sqrt{5}) + 3(2 - i\sqrt{5}) - (2 - i\sqrt{5})(2 + i\sqrt{5})$$
$$1 + i\sqrt{5} = 3 - (2 - i\sqrt{5}).$$

Furthermore, setting $\mathfrak{q} = (3, 1 + i\sqrt{5})$ and $\mathfrak{p} = (3, 1 - i\sqrt{5})$ we have

$$\mathfrak{q}\mathfrak{p} = (3, 1 + i\sqrt{5}) \cdot (3, 1 - i\sqrt{5}) = (9, 3(1 - i\sqrt{5}), 3(1 + i\sqrt{5}), 6) = (3)$$

as we wanted! Let us check if the other predictions we made are correct:

$$\begin{aligned}\mathfrak{q}^2 &= (9, 3 + 3i\sqrt{5}, -4 + 2i\sqrt{5}) \supset (9, 7 + i\sqrt{5}) \supset (2 - i\sqrt{5}) \\ &\supset (N(2 - i\sqrt{5}), -2(2 - i\sqrt{5}), -3(2 - i\sqrt{5})) \supset (9, -4 + 2i\sqrt{5}, -3(2 - i\sqrt{5}) + 9) \\ &= \mathfrak{q}^2,\end{aligned}$$

hence $\mathfrak{q}^2 = (\beta)$. One can check directly that $(\alpha) = \mathfrak{p}^3\mathfrak{q}$ (exercise), or observe that $(\alpha) = (3\bar{\beta}) = \mathfrak{p}\mathfrak{q}\bar{\mathfrak{q}}^2 = \mathfrak{p}\mathfrak{q}\mathfrak{p}^2$.

Exercise 7.2.2. Express 2 and 3 as products of ideal numbers, and verify that the two factorisations $2 \cdot 3 = 6 = (1+i\sqrt{5})(1-i\sqrt{5})$ can be refined into the same factorisation as product of ideal numbers.

7.3. **Ideals.** Our calculations suggest that ideal numbers as defined in (7.2.1.2) are the objects we were looking for, suggesting a refinement of the "separating powers trick" and restoring the lost property of unique factorisation. Proving that this happens in general will require some non trivial work - but the point of the above experiments was to convince you that ideal numbers should be the right objects to work with. However, although we seem to have discovered the correct objects, you may notice something a bit dissatisfying in our current definition of ideal number: the issue is that, as it already emerged in the previous computations, a given ideal number may have several distinct presentations in the form $(\alpha_1, \dots, \alpha_n)$. It would be desirable to have a more intrinsic definition of these objects.[7] To find it you may review the material we discussed so far in the

[6] It turns out that this is the case, and it is not too hard to prove, as we will soon see. If you can already show this, you may try to replace $\mathbb{Z}[i\sqrt{5}]$ by $\mathbb{Z}[\zeta_p]$; the statement is still true but harder.

[7] When one is trying to understand new mathematical objects it is crucial, and by no means easy, to single out their key properties and capture them via a good definition.

course, look for situations where ideal numbers have already appeared, and check which of their properties we crucially used in our arguments. You will notice that ideal numbers showed up when we proved the Bézout property of $\mathbb{Z}$ and more generally of Euclidean domains (see 5.2.6). With our new terminology, what we proved amounts to say that every ideal number I in a Euclidean domain A is of the form $I = (a)$ for some $a \in A$. In fact when proving existence of factorisation we established the same property also for more general subsets $I \subset A$. In our arguments however we did not quite use the definition of ideal numbers we gave, but only the following two properties:

(1) $a, b \in I \Rightarrow a + b \in I$
(2) $a \in I, q \in A \Rightarrow qa \in I$.

This motivates the following

Definition 7.3.1. *Let A be a ring. An* ideal *of A is an additive subgroup $I \subset A$ such that $aI \subset I$ for every $a \in A$. In other words, it is an A-submodule of A. An ideal I is called* principal *if it is of the form $I = (a) = aA$ for some $a \in A$.*

Notation 7.3.2. Let I be an ideal in a ring A. We say that two elements $a, b \in A$ are *congruent modulo* I if $a - b \in I$; we denote this by

$$a \equiv b \pmod{I}.$$

Exercise 7.3.3. Go back to 5.2.6 and verify that the argument used to prove the Bézout property proves that every ideal in a Euclidean domain is principal. Furthermore, show that an integral domain A is a UFD if and only if every irreducible element in A is prime and A does not contain infinite ascending chains of principal ideals.

Remark 7.3.4. Let $A = \mathbb{Z}[i\sqrt{5}]$; clearly every ideal number of A is an ideal. Conversely, let $I \subset A$ be a non zero ideal. Let $\alpha \in I \smallsetminus \{0\}$ and $a = N(\alpha) = \alpha\bar{\alpha} \in \mathbb{Z} \smallsetminus \{0\}$; then $aA \subset I$. Notice that A/aA is finite, as every element $x + i\sqrt{5}y \in A$ has a representative in A/aA such that $0 \leq x, y < a$. It follows that $aA \subset I$ has finite index, so I is of the form (7.2.1.2). Hence ideals and ideal numbers are the same thing in this case; we will see that this holds true more generally. In fact, we will later show (see 25.2.5) that for most rings of interest to us every ideal is of the form (α, β), so for our purposes our original definition was correct after all.

7.3.5. *Prime ideals.* The next step is to define what *prime ideals* should be. The idea is to impose that they satisfy a version of Euclid lemma:

Definition 7.3.6. *Let A be a ring.*

(1) *An ideal $I \subset A$ is called a* prime ideal *if $I \neq A$ and, for every $a, b \in A$, we have*

$$ab \in I \Rightarrow a \in I \text{ or } b \in I.$$

(2) *An ideal $I \subset A$ is a* maximal ideal *if $I \neq A$ and for every ideal $I \subset J \subset A$ we have $J = I$ or $J = A$.*

The set of prime ideals of a ring A is denoted by Spec A.

The algebraist's point of view on ideals is the following:

Proposition 7.3.7. (1) *Let A, B be rings and $f : A \to B$ be a morphism of rings (i. e. $f(1) = 1$ and for all $a, b \in A$, $f(a + b) = f(a) + f(b)$ and $f(ab) = f(a)f(b)$). Then*

$$ker(f) = \{a \in A \mid f(a) = 0\} \subset A$$

is an ideal.

(2) *If $I \subset A$ is an ideal then the expressions*

$$(a + I) + (b + I) = (a + b) + I,$$
$$(a + I)(b + I) = ab + I$$

are well-defined operations that make A/I into a ring. The natural map $A \to A/I$ is a surjective ring homomorphism.

(3) *In the situation of* (1)*, the subset $f(A) \subset B$ is a subring, and f induces a ring isomorphism $A/ker(f) \simeq f(A)$.*

Proof. Exercise. □

Lemma 7.3.8. *Let A be a ring and $I \subset A$ an ideal.*

$$I \text{ is prime } \Leftrightarrow A/I \text{ is an integral domain}$$
$$I \text{ is maximal } \Leftrightarrow A/I \text{ is a field};$$

in particular, every maximal ideal is prime.

Proof. The ring A/I is a domain if and only if $A/I \neq 0$ and, for every $a, b \in A$ such that $ab \equiv 0 \pmod{I}$ either $a \equiv 0 \pmod{I}$ or $b \equiv 0 \pmod{I}$; in other words, A/I is a domain if and only if $I \neq A$ and $ab \in I \Rightarrow a \in I$ or $b \in I$, which are the conditions defining prime ideals.

If I is maximal and $a \in A \smallsetminus I$ then $J = \{ax + i, x \in A, i \in I\}$ is an ideal of A such that $I \subsetneq J$, hence $J = A$. It follows that we can write $1 = ab + i$ for some $b \in A$ and $i \in I$. In other words, $ab \equiv 1 \pmod{I}$. We have proved that every non zero element in A/I is a unit, hence A/I is a field.

Conversely, assume that there exists an ideal J such that $I \subsetneq J \subsetneq A$ and pick $j \in J \smallsetminus I$. We claim that the image of j in A/I is not invertible. Indeed, if we had $ja \equiv 1 \pmod{I}$ for some $a \in A$ then $1 \in ja + I \subset J$ hence $J = A$, contradiction. □

Remark 7.3.9. (for the attentive reader) Notice that in 5.1.1 we declared that integral domains and fields are *non zero* rings; consistently, we require that prime and maximal ideals are *proper* ideals, i.e. they are not equal to the full ring. A motivation for this convention comes from unique factorisation: we do not want (1) to be regarded as a prime ideal, much in the same way as 1 is not a prime number in $\mathbb{Z}$.

Remark 7.3.10. While in this lecture we have focused on the ring $\mathbb{Z}[i\sqrt{5}]$, you are encouraged to make further experiments with other quadratic rings. For example, you can show that none of the rings $\mathbb{Z}[\sqrt{-k}], k \geq 3$ is a UFD. The outcome of the experiments in this lecture is that for such rings we should investigate unique factorisation of *ideals* instead of *numbers*; this will be our main aim in the next few lectures. Nonetheless, being a UFD is an important property enjoyed by several other interesting rings, such as the ring of polynomial functions on the sphere, $\mathbb{R}[X, Y, Z]/(X^2 + Y^2 + Z^2 - 1)$. More information about UFDs can be found in the survey [30].

8. Lecture 8: More on ideals

In this lecture we will study further properties of ideals and investigate how prime numbers decompose in certain imaginary quadratic rings.

8.1. **Operations with ideals.** Let A be a ring and $I, J \subset A$ two ideals. We define

$$I + J = \{r + s \mid r \in I, s \in J\}$$
$$I \cap J = \{r \mid r \in I, r \in J\}$$
$$IJ = \left\{ \sum_{i=1}^{n} r_i s_i \mid r_i \in I, s_i \in J, n \in \mathbb{Z}_{>0} \right\}.$$

The ideal $I+J$ is the smallest ideal containing I and J, and can be thought of as a *greatest common divisor* of I and J. If $a_1, \ldots, a_n \in A$ then $(a_1) + \ldots + (a_n)$ is the *ideal generated by* $a_1, \ldots, a_n$, and is denoted by $(a_1, \ldots, a_n)$ (generalising (7.2.1.2)). An ideal I is called *finitely generated* if it is of the form $I = (a_1, \ldots, a_n)$ for some $n \geq 0$ and $a_1, \ldots, a_n \in A$.

Dually, $I \cap J$ is a *least common multiple* of I and J. Notice that $IJ \subset I \cap J$, and the inclusion is proper in general: for example, if $A = \mathbb{Z}$, then $(2) \cdot (4) = (8) \subsetneq (2) \cap (4) = (4)$.

Definition 8.1.1. *Let A be a ring and I, J ideals of A. We say that I and J are* relatively prime ideals *(or* coprime *ideals) if $I + J = A$.*

Proposition 8.1.2. *(Chinese remainder theorem) Let A be a ring and $I, J \subset A$* coprime *ideals. Then $IJ = I \cap J$, and the natural map $A \to A/I \times A/J$ induces an isomorphism*

$$A/IJ \xrightarrow{\sim} A/I \times A/J.$$

Proof. If I and J are coprime we can write $1 = i + j$ for some $i \in I, j \in J$. Let $x \in I \cap J$. Then

$$x = x \cdot 1 = x(i + j) = ix + jx \in IJ + JI = IJ$$

hence $IJ = I \cap J$. Since the kernel of the natural map $A \to A/I \times A/J$ is $I \cap J$, we deduce that the induced map $p : A/IJ \to A/I \times A/J$ is injective. On the other hand $p(j) = (1, 0)$ and $p(i) = (0, 1)$, hence the map p is surjective. □

Example 8.1.3. If $A = \mathbb{Z}$ then every ideal is of the form (m) for some $m \in \mathbb{Z}$, and the above result is the usual Chinese remainder theorem. Concretely, it asserts that if $m, n \in \mathbb{Z}_{>0}$ are coprime then for every $a, b \in \mathbb{Z}$ the congruences

$$X \equiv a \pmod{m}$$
$$X \equiv b \pmod{n}$$

have a unique solution modulo mn.

Exercise 8.1.4. Let $P(X) \in \mathbb{Z}[X]$.

(1) Show that the following assertions are equivalent:
   (a) The equation $P(X) = 0$ has a solution in $\mathbb{Z}/n\mathbb{Z}$ for every $n \geq 1$.
   (b) For every prime p and every integer $r \geq 1$ the equation $P(X) = 0$ has a solution in $\mathbb{Z}/p^r\mathbb{Z}$.

(2) Now fix a prime p. Show that the following statements are equivalent:
   (a) For every integer $r \geq 1$ the equation $P(X) = 0$ has a solution in $\mathbb{Z}/p^r\mathbb{Z}$.
   (b) There is a sequence of integers $(x_r)_{r \geq 1}$ such that $P(x_r) \equiv 0 \pmod{p^r}$ for every $r \geq 1$ and $x_r \equiv x_{r-1} \pmod{p^{r-1}}$ for every $r \geq 2$.

(3) (Bonus) Let $P(X) = X^2 + 1$, $p = 5$ and $x_1 = 2$. Can you construct a sequence $(x_r)_{r \geq 1}$ as in (2)(b) with $x_1 = 2$? How could you represent such a sequence as a "unique number"? (Hint: have a look at 1.2.3)

8.2. **More properties of rings.** Rings in which all ideals are of the form (7.2.1.2) deserve a special name.

Definition 8.2.1. *Let A be a ring.*

(1) *A is called a* noetherian ring *if every ideal of A is finitely generated. Equivalently, A is noetherian if every infinite ascending chain of ideals of A*

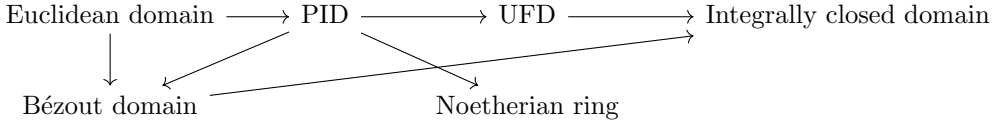
$$I_1 \subset I_2 \subset \ldots \subset I_k \subset \ldots$$

is stationary, i.e. there exists k_0 such that $I_k = I_{k_0}$ for every $k \geq k_0$.

(2) *A is called a* principal ideal domain *(PID) if it is an integral domain and every ideal of A is principal.*

Exercise 8.2.2. Check that the two conditions defining Noetherian rings are indeed equivalent. Show that a Bézout domain is Noetherian if and only if it is a PID. Finally, adapt the proof of Euclidean ⇒ UFD in 5.2.6 to show that every PID is a UFD.

Let us refine the picture at the beginning of Lecture 7; the algebraically inclined reader may check that no arrow in the following picture can be reversed, nor did we forget any arrow.



8.3. **Examples.**

8.3.1. *The ring* $\mathbb{Z}[i]$. We proved in Lecture 3 that $\mathbb{Z}[i]$ is a Euclidean domain, hence it is a PID. A non-zero ideal $(\alpha) \subset \mathbb{Z}[i]$ is prime if and only if $\alpha \in \mathbb{Z}[i]$ is prime ($\Leftrightarrow$ irreducible). Furthermore, we showed in 3.2.2 that $2\mathbb{Z}[i] = (1+i)^2\mathbb{Z}[i]$, and for $p \neq 2$ prime the following equivalences hold:

$$p\mathbb{Z}[i] \text{ is prime } \Leftrightarrow p \equiv 3 \pmod 4 \Leftrightarrow p \text{ is not of the form } a^2+b^2$$
$$p\mathbb{Z}[i] = (\pi)(\bar{\pi}), (\pi) \text{ prime } \Leftrightarrow p \equiv 1 \pmod 4 \Leftrightarrow p \text{ is of the form } a^2+b^2.$$

8.3.2. *The ring* $\mathbb{Z}[\zeta_3]$. The ring $\mathbb{Z}[\zeta_3]$ is Euclidean (cf. exercise 5.2.10), hence a PID; let us determine the factorisation of prime numbers in $\mathbb{Z}[\zeta_3]$. We have $3\mathbb{Z}[\zeta_3] = (i\sqrt{3}\mathbb{Z}[\zeta_3])^2$. Let p be a prime different from 3. Then p is irreducible in $\mathbb{Z}[\zeta_3]$ if and only if $p\mathbb{Z}[\zeta_3]$ is a prime ideal; by Lemma 7.3.8 this is equivalent to the fact that $\mathbb{Z}[\zeta_3]/(p)$ is an integral domain. Now

$$\mathbb{Z}[\zeta_3] = \mathbb{Z}[X]/(X^2+X+1) \Rightarrow \mathbb{Z}[\zeta_3]/(p) = \mathbb{F}_p[X]/(X^2+X+1).$$

Hence there are two alternatives: either $X^2+X+1 \in \mathbb{F}_p[X]$ is irreducible, in which case p is irreducible in $\mathbb{Z}[\zeta_3]$, or $X^2+X+1 \in \mathbb{F}_p[X]$ has a root in $\mathbb{F}_p$, in which case p must factor in $\mathbb{Z}[\zeta_3]$ as a product of two irreducible elements: $p = \pi\bar{\pi}$. Furthermore in the latter case writing $\pi = a + \zeta_3 b$ with $a, b \in \mathbb{Z}$ we must have $N(\pi) = a^2 - ab + b^2 = p$.

As $X^2+X+1 = \frac{X^3-1}{X-1}$ we see that X^2+X+1 has a root in $\mathbb{F}_p$ if and only if $\mathbb{F}_p^\times$ contains an element of order 3, which happens if and only if $p \equiv 1 \pmod 3$ - to show this, use either that the group $\mathbb{F}_p^\times$ is cyclic or that $X^{p-1}-1$ has all its roots in $\mathbb{F}_p^\times$. Hence we find that, for a prime p different from 3:

$$p\mathbb{Z}[\zeta_3] \text{ is prime } \Leftrightarrow p \equiv 2 \pmod 3 \Leftrightarrow p \text{ is not of the form } a^2-ab+b^2$$
$$p\mathbb{Z}[\zeta_3] = (\pi)(\bar{\pi}), (\pi) \text{ prime } \Leftrightarrow p \equiv 1 \pmod 3 \Leftrightarrow p \text{ is of the form } a^2-ab+b^2.$$

8.3.3. *The ring* $\mathbb{Z}[i\sqrt{5}]$. Let us now examine the ring $A = \mathbb{Z}[i\sqrt{5}] = \mathbb{Z}[X]/(X^2+5)$. By Lemma 7.3.8 the ideal generated by a prime p is a prime ideal if and only if $A/(p)$ is an integral domain. As

$$A/(p) = \mathbb{F}_p[X]/(X^2+5)$$

we see that:

(1) The ideal $5A$ is not prime, since $\mathbb{F}_5[X]/(X^2+5) = \mathbb{F}_5[X]/(X^2)$ is not an integral domain. In fact we have $X \neq 0 \in \mathbb{F}_5[X]/(X^2)$ and $X^2 = 0 \in \mathbb{F}_5[X]/(X^2)$, i. e. $X \in \mathbb{F}_5[X]/(X^2)$ is a non-zero nilpotent element.
The equality $5A = (i\sqrt{5}A)^2$ holds, and $(i\sqrt{5}A)$ is prime.

(2) Similarly, $\mathbb{F}_2[X]/(X^2+5) = \mathbb{F}_2[X]/(X+1)^2$ is not an integral domain, hence $2A$ is not prime. In fact, $2A = (2, 1+i\sqrt{5})^2$ and $(2, 1+i\sqrt{5})$ is prime since $A/(2, 1+i\sqrt{5}) = \mathbb{F}_2[X]/(X+1) = \mathbb{F}_2$.

(3) Let p be a prime coprime to 10. If $\left(\frac{-5}{p}\right) = -1$ then X^2+5 is irreducible in $\mathbb{F}_p[X]$, hence $\mathbb{F}_p[X]/(X^2+5)$ is an integral domain (better, a field) and pA is prime (better, maximal).
(4) If $\left(\frac{-5}{p}\right) = 1$ then, letting $a \in \mathbb{Z}$ be an element such that $a^2 \equiv -5 \pmod p$, we have $X^2+5 = (X-a)(X+a) \in \mathbb{F}_p[X]$. Using the Chinese remainder theorem we find

$$A/(p) = \mathbb{F}_p[X]/(X-a) \times \mathbb{F}_p[X]/(X+a)$$

and $pA = (p, i\sqrt{5}+a)(p, i\sqrt{5}-a)$. As $A/(p, i\sqrt{5}\pm a) = \mathbb{F}_p[X]/(X \pm a) \simeq \mathbb{F}_p$ we see that the ideals $(p, i\sqrt{5}\pm a)$ are maximal, hence prime.

A special case of the quadratic reciprocity law asserts that, for a prime p coprime to 10:

$$\left(\frac{-5}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{5}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{5}\right) = 1 \Leftrightarrow \begin{cases} p \equiv 1 \pmod 4 \text{ and } & p \equiv \pm 1 \pmod 5 \\ p \equiv -1 \pmod 4 \text{ and } & p \equiv \pm 2 \pmod 5 \end{cases};$$

it follows that for such a p

$$p\mathbb{Z}[i\sqrt{5}] \text{ is prime } \Leftrightarrow p \equiv 11, 13, 17, 19 \pmod{20}$$
$$p\mathbb{Z}[i\sqrt{5}] = \mathfrak{p}_1\mathfrak{p}_2, \quad \mathfrak{p}_1, \mathfrak{p}_2 \text{ prime ideals } \Leftrightarrow p \equiv 1, 3, 7, 9 \pmod{20}.$$

Remark 8.3.4. Let A be one of the rings $\mathbb{Z}[i], \mathbb{Z}[\zeta_3], \mathbb{Z}[i\sqrt{5}]$.

(1) We saw above that prime numbers either generate prime ideals in A or split as products of two (possibly equal) prime ideals. Furthermore, as a consequence of quadratic reciprocity we found that a suitable *congruence condition* on primes determines which of the cases occurs.
(2) For $A = \mathbb{Z}[i]$ or $\mathbb{Z}[\zeta_3]$ we found a *quadratic form* $f_A(X,Y) = tX^2 + uXY + vY^2 \in \mathbb{Z}[X,Y]$ such that, for every prime p:

$$pA \text{ is not prime } \Leftrightarrow \exists\, a, b \in \mathbb{Z} \mid p = f_A(a,b).$$

What about $A = \mathbb{Z}[i\sqrt{5}]$? It seems reasonable to look at the form $f_A(X,Y) = X^2+5Y^2$. However notice that, if p is a prime coprime to 10,

$$p = a^2+5b^2,\ a, b \in \mathbb{Z} \Rightarrow p \equiv a^2+b^2 \equiv 1 \pmod 4$$
$$p \equiv a^2 \equiv \pm 1 \pmod 5.$$

Hence

$$p = a^2+5b^2 \Rightarrow p \equiv 1, 9 \pmod{20}; \tag{8.3.4.1}$$

every prime $p \equiv 3, 7 \pmod{20}$ splits as a product of two prime *ideals* in $\mathbb{Z}[i\sqrt{5}]$, but it is not of the form a^2+5b^2. The point is that

$$p = a^2+5b^2,\ a, b \in \mathbb{Z} \Leftrightarrow p\mathbb{Z}[i\sqrt{5}] = \mathfrak{p}_1\bar{\mathfrak{p}}_1, \quad \mathfrak{p}_1 = (a+i\sqrt{5}b) \text{ prime } \textit{principal} \text{ ideal}$$

and the issue is that not every prime ideal in $\mathbb{Z}[i\sqrt{5}]$ is principal. We will see that the implication in (8.3.4.1) is in fact an equivalence; in other words, p splits as a product of two *principal* prime ideals if and only if $p \equiv 1, 9 \pmod{20}$.

8.3.5. *The ring* $\mathbb{Z}[i\sqrt{3}]$. Finally, let us give one more reason why the ring $A = \mathbb{Z}[i\sqrt{3}]$ has worse arithmetic properties than $\mathbb{Z}[\zeta_3]$. We claim that the ideal $2A$ does *not* factor as a product of prime ideals in A. Indeed, we have $A = \mathbb{Z}[X]/(X^2+3)$ hence $A/2A = \mathbb{F}_2[X]/(X^2+3) = \mathbb{F}_2[X]/(X+1)^2$, so the ideal $2A$ is not prime. If we could factor it as a product of prime ideals, each of them should contain $2A$. There is only one prime ideal in the ring $A/2A = \mathbb{F}_2[X]/(X+1)^2$, generated by $X+1$, hence the only prime ideal containing $2A$ is $I = (2, i\sqrt{3}+1)$. Now we compute

$$I^2 = (4, 2+2i\sqrt{3}, -2+2i\sqrt{3}) = (2)(2, 1+i\sqrt{3}) \subsetneq (2)$$

hence no power of I equals (2).

Exercise 8.3.6. Show that primes p other than 2 behave in $A = \mathbb{Z}[i\sqrt{3}]$ as they do in $\mathbb{Z}[\zeta_3]$, i. e.:

(1) The ideal $3A$ is the square of a principal prime ideal.
(2) If $p \equiv 2 \pmod 3$ and p is different from 2 then pA is a prime ideal.
(3) If $p \equiv 1 \pmod 3$ then $pA = (\pi)(\bar{\pi})$ for some prime ideal $(\pi) \subset A$.

9. Lecture 9: Back to the Mordell equation

In this lecture we will formulate a refined version of the separating powers trick, and show that, when it holds, it allows to find all the integral solutions of certain Mordell equations. We will also take a quick look at new phenomena arising when looking at rational points on Mordell curves.

9.1. **A refined separating powers trick.** Let us now come back to our study of the Mordell equation $Y^2 + k = X^3$. The case $k = 5$ led us to study the "separating powers trick" in $\mathbb{Z}[i\sqrt{5}]$ (cf. 7.2); however in the formulation given in (7.2.0.1) it failed. The outcome of 7.2 is that the right notion of coprime elements should not be that their only common factors are units, but instead that the ideal they generate is the full ring. This suggests the following refinement of the "separating powers trick" condition.

Definition 9.1.1. *Let $n \geq 1$. We say that a ring A satisfies property $SP(n)$ (=separating powers trick with exponent n) if for every $a, b \in A \smallsetminus \{0\}$ such that $(a, b) = A$, if*

$$ab = c^n \text{ for some } c \in A \smallsetminus \{0\}$$

then there exist $a_1, b_1 \in A$ and $u, v \in A^\times$ such that $a = ua_1^n$ and $b = vb_1^n$.

Proposition 9.1.2. *Let $k \in \mathbb{Z}_{>0}$ be squarefree and congruent to 1 or 2 modulo 4. Assume that the ring $\mathbb{Z}[i\sqrt{k}]$ satisfies property $SP(3)$. Then:*

1. *if there exists an integer a such that $k = 3a^2 \pm 1$ then the only integer solutions of the equation $Y^2 + k = X^3$ are*
$$(x, y) = (a^2 + k, \pm a(a^2 - 3k)).$$
2. *If there is no integer a such that $k = 3a^2 \pm 1$ then the equation $Y^2 + k = X^3$ has no integer solution.*

Proof. Let $x, y \in \mathbb{Z}$ be such that $y^2 + k = x^3$. Reducing modulo 4 we see that x must be odd. Furthermore $\gcd(k, x) = 1$. Indeed, if there was a prime p dividing k and x then $p \mid y$ hence $p^2 \mid x^3 - y^2 = k$, contradicting the assumption that k is squarefree.

Write

$$(y + i\sqrt{k})(y - i\sqrt{k}) = x^3; \tag{9.1.2.1}$$

let $A = \mathbb{Z}[i\sqrt{k}]$. We claim that $I = (y + i\sqrt{k}, y - i\sqrt{k})$ equals A. Indeed I contains $2i\sqrt{k}$, hence $4k$; furthermore $x^3 = (y + i\sqrt{k})(y - i\sqrt{k}) \in I$. As x is coprime with $4k$ we deduce that $1 \in I$, hence $I = A$.

We can restrict ourselves to $k > 1$: the case $k = 1$ was already dealt with in 4.1.1. As $k > 1$ the only integers (m, n) satisfying $m^2 + kn^2 = 1$ are $(\pm 1, 0)$, hence $A^\times = \{\pm 1\} = (A^\times)^3$. This fact and the assumption that $SP(3)$ holds true in A imply that there exist $a, b \in \mathbb{Z}$ such that $y + i\sqrt{k} = (a + ib\sqrt{k})^3$, hence

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

The second equation forces $b = \pm 1$ hence $k = 3a^2 \pm 1$. Hence if k is not of the form $3a^2 \pm 1$ there is no integer solution to (9.1.2.1). If there exists $a \geq 0$ such that $k = 3a^2 \pm 1$ then we find $y = \pm a(a^2 - 3k)$ and $x = a^2 + k$. □

Example 9.1.3.
1. Let $k = 109$. Then $k - 1 = 3 \cdot 6^2$ hence we can take $a = 6$. If the ring $\mathbb{Z}[i\sqrt{109}]$ satisfied $SP(3)$, then the *only* integral solutions to $Y^2 = X^3 - 109$ should be $(145, \pm 1746)$. However $109 + 16 = 125$ hence $(5, \pm 4)$ are further solutions of our equation. It follows that $\mathbb{Z}[i\sqrt{109}]$ does *not* satisfy $SP(3)$.
2. Let $k = 61$. Then k is not of the form $3a^2 \pm 1$ for any $a \in \mathbb{Z}$. Nonetheless we have $8^2 = 5^3 - 61$ hence $(5, \pm 8)$ are solutions of the equation $Y^2 = X^3 - 61$. So the ring $\mathbb{Z}[i\sqrt{61}]$ does not satisfy $SP(3)$.
3. Let $k = 170$; then k is not of the form $3a^2 \pm 1$, but the equation $Y^2 = X^3 - 170$ has the integral solutions $(59, \pm 453)$. Hence $\mathbb{Z}[i\sqrt{170}]$ does not satisfy $SP(3)$.

Question 9.1.4. How do we determine whether a ring of the form $\mathbb{Z}[i\sqrt{k}]$, for $k \equiv 1, 2 \pmod 4$ squarefree, satisfies $SP(3)$? How do we find the missing solutions when $SP(3)$ is not satisfied?

Let us examine again the proof of Proposition 9.1.2 and think about the first question. The beginning of the argument works in general, showing that if $y^2 = x^3 - k$ then $(y + i\sqrt{k}, y - i\sqrt{k}) = \mathbb{Z}[i\sqrt{k}]$. Our computations in the last lecture seem to suggest that unique factorisation of ideals of $\mathbb{Z}[i\sqrt{k}]$ into a product of prime ideals may hold, if $k > 0$ is squarefree and *not* congruent to 3 modulo 4. If we believe that this is the case, then we deduce that there exists an ideal $\mathfrak{a} \subset \mathbb{Z}[i\sqrt{k}]$ such that $(y + i\sqrt{k}) = \mathfrak{a}^3$. We would like to deduce from this that $\mathfrak{a}$ is a *principal* ideal. Hence our first question is reduced (modulo believing in unique factorisation of ideals) to:

Question 9.1.5. Which rings $\mathbb{Z}[i\sqrt{k}]$ (with $0 < k \equiv 1, 2 \pmod 4$ squarefree) have the property that, for every ideal $\mathfrak{a}$,

$$\mathfrak{a}^3 \text{ principal } \Rightarrow \mathfrak{a} \text{ principal?}$$

In the next lectures we will study more in depth ideals in quadratic rings and give a precise measure of failure of unique factorisation, which will allow us to answer this question.

9.2. **Constructing new (rational) solutions from old ones.** Up to now we have always restricted to the study of *integral* solutions of Diophantine equations. In this section we wish to examine *rational* solutions of certain Diophantine equations, highlighting some new phenomena.

9.2.1. *Rational points on the circle.* In lecture 2 we studied the integral solutions of the equation $X^2 + Y^2 = n$ for a given positive integer n. Let us now determine the *rational* solutions. Those correspond to elements $x + iy \in \mathbb{Q}(i)$ of norm n. Notice that if α, β are two such elements, then $\frac{\alpha}{\beta} \in \mathbb{Q}(i)$ is an element of norm 1. Conversely if $u \in \mathbb{Q}(i)$ satisfies $N(u) = 1$ and $x + iy$ has norm n then $u(x + iy)$ has norm n. Hence rational solutions of the equation $X^2 + Y^2 = n$, if they exist, are in bijection with the set

$$\{u \in \mathbb{Q}(i) \mid N(u) = 1\}.$$

Unique factorisation in $\mathbb{Z}[i]$ plus the description of irreducible elements in $\mathbb{Z}[i]$ imply that every element $u \in \mathbb{Q}(i)$ can be written uniquely in the form

$$u = i^k (1+i)^a \prod_{p \equiv 1 \pmod 4} \pi_p^{b_p} \bar{\pi}_p^{c_p} \prod_{p \equiv 3 \pmod 4} p^{d_p}, \quad k \in \mathbb{Z}/4\mathbb{Z},\ a, b_p, c_p, d_p \in \mathbb{Z}$$

where π_p was defined in Theorem 3.2.3. We obtain

$$N(u) = 2^a \prod_{p \equiv 1 \pmod 4} p^{c_p + b_p} \prod_{p \equiv 3 \pmod 4} p^{2d_p}$$

hence $N(u) = 1$ if and only if $a = 0$, $c_p = -b_p$ for $p \equiv 1 \pmod 4$ and $d_p = 0$ for $p \equiv 3 \pmod 4$. It follows from this (and the fact that $i = \frac{1+i}{1-i}$) that

$$\{u \in \mathbb{Q}(i) \mid N(u) = 1\} = \left\{ \frac{\alpha}{\bar{\alpha}}, \alpha \in \mathbb{Z}[i] \right\} = \left\{ \frac{a^2 - b^2}{a^2 + b^2} + i \frac{2ab}{a^2 + b^2}, (a, b) \in \mathbb{Z}^2 \smallsetminus \{0\} \right\}. \tag{9.2.1.1}$$

In particular, the equation $X^2 + Y^2 = n$ has either zero or *infinitely many* rational solutions, whereas of course it has only finitely many integral solutions.

Exercise 9.2.2. If you like algebra, apply Hilbert Theorem 90 to to extension $\mathbb{Q}(i)/\mathbb{Q}$ to give another proof of (9.2.1.1). If you prefer geometry, intersect the family of lines in the plane passing through $(1, 0)$ with the circle $X^2 + Y^2 = 1$ to prove (9.2.1.1). In both cases, try to generalise your argument to other situations.

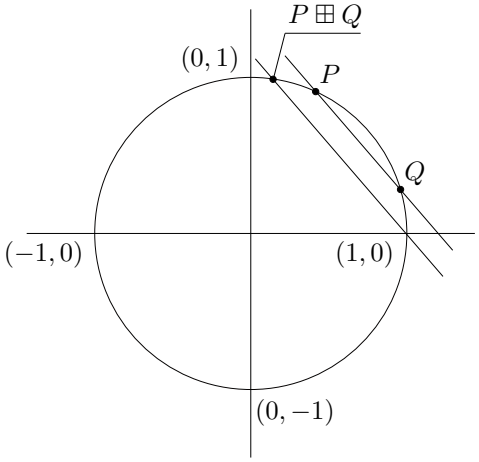
The group law. The unit circle

$$S^1 = \{z \in \mathbb{C} \mid z\bar{z} = 1\} \subset \mathbb{C} \smallsetminus \{0\}$$

inherits a natural group structure from $\mathbb{C}^\times = \mathbb{C} \smallsetminus \{0\}$. If $z = x + iy$ and $z' = x' + iy'$, then $zz' = (xx' - yy') + i(xy' + x'y)$; the inverse of $z = x + iy$ is $\bar{z} = x - iy$. Since the coordinates of zz' and z^{-1} are *polynomial expressions* with integer coefficients of the coordinates of z, z', the rule

$$(x, y) \boxplus (x', y') = (xx' - yy', xy' + x'y)$$

endows the set $S^1(A)$ of solutions of the equation $X^2 + Y^2 = 1$ with coordinates in an *arbitrary* ring A with the structure of an abelian group. If $A \subset \mathbb{R}$ the group operation can be visualised as follows:



In particular, given any two rational solutions (x, y) and (x', y') of the equation $X^2 + Y^2 = 1$, we can obtain the rational solution $(x, y) \boxplus (x', y')$ (and find more solutions iterating the process).

Exercise 9.2.3. Show that the group $S^1(\mathbb{Q})$ is isomorphic to $\mathbb{Z}/4\mathbb{Z} \oplus \bigoplus_{p \equiv 1 \pmod 4} \mathbb{Z}$. What about the group $S^1(\mathbb{Q}(i))$?

Question 9.2.4. What about rational points on an ellipse with equation $X^2 + kY^2 = n$ with $k > 0$ squarefree and $n > 0$?

9.2.5. *Rational points on Mordell curves.* (References: [23, 24]) Let us now consider the Mordell equation $Y^2 = X^3 + k$ for $k \in \mathbb{Z}$. To fix ideas, let us first look at the case $k = -2$. We proved in 4.2 that the only *integral* solutions are $(3, \pm 5)$. What about rational solutions? Let us set $P_0 = (x_0, y_0) = (3, 5)$. Can we use this solution to produce a new solution $P_1 = (x_1, y_1)$? Here's an idea: the plane curve

$$E_{-2} : Y^2 = X^3 - 2 \tag{9.2.5.1}$$

is the vanishing locus of a polynomial of degree 3. Furthermore it is *smooth*: the derivative of $Y^2 - X^3 + 2$ with respect to X (resp. Y) is $-3X^2$ (resp. $2Y$), and every $(x, y) \in \mathbb{R}^2$ such that $y^2 = x^3 - 2$ satisfies $(x, y) \neq (0, 0)$. Therefore we can draw the *tangent line* to E_{-2} at the point P_0: it is the line with equation:

$$-3x_0^2(X - x_0) + 2y_0(Y - y_0) = 0 \Leftrightarrow X = 3 + t,\ Y = 5 + \frac{27}{10}t. \tag{9.2.5.2}$$

Let us compute the points of intersection between this line and the curve E_{-2}. We get:

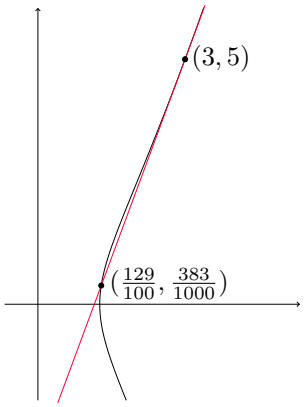
$$\left(5 + \frac{27}{10}t\right)^2 = (3 + t)^3 - 2 \Leftrightarrow t^3 + \frac{171}{100}t^2 = 0.$$

The solution $t = 0$ corresponds to the point P_0 we started with; it appears with multiplicity 2 because the line with equation (9.2.5.2) is tangent to E_{-2} at P_0. The third solution $t = \frac{-171}{100}$ yields

$$P_1 = (x_1, y_1) = \left(\frac{129}{100}, \frac{383}{1000}\right).$$

We have found a new solution!

There is no reason to stop here: we can repeat the process replacing P_0 with P_1. The tangent line to E_{-2} at P_1 has *rational slope* μ, as P_1 has rational coordinates; writing its equation in the form $X = x_1 + t$, $Y = y_1 + \mu t$ and intersecting the line with E_{-2} we will find an equation of degree 3 in t with *rational* coefficients. Since the line is tangent to E_{-2} at P_1 the solution $t = 0$ *must* appear with multiplicity at least 2; in other words the resulting equation will have the form $t^3 + pt^2$ for some $p \in \mathbb{Q}$. It follows that the remaining root $t = -p$ is rational, giving a further rational solution $P_2 = (x_2, y_2)$ of our original equation.



Remark 9.2.6. (1) In the above discussion, there is nothing special with the value $k = -2$. All we are using is that we have an equation of degree 3 in two variables, and that the curve it defines is smooth, so that its tangent line at every point is defined. In particular a similar process could be carried out with any equation of the form $Y^2 = X^3 + k$ for $k \neq 0$ (provided that we have at least one rational solution); it is natural to wonder whether the process produces *infinitely many rational solutions*. This is indeed the case for $k = -2$; the general situation was studied by Mordell [24], who showed that for most k if the equation $Y^2 = X^3 + k$ has a solution (x, y) with $xy \neq 0$ then the process described above gives rise to infinitely many rational solutions. The problem of finding for which k a solution $(x, y) \in \mathbb{Q}^2$ with $xy \neq 0$ does exist is open.

(2) More generally, given two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on $E_k : Y^2 = X^3 + k$ with rational coordinates and such that $y_2 \neq -y_1$, the line passing through P_1 and P_2 is defined by an equation with rational coefficients; hence the third intersection point $P_1 \boxplus P_2 = (x_3, y_3)$ between this line and E_k still has rational coordinates. It turns out that a slight modification of the rule $\boxplus$ endows the set of rational solutions of the equation $Y^2 = X^3 + k$ (plus a "point at infinity") with a *group structure*. Furthermore, Mordell proved that there always exist finitely many rational points $P_1, \dots, P_n$ of E_k such that any other rational point can be obtained from $P_1, \dots, P_n$ performing a finite number of chord or tangent constructions. In other words, the group of rational points of E_k is a *finitely generated* abelian group. Notice that this was not the case for the circle (cf. Exercise 9.2.3).

Observe that to define the composition law $\boxplus$ one only needs to start with an equation in two variables of degree 3, such that the tangent to every point of the corresponding curve is defined. For example, one could play the same game as above with the rational solutions of equations of the form $nY^2 = X^3 - X$, which appeared in lecture 1 in connection with the congruent number problem.

The content of this section is just a glimpse of a rich area of Mathematics; the interested reader can learn more from [32], and think about the following question.

Question 9.2.7. (1) What about rational solutions of the equation $Y^2 = X^3$? Notice that this equation has degree 3, but it defines a curve with a *cusp* at $(0, 0)$, where the tangent line is not defined.

(2) Can one similarly construct new rational solutions from old ones in the case of equations of degree 2 or 3 in more variables? For example, what about rational solutions of the equations

$$X^2 + Y^2 + Z^2 = a$$
$$X^3 + Y^3 + Z^3 = a, a \in \mathbb{Z} \text{ (e. g. } a = 1)?$$

10. Lecture 10: Ideals in quadratic rings and binary quadratic forms

In this lecture we will look at further examples of quadratic forms, examine which primes they represent and discover some new phenomena.

10.1. **Representing primes by quadratic forms: examples.**

Definition 10.1.1. *Let $k \in \mathbb{Z}$ be an integer which is not a square. A ring of the form*

$$\mathbb{Z}[\sqrt{k}] = \left\{a + b\sqrt{k} \mid a, b \in \mathbb{Z}\right\}$$
$$\mathbb{Z}\left[\frac{1+\sqrt{k}}{2}\right] = \left\{a + b\frac{1+\sqrt{k}}{2} \mid a, b \in \mathbb{Z}\right\}$$
$$= \left\{\frac{a+b\sqrt{k}}{2} \mid a, b \in \mathbb{Z}, a \equiv b \pmod 2\right\} \text{ if } k \equiv 1 \pmod 4$$

is called a quadratic ring. *It is called* real quadratic *(resp.* imaginary quadratic*) if $k > 0$ (resp. $k < 0$).*

Remark 10.1.2. Notice that as k is not a square then $\sqrt{k} \notin \mathbb{Q}$, hence every element in a quadratic ring of the form $\mathbb{Z}[\sqrt{k}]$ can be uniquely written as $a+b\sqrt{k}$ for some $a, b \in \mathbb{Z}$. Similarly, every element in a quadratic ring of the form $\mathbb{Z}\left[\frac{1+\sqrt{k}}{2}\right]$ can be uniquely written as $a + b\frac{1+\sqrt{k}}{2}$. Furthermore the fraction field of a quadratic ring $A = \mathbb{Z}[\sqrt{k}]$ or $\mathbb{Z}\left[\frac{1+\sqrt{k}}{2}\right]$ is the field $K = \mathbb{Q}(\sqrt{k}) = \{a + b\sqrt{k} \mid a, b \in \mathbb{Q}\}$.

Motivated by question 9.1.5 we wish to study ideals in quadratic rings. In lecture 8 we observed that, for a prime $p > 5$:

(p) splits as a product of two prime ideals in $\mathbb{Z}[i] \Leftrightarrow p \equiv 1 \pmod 4 \Leftrightarrow p = a^2 + b^2$

(p) splits as a product of two prime ideals in $\mathbb{Z}[\zeta_3] \Leftrightarrow p \equiv 1 \pmod 3 \Leftrightarrow p = a^2 - ab + b^2$

(p) splits as a product of two prime ideals in $\mathbb{Z}[i\sqrt{5}] \Leftrightarrow p \equiv 1, 3, 7, 9 \pmod{20} \Leftarrow p = a^2 + 5b^2$.

We also remarked that the last implication cannot be reversed: primes different from 5 and of the form $a^2 + 5b^2$ must be congruent to $1, 9$ modulo 20, and coincide with primes which are products of *principal* prime ideals in $\mathbb{Z}[i\sqrt{5}]$. What about primes congruent to $3, 7 \pmod{20}$?

Here's a little experiment:

3	6	$1^2 + 5 \cdot 1^2$
7	14	$3^2 + 5 \cdot 1^2$
23	46	$1^2 + 5 \cdot 3^2$
43	86	$9^2 + 5 \cdot 1^2$
47	94	$7^2 + 5 \cdot 3^2$
67	134	$3^2 + 5 \cdot 5^2$
83	166	$11^2 + 5 \cdot 3^2$

It seems that whenever $p \equiv 3, 7 \pmod{20}$ the number $2p$ is of the form $x^2 + 5y^2$. Let us reformulate this condition as follows: for $n > 0$ odd, if $2n = a^2 + 5b^2$ with $\gcd(a, b) = 1$ then letting $a = 2u + v$, $b = v$ we find $n = 2u^2 + 2uv + 3v^2$ and $\gcd(u, v) = 1$. Reversing the argument we obtain:

$$2n = a^2 + 5b^2,\ a, b \in \mathbb{Z}, \gcd(a, b) = 1 \Leftrightarrow n = 2u^2 + 2uv + 3v^2,\ u, v \in \mathbb{Z}, \gcd(u, v) = 1.$$

Using this observation we see that the above table suggests that, for a prime p coprime to 10:

$$p \equiv 1, 9 \pmod{20} \Leftrightarrow p = a^2 + 5b^2\ a, b \in \mathbb{Z} \tag{10.1.2.1}$$
$$p \equiv 3, 7 \pmod{20} \Leftrightarrow p = 2u^2 + 2uv + 3v^2 \Leftrightarrow 2p = a^2 + 5b^2\ a, b \in \mathbb{Z}.$$

Let us prove the implication from right to left in the bottom line: take a prime $p > 5$ and of the form $2u^2 + 2uv + 3v^2$; then either $u \not\equiv 0 \pmod p$ or $v \not\equiv 0 \pmod p$. In the first case notice that $p \mid (6v + 2u)^2 + 20u^2$; in the second case $p \mid (4u + 2v)^2 + 20v^2$. In both cases we obtain

$\left(\frac{-20}{p}\right) = 1$, i.e. $p \equiv 1, 3, 7, 9 \pmod{20}$. Furthermore $2p$ is of the form $a^2 + 5b^2$, hence congruent to $\pm 1 \pmod 5$. It follows that p must be congruent to $3, 7 \pmod{20}$.

In other words, if we can show that

$$p \equiv 1, 3, 7, 9 \pmod{20} \Leftrightarrow p = a^2 + 5b^2 \text{ or } p = 2a^2 + 2ab + 3b^2,\ a, b \in \mathbb{Z}$$

then the equivalences in (10.1.2.1) will follow (in fact, the implication $\Rightarrow$ is the only one we don't know yet).

One may wonder why we looked at $2p$ when $p \equiv 3, 7 \pmod{20}$, and investigate what happens for other multiples of p.

Exercise 10.1.3. Can you write the first few numbers of the form $3p, p \equiv 3, 7 \pmod{20}$ in the form $a^2 + 5b^2$? What about numbers of the form $7p$?

Let $f(X, Y) = X^2 + 5Y^2$ and $g(X, Y) = 2X^2 + 2XY + 3Y^2 \in \mathbb{Z}[X, Y]$. Since the norm map on $\mathbb{Z}[i\sqrt{5}]$ is multiplicative we know that

$$p = f(a_0, b_0), q = f(a_1, b_1) \Rightarrow pq = f(a_2, b_2), a_i, b_i \in \mathbb{Z}, i = 0, 1, 2.$$

However the above experiments reveal that the story for g seems to be different: it looks like

$$p = g(a_0, b_0), q = g(a_1, b_1) \Rightarrow pq = f(a_2, b_2),\ a_i, b_i \in \mathbb{Z}, i = 0, 1, 2.$$

In fact the above implication holds true, because of the following "composition formula" discovered by Lagrange:

$$(2a_0^2 + 2a_0b_0 + 3b_0^2)(2a_1^2 + 2a_1b_1 + 3b_1^2) = (2a_0a_1 + a_0b_1 + b_0a_1 + 3b_0b_1)^2 + 5(a_0b_1 - b_0a_1)^2.$$

We have learned:

(1) $p \equiv 1, 3, 7, 9 \pmod{20} \Leftrightarrow p$ splits as product of two prime ideals in $\mathbb{Z}[i\sqrt{5}]$.
(2) It seems that

p splits as product of two *non principal* ideals $\Leftrightarrow p = g(a, b) \Leftrightarrow p \equiv 3, 7 \pmod{20}$.

(3) After observing that, for $a, b \in \mathbb{Z}$,

$$N(2a + (1 + i\sqrt{5})b) = 2(2a^2 + 2ab + 3b^2) = 2g(a, b), \tag{10.1.3.1}$$

the following formulas seem somehow parallel

$$g(a_0, b_0)g(a_1, b_1) = f(a_2, b_2) \text{ vs } (2, 1 + i\sqrt{5})(2, 1 + i\sqrt{5}) = (2). \tag{10.1.3.2}$$

This suggests the existence of a relation between ideals in quadratic rings and quadratic forms in two variables with integer coefficients. We will now study this relation and explain the above phenomena.

10.2. **Quadratic forms.**

Definition 10.2.1. *An* integral binary quadratic form *is a homogeneous polynomial of degree two in two variables with integral coefficients*

$$f(X, Y) = aX^2 + bXY + cY^2 \in \mathbb{Z}[X, Y].$$

It is called primitive *if* $\gcd(a, b, c) = 1$. *The* discriminant *of* $f(X, Y)$ *is the integer* $\Delta(f) = b^2 - 4ac$. *We say that* f *is* definite *(resp.* indefinite*) if* $\Delta(f) < 0$ *(resp.* $\Delta(f) > 0$*).*

10.2.2. *Basic properties.* Recall that a quadratic form $f(X, Y) = aX^2 + bXY + cY^2 \in \mathbb{R}[X, Y]$ is said to be *positive definite* if for every $(x, y) \in \mathbb{R}^2 \smallsetminus \{0\}$ the inequality $f(x, y) > 0$ holds. In particular if f is positive definite then $a, c > 0$.

Now let $f(X, Y) = aX^2 + bXY + cY^2$ be an integral binary quadratic form.

(1) We have

$$4af(X, Y) = (2aX + bY)^2 - \Delta(f)Y^2$$
$$4cf(X, Y) = (2cY + bX)^2 - \Delta(f)X^2. \tag{10.2.2.1}$$

If $\Delta(f) < 0$ then, for $n \in \mathbb{Z}_{>0}$, the set of real solutions of the equation $f(X, Y) = n$ is either empty or an ellipse. For $(x, y) \in \mathbb{R}^2$ we have $f(x, y) = 0 \Leftrightarrow x = y = 0$; furthrmore exactly one of f and $-f$ is positive definite.

If $\Delta(f) > 0$ then, for $n \in \mathbb{Z}_{>0}$, the set of real solutions of the equation $f(X,Y) = n$ is an hyperbola. Both f and $-f$ take positive and negative values when evaluated at $(x,y) \in \mathbb{R}^2$.

(2) We have $\Delta(f) = b^2 - 4ac \equiv 0, 1 \pmod 4$. Conversely, for every integer $d \equiv 0, 1 \pmod 4$ there exists an integral binary quadratic form with discriminant d:

$$X^2 - \frac{d}{4}Y^2 \quad \text{if } d \equiv 0 \pmod 4$$
$$X^2 + XY + \frac{1-d}{4}Y^2 \quad \text{if } d \equiv 1 \pmod 4.$$

The form constructed above is called the *fundamental form* of discriminant d.

Remark 10.2.3. Notice that when d is not a square the fundamental form can be interpreted as follows. Say $d \equiv 1 \pmod 4$; then $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] = \left\{a + \frac{1+\sqrt{d}}{2}b, a, b \in \mathbb{Z}\right\}$. The fundamental form evaluated at the integers is related to the norm map

$$N : \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] \to \mathbb{Z}$$
$$\alpha = a + \frac{1+\sqrt{d}}{2}b \mapsto \alpha\alpha' = a^2 + ab + b^2\frac{1-d}{4}.$$

Similarly, if $d \equiv 0 \pmod 4$ then the fundamental form is related to the norm map on $\mathbb{Z}\left[\frac{1}{2}\sqrt{d}\right]$.

Definition 10.2.4. *Let $f(X,Y)$ be a binary integral quadratic form and $n \in \mathbb{Z}$. We say that f* represents *n if there exist $x, y \in \mathbb{Z}$ such that $f(x,y) = n$.*

In what follows we will restrict to the study of quadratic forms whose discriminant is *not* a square in $\mathbb{Q}$. In view of equation (10.2.2.1) this amounts to considering polynomials which are *irreducible* over $\mathbb{Q}$.

Lemma 10.2.5. *Let $d \equiv 0, 1 \pmod 4$ be an integer which is not a square and let p be a prime number. The following assertions are equivalent:*

(1) *There exists a form $f(X,Y)$ such that $\Delta(f) = d$ representing p.*
(2) *The congruence*

$$x^2 \equiv d \pmod{4p}$$

has a solution.

Proof. (1) $\Rightarrow$ (2)**:** Let $x, y \in \mathbb{Z}$ such that and $f(x,y) = p$; then we must have $\gcd(x,p) = 1$ or $\gcd(y,p) = 1$; without loss of generality we may assume that $\gcd(y,p) = 1$. Using (10.2.2.1) we obtain

$$4ap = (2ax + by)^2 - \Delta(f)y^2 \Rightarrow (2ax + by)^2 \equiv \Delta(f)y^2 \pmod p.$$

As $\gcd(y,p) = 1$ it follows that $\Delta(f)$ is a square modulo p. On the other hand $\Delta(f) = b^2 - 4ac$ is always a square modulo 4, hence we are done if p is odd by the Chinese remainder theorem. If $p = 2$ the above equation yields $(2ax + by)^2 \equiv \Delta(f)y^2 \pmod 8$; since $\gcd(y,2) = 1$ we are done.

(2) $\Rightarrow$ (1)**:** Take $b \in \mathbb{Z}$ such that $b^2 \equiv d \pmod{4p}$, and set $c = \frac{b^2-d}{4p}, a = p$. Then $f(X,Y) = aX^2 + bXY + cY^2$ satisfies

$$\Delta(f) = b^2 - 4p\frac{b^2-d}{4p} = d, \;\; f(1,0) = p.$$

□

Example 10.2.6. We can reformulate some of our previous results as follows. Let $p > 5$ be a prime number.

$f(X,Y) = X^2 + Y^2, \Delta(f) = -4, f$ represents $p \Leftrightarrow p \equiv 1 \pmod 4$

$f(X,Y) = X^2 - XY + Y^2, \Delta(f) = -3, f$ represents $p \Leftrightarrow p \equiv 1 \pmod 3$

$f(X,Y) = X^2 + 5Y^2, \Delta(f) = -20, f$ represents $p \overset{?}{\Longleftrightarrow} p \equiv 1, 9 \pmod{20}$

$f(X,Y) = 2X^2 + 2XY + 3Y^2, \Delta(f) = -20, f$ represents $p \overset{?}{\Longleftrightarrow} p \equiv 3, 7 \pmod{20}$.

11. Lecture 11: Abelian groups and fractional ideals

Our aim is now to generalise (10.1.3.1), constructing quadratic forms from ideals in quadratic rings. To do this we will need to know some basic properties of abelian groups, which we will prove first. It will also be convenient to extend slightly the notion of ideal introducing fractional ideals.

11.1. **Finitely generated abelian groups.**

Definition 11.1.1. *An abelian group (= $\mathbb{Z}$-module) M is called* free of rank n *if it is isomorphic to $\mathbb{Z}^n$. In other words, M is free of rank n if there exist $e_1, \dots, e_n \in M$ such that every $m \in M$ can be written uniquely in the form $m = k_1 e_1 + \dots + k_n e_n$, where $k_1, \dots, k_n \in \mathbb{Z}$. In this case $(e_1, \dots, e_n)$ is called a* basis *of M.*

For example, quadratic rings as defined in 10.1.1 are free abelian groups of rank 2. We will need to know that ideals of these rings have the same property. To show this we will use the following general property of abelian groups.

Proposition 11.1.2. *Let M be a free abelian group of rank n and let $M' \subseteq M$ be a subgroup. Then M' is a free abelian group of rank at most n.*

Proof. We prove the proposition by induction on n. The case $n = 1$ follows from the fact that $\mathbb{Z}$ is a principal ideal domain. Let $n \geq 2$ and $M' \subset M \simeq \mathbb{Z}^n$. If $M' \neq 0$ then there exists a projection map $p : M \simeq \mathbb{Z}^n \to \mathbb{Z}$ such that $p(M')$ is non-zero, so that $p(M') = a\mathbb{Z}$ for some $a \neq 0$. Furthermore $N = ker(p) \cap M'$ is contained in a free abelian group of rank strictly less than n, hence N is free of rank at most $n-1$ by induction. Let $b \in M'$ be an element such that $p(b) = a$. Then we claim that the map

$$N \oplus \mathbb{Z} \to M'$$
$$(r, s) \mapsto r + sb$$

is an isomorphism. Indeed it is injective as $N \cap \mathbb{Z}b = \{0\}$; to show surjectivity, take $m \in M'$, and observe that $p(m) = ka = kp(b)$ for some $k \in \mathbb{Z}$, hence $m - kb \in N$. □

Notation 11.1.3. Let M be an abelian group and $M' \subset M$ a subgroup. We will denote by $[M : M']$ the index of M' in M, i. e. the cardinality of the quotient abelian group M/M'. If A is a ring and $I \subset A$ is an ideal then $[A : I]$ denotes the index of $(I, +)$ in $(A, +)$ (which coincides with the cardinality of the quotient ring A/I).

11.1.4. *Finitely generated abelian groups.* Let M be a free abelian group of rank n, and $M' \subset M$ a non-zero subgroup. By the previous proposition we know that M' is free, of rank $0 < n' \leq n$. We may pick a basis $(e_1, \dots, e_n)$ of M, and a basis $(e'_1, \dots, e'_{n'})$ of M'; having made these choices, we can construct a matrix $\mathbf{A}$ with integral coefficients whose i-th column consists of the coordinates of e'_i with respect to the basis $(e_1, \dots, e_n)$. Multiplying $\mathbf{A}$ by an invertible matrix with $\mathbb{Z}$-coefficients on the left (resp. right) amounts to changing the chosen basis of M (resp. M'). The existence of the Smith normal form of a matrix with integral coefficients translates into the following properties of abelian groups.

(1) Let M be a free abelian group of rank n and $M' \subset M$ a subgroup of rank $0 < n' \leq n$. Then there exist a basis $e_1, \dots, e_n$ of M and positive integers $d_1 \mid d_2 \mid \dots \mid d_{n'}$ such that $d_1 e_1, \cdots, d_{n'} e_{n'}$ is a basis of M'.

Proof. Choose bases of M and M', construct the matrix $\mathbf{A}$ as above and reduce it to the Smith normal form. □

(2) Let M be a free abelian group of rank n and $M' \subset M$ a subgroup. Then the index $[M : M']$ of M' in M is finite if and only if M' has rank n. If this is the case, let $(e_1, \dots, e_n)$ be a basis of M and $(e'_1, \dots, e'_n)$ a basis of M'. Let $\mathbf{A}$ be the matrix whose i-th column consists of the coordinates of e'_i with respect to the basis $(e_1, \dots, e_n)$. Then $[M : M'] = |det(\mathbf{A})|$.

Proof. Choose a basis of M as in (1); then $M/M' \simeq \mathbb{Z}/d_1\mathbb{Z} \times \dots \times \mathbb{Z}/d_{n'}\mathbb{Z} \times \mathbb{Z}^{n-n'}$. Hence M' has finite index in M if and only if $n = n'$. In this case we have $[M : M'] = d_1 \cdot d_2 \cdots d_n = |\det(\mathbf{A})|$ where the last equality holds because multiplying by a matrix in $GL_n(\mathbb{Z})$ preserves the absolute value of the determinant. □

(3) Every *finitely generated abelian group* G is isomorphic to

$$\mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}$$

for some $r \in \mathbb{Z}_{\geq 0}$ and $d_1, \ldots, d_k \in \mathbb{Z}_{>0}$.

Proof. The fact that G is finitely generated means that there is a surjection $\mathbb{Z}^n \twoheadrightarrow G$ for some n. The kernel N is free of rank $k \leq n$; choosing a basis of $\mathbb{Z}^n$ as in (1) we find that $G \simeq \mathbb{Z}^n/N \simeq \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z} \times \mathbb{Z}^{n-k}$. □

Exercise 11.1.5. (1) Let $f : \mathbb{Z}^n \to \mathbb{Z}^n$ be a group homomorphism. Prove that f is surjective $\Leftrightarrow$ f is bijective $\Leftrightarrow$ $\det(f) \in \{\pm 1\}$.

(2) Let p be a prime.
   (a) Show that subgroups of $\mathbb{Z}^2$ of index p are in bijection with one dimensional subspaces of the $\mathbb{F}_p$-vector space $(\mathbb{Z}/p\mathbb{Z})^2$. Prove that there are $p+1$ such subgroups.
   (b) Let $\Gamma_0(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid c \equiv 0 \pmod{p} \right\} \subset SL_2(\mathbb{Z})$. Show that $\Gamma_0(p) \subset SL_2(\mathbb{Z})$ is the set of matrices sending $\mathbb{Z} \oplus p\mathbb{Z}$ to itself. Prove that $\Gamma_0(p)$ has index $p+1$ in $SL_2(\mathbb{Z})$.

11.2. **Fractional ideals.** It will be convenient to generalise slightly the notion of ideal.

Definition 11.2.1. *Let A be an integral domain and K its fraction field. A* fractional ideal *is an additive subgroup I of K such that there exists $\alpha \in K^\times$ such that $\alpha I \subset A$ is a non-zero ideal.*

Example 11.2.2. (1) If A, K are as above, then for any $a \in K^\times$ the group aA is a fractional ideal. It is called the *principal fractional ideal* generated by a.

(2) A fractional ideal is an ideal if and only if it is contained in A - hence fractional ideals are non-zero ideals of A "rescaled by a non-zero constant".

(3) If I, J are fractional ideals of A then

$$IJ = \{i_1 j_1 + \ldots + i_n j_n, n \in \mathbb{Z}_{\geq 1}, i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J\} \subset K$$

is a fractional ideal.

Lemma 11.2.3. *Let A be a quadratic ring and K its fraction field.*

(1) *For every non-zero ideal $I \subset A$ the quotient A/I is finite.*
(2) *Let $I \subset K$ be a fractional ideal of A. Then I is of the form $\alpha\mathbb{Z} \oplus \beta\mathbb{Z}$ for some $\alpha, \beta \in K^\times$ which are linearly independent over $\mathbb{Q}$. In particular it is a free abelian group of rank 2.*
(3) *Let $I \subset K$ be a fractional ideal of A. Then there exists $k \in \mathbb{Z}$ such that $kI \subset A$ is a non-zero ideal. Furthermore the quantity*

$$\frac{[A : kI]}{k^2}$$

does not depend on the choice of k.

Proof. (1) Let $\alpha \in I$ be a non zero element. Then I contains $N(\alpha) = \alpha\alpha' \in \mathbb{Z} \smallsetminus \{0\}$. As $A/N(\alpha)A$ is finite, it follows that A/I is finite.

(2) Let I be fractional ideal of A. By definition there is $\gamma \in K^\times$ such that $J = \gamma I \subset A$ is a non-zero ideal. As A is a free abelian group of rank 2, by consequence (2) in 11.1.4 and (1) we know that J is a free abelian group of rank 2, hence the same is true for I. Finally, if $I = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$, then α, β are linearly independent over $\mathbb{Z}$, hence over $\mathbb{Q}$.

(3) Let $\alpha \in K^\times$ is such that $\alpha I \subset A$ is a non-zero ideal. Then $a\alpha I \subset A$ is a non-zero ideal for any non-zero integer a; choosing a such that $a\alpha \in A$ we have that $N(a\alpha)I \subset A$ is also a non-zero ideal, hence any non-zero integer multiple of $N(a\alpha)$ does the job. Furthermore, if $l \in \mathbb{Z} \smallsetminus \{0\}$ also satisfies $lI \subset A$ then

$$[A : klI] = [A : kI][kI : klI] = [A : kI][I : lI] = [A : kI]l^2$$

where the first equality follows from multiplicativity of indices in a tower, the second from the fact that multiplication by k induces an isomorphism of abelian groups $I/lI \simeq kI/klI$, and the last holds as I is a free abelian group of rank 2. Symmetrically, we have

$$[A : klI] = [A : lI][lI : klI] = [A : lI][I : kI] = [A : lI]k^2.$$

Comparing the two expressions we find

$$[A : lI]k^2 = [A : kI]l^2 \Rightarrow \frac{[A : lI]}{l^2} = \frac{[A : kI]}{k^2}.$$

□

11.2.4. *Norm of fractional ideal.* Let A be a quadratic ring, K its fraction field and I a fractional ideal of A. The *norm* of I is defined as the quotient

$$N(I) = \frac{[A : kI]}{k^2}$$

for any integer k such that $kI \subset A$ is a non-zero ideal. The previous lemma ensures that this is a well-posed definition. For example, if $I \subset A$ is a non-zero ideal, we can take $k = 1$ and we find $N(I) = [A : I]$. The norm function enjoys the following key properties:

(1) If $\alpha \in K^\times$ then $N(\alpha A) = |N(\alpha)|$.
(2) For every $\alpha \in K^\times$ and every fractional ideal I,

$$N(\alpha I) = |N(\alpha)|N(I).$$

Notice that (2) follows from (1). Indeed, assume (1). As (2) holds for $\alpha \in \mathbb{Z} \smallsetminus \{0\}$, it suffices to check it for $\alpha \in A \smallsetminus \{0\}$. Let I be a fractional ideal and $k \in \mathbb{Z}$ such that $kI \subset A$ is a non-zero ideal. Then $k(\alpha I) \subset A$ is also a non-zero ideal, and we have

$$[A : k\alpha I] = [A : \alpha A][\alpha A : k\alpha I] = [A : \alpha A][A : kI]$$

where the last equality holds because multiplication by α induces an isomorphism of abelian groups $A/kI \simeq \alpha A/k\alpha I$. Dividing the left and right hand side by k^2 and using (1) we obtain (2).

Let us now show (1). After multiplying by a suitable integer $k \in \mathbb{Z}$ we may assume that $\alpha \in A \smallsetminus \{0\}$, hence we have $N(\alpha A) = [A : \alpha A]$. We learn from 11.1.4 (2) that $[A : \alpha A]$ is the absolute value of the determinant of the $\mathbb{Z}$-linear map

$$\begin{aligned} m_\alpha : A &\to A \\ x &\mapsto \alpha x. \end{aligned}$$

Therefore we have to check that $|\det(m_\alpha)| = |N(\alpha)|$. In fact, the following stronger statement holds:

Lemma 11.2.5. *Let $K = \mathbb{Q}(\sqrt{k})$ be a quadratic field and $\alpha \in K$. Then*

$$\begin{aligned} \det(m_\alpha) &= N(\alpha) \\ \mathrm{tr}(m_\alpha) &= Tr(\alpha). \end{aligned} \tag{11.2.5.1}$$

Proof. This can be checked by a direct computation (exercise); let us give a more scientific argument which will be useful later. For $\alpha \in \mathbb{Q}$ the statement is easily verified; let us now consider $\alpha \in K \smallsetminus \mathbb{Q}$. Recall that the ring $\mathbb{Q}[X]$ is Euclidean, so the ideal $I_\alpha = \{P \in \mathbb{Q}[X] \mid P(\alpha) = 0\}$ is principal; hence I_α is generated by a *unique* monic polynomial Q. We claim that $Q = P_\alpha = (X - \alpha)(X - \alpha') = X^2 - Tr(\alpha)X + N(\alpha)$. Indeed, $P_\alpha(\alpha) = 0$ hence $P_\alpha \in I_\alpha$ and $Q \mid P_\alpha$; however P_α is irreducible and $Q \neq 1$, hence $Q = P_\alpha$.

On the other hand, by the Cayley-Hamilton theorem the endomorphism m_α is killed by its characteristic polynomial $P_{m_\alpha} = X^2 - Tr(m_\alpha)X + det(m_\alpha)$: in other words, $P_{m_\alpha}(m_\alpha) : A \to A$ is the zero map. Evaluating it at $1 \in A$ we find

$$P_{m_\alpha}(\alpha) = 0$$

hence $P_{m_\alpha} \in I_\alpha$. As $P_{m_\alpha} \in \mathbb{Q}[X]$ is monic of degree two, we deduce that $P_{m_\alpha} = P_\alpha$, proving the lemma. □

Exercise 11.2.6. (Matrix representation of quadratic fields) Let $K = \mathbb{Q}(\sqrt{k})$ be a quadratic field. Let us denote by $End_{\mathbb{Q}}(K)$ be the (non-commutative) ring of endomorphisms of K as a $\mathbb{Q}$-vector space - choosing a basis of K, for example $(1, \sqrt{k})$, we can identify $End_{\mathbb{Q}}(K)$ with $M_2(\mathbb{Q})$.

(1) Prove that the map

$$K \to End_{\mathbb{Q}}(K)$$
$$\alpha \mapsto m_\alpha$$

is an injective ring homomorphism.

(2) Identify $End_{\mathbb{Q}}(K)$ with $M_2(\mathbb{Q})$ as above. Then $K^\times$ acts on $\mathbb{C}^2$ via the composition of the maps $K^\times \to GL_2(\mathbb{Q}) \to GL_2(\mathbb{C})$ and the natural action of $GL_2(\mathbb{C})$ on $\mathbb{C}^2$. Let

$$\mathbb{P}^1(\mathbb{C}) = \{\text{1-dimensional vector subspaces of } \mathbb{C}^2\}$$

and let $\mathbb{P}^1(\mathbb{R}) \subset \mathbb{P}^1(\mathbb{C})$ be the subset consisting of one dimensional vector spaces generated by an element (a, b) with either $b = 0$ or $\frac{a}{b} \in \mathbb{R}$. Prove that if $k > 0$ (resp. $k < 0$) then there are two elements of $\mathbb{P}^1(\mathbb{R})$ (resp. two elements of $\mathbb{P}^1(\mathbb{C}) \smallsetminus \mathbb{P}^1(\mathbb{R})$) fixed by the action of $K^\times$.

12. Lecture 12: Problem session II

(1) If you have not seen it before, prove Proposition 7.3.7.
(2) Recall that Spec A denotes the set of prime ideals of a ring A; this is called the *spectrum* of the ring A. Determine the spectrum of the following rings:
   (a) $A = \mathbb{C}[X]$;
   (b) $A = \mathbb{R}[X]$;
   (c) $A = \mathbb{Z}/n\mathbb{Z}, n \geq 1$.

   Bonus: let $A = \mathcal{C}([0,1], \mathbb{R})$ be the ring of continuous functions from the closed interval $[0,1]$ to $\mathbb{R}$. Prove that every maximal ideal of A is of the form $m_a = \{f \in A \mid f(a) = 0\}$ for some $a \in [0,1]$.
(3) (a) Let $A = \mathbb{Z}[\sqrt{k}]$ where k is not a square and let $\mathfrak{p} \in \text{Spec } A$ be a prime ideal of A. Prove that $\mathfrak{p} \cap \mathbb{Z}$ is a prime ideal of $\mathbb{Z}$.
   (b) More generally, let $f : A \to B$ be a morphism of rings and $\mathfrak{p} \in \text{Spec } B$. Prove that $f^{-1}(\mathfrak{p}) \in \text{Spec } A$. What if we replace prime ideals by maximal ideals?
   (c) Let A be a ring, $I \subset A$ an ideal and $q : A \to A/I$ the projection map. Prove that the map sending $\mathfrak{p}$ to $q^{-1}(\mathfrak{p})$ induces a bijection

$$\text{Spec } A/I \xrightarrow{\sim} \{\mathfrak{q} \in \text{Spec } A \mid \mathfrak{q} \supset I\}.$$

(4) Let A be a ring.
   (a) Show that an element $a \in A$ is a unit if and only if there is no maximal ideal of A containing a (you may assume the fact that every non-zero ring contains a maximal ideal; you can prove this fact if A is noetherian, and in general if you accept the axiom of choice).
   (b) Take an element a belonging to the intersection of all the maximal ideals of A. Prove that for every $b \in A$ the element $1 + ab$ is a unit.
   (c) Use the previous point to show that there are infinitely many prime numbers; then compare your proof with Euclid's classical argument.
   (d) Prove the converse of (b).
(5) Let $A = \mathbb{Z}\left[\frac{1+i\sqrt{7}}{2}\right]$.
   (a) Let $I \subset A$ be the ideal generated by $11, \frac{-13+i\sqrt{7}}{2}$. Show that $(11, \frac{-13+i\sqrt{7}}{2})$ is a basis of I as an abelian group; compute the norm of I.
   (b) Show that I is principal generated by $-2 + i\sqrt{7}$.
(6) In this exercise we will study the equation $Y^2 = X^5 - 227$ and prove that the ring $A = \mathbb{Z}\left[\frac{1+i\sqrt{227}}{2}\right]$ does not satisfy $SP(5)$.
   (a) Let $x, y \in \mathbb{Z}$ be such that $y^2 + 227 = x^5$. Show that x is odd.
   (b) Prove that $(y + i\sqrt{227}, y - i\sqrt{227}) = A$.
   (c) Assume that A satisfies $SP(5)$. Show that there exist $a, b \in \mathbb{Z}$ such that $a \equiv b \pmod 2$ and

$$y + i\sqrt{227} = \left(\frac{a + bi\sqrt{227}}{2}\right)^5.$$

   (d) Deduce that, if A satisfies $SP(5)$, then there are no x, y as in (1). Finally produce an integral solution of the equation $Y^2 = X^5 - 227$, and conclude that A does not safisfy $SP(5)$.
(7) Let $d \equiv 0, 1 \pmod 4$ not a square and let P_d be the set of prime numbers p such that there exists $f \in Q(d)$ representing p. The aim of this exercise is to prove that the set P_d is infinite.

   We say that an integral binary quadratic form $f(X, Y)$ *properly represents* n if there exist two *coprime* integers (x, y) such that $f(x, y) = n$.
   (a) Let n be a positive integer. Prove that there exists a form $f \in Q(d)$ properly representing n if and only if the congruence

$$X^2 \equiv d \pmod{4n}$$

   has a solution (this is a generalisation of Proposition 10.2.5).

(b) Deduce that if n is properly represented by a form in $Q(d)$ then so is every positive divisor of n.
(c) Let $q(X,Y)$ be the principal form of discriminant d and $(p_1,\dots,p_k)$ a finite list of prime numbers (not necessarily distinct). Let $n = q(1, p_1 \cdot \dots \cdot p_k)$. Show that if k is large enough then $n \neq \pm 1$.
(d) Let p be a prime factor of n; deduce that p is properly represented by a form in $Q(d)$.
(e) Show that p is different from every $p_i, i = 1,\dots,k$. Conclude that P_d is infinite.

Bonus exercise: ideals and geometry. In this exercise we study the spectrum of the ring $A = \mathbb{C}[X,Y]$; we will see that it captures some geometric information about the complex plane. Notice that every element of A can be seen as an element of the Euclidean domain $B = \mathbb{C}(X)[Y]$, where $\mathbb{C}(X)$ is the fraction field of $\mathbb{C}[X]$. You can make use of the following algebraic properties of the ring A: it is a UFD, and irreducible elements are polynomials $P(X,Y)$ which are irreducible in $\mathbb{C}(X)[Y]$ and such that, seeing $P(X,Y)$ as a polynomial in the variable Y with coefficients in $\mathbb{C}[X]$, these coefficients have no common factor.

(1) Let $\mathfrak{p}$ be a prime ideal in $\mathbb{C}[X,Y]$. Show that either $\mathfrak{p} \cap \mathbb{C}[X] = (0)$ or $\mathfrak{p} \cap \mathbb{C}[X] = (X-a)\mathbb{C}[X]$ for some $a \in \mathbb{C}$.
(2) Show that the map $\mathrm{Spec}B \to \mathrm{Spec}A$ induced by the inclusion $A \subset B$ yields a bijection
$$\mathrm{Spec}B \to \{\mathfrak{p} \in \mathrm{Spec}A \mid \mathfrak{p} \cap \mathbb{C}[X] = (0)\}$$
whose inverse sends $\mathfrak{q}$ to $\mathfrak{q}B$.
(3) Take $\mathfrak{p} \in \mathrm{Spec}A$ such that $\mathfrak{p} \cap \mathbb{C}[X] = (0)$. Show that $\mathfrak{p} = (P)$ for some irreducible polynomial $P \in \mathbb{C}[X,Y]$.
(4) Let $a \in \mathbb{C}$; show that the map $\mathbb{C}[X,Y] \to \mathbb{C}[Y]$ sending X to a induces a bijection
$$\mathrm{Spec}\mathbb{C}[Y] \to \{\mathfrak{p} \in \mathrm{Spec}\mathbb{C}[X,Y] \mid \mathfrak{p} \supset (X-a)\}$$
(5) Deduce that prime ideals of $\mathbb{C}[X,Y]$ are of the following form (their geometric interpretation is given on the right):
$$(0) \rightsquigarrow \text{the generic point;}$$
$$(X-a, Y-b) \rightsquigarrow \text{the point } (a,b) \in \mathbb{C}^2;$$
$$(P(X,Y)) \rightsquigarrow \text{the curve with equation } P = 0, \text{ where } P(X,Y) \in \mathbb{C}[X,Y] \text{ is irreducible.}$$
(6) Let $P(X,Y) \in \mathbb{C}[X,Y]$ be irreducible, and $(a,b) \in \mathbb{C}^2$; what is the formulation in terms of ideals of the fact that (a,b) belongs to the curve $P = 0$?
(7) Follow the above strategy to describe the spectrum of $\mathbb{Z}[X]$. You can find some pictures in [25, Chapter II].

13. Lecture 13: The dictionary between ideals and quadratic forms

In this lecture we will describe the dictionary between ideals in quadratic rings and binary integral quadratic forms.

13.1. **The main actors.** Fix throughout this lecture an integer $d \equiv 0, 1 \pmod 4$ which is *not* the square of an integer. If $d > 0$ then $\sqrt{d}$ will denote (as usual) the positive square root of d. If $d < 0$ we will instead set $\sqrt{d} = i\sqrt{-d}$, where $\sqrt{-d} > 0$.

Motivated by Remark 10.2.3 we define the *quadratic ring of discriminant* d to be the ring

$$\mathcal{O}_d = \begin{cases} \mathbb{Z}[\frac{\sqrt{d}}{2}] & = \left\{a + b\frac{\sqrt{d}}{2}, a, b \in \mathbb{Z}\right\} \text{ if } d \equiv 0 \pmod 4 \\ \mathbb{Z}[\frac{1+\sqrt{d}}{2}] & = \left\{a + b\frac{1+\sqrt{d}}{2} a, b \in \mathbb{Z}\right\} \text{ if } d \equiv 1 \pmod 4. \end{cases}$$

We will denote by K the fraction field of $\mathcal{O}_d$; if $\alpha_1, \alpha_2 \in K$ are linearly independent over $\mathbb{Q}$, we will say that (α_1, α_2) is a *positive basis* of K if

$$(\alpha_1, \alpha_2) = (1, \sqrt{d}) \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}, \text{ with } \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} > 0.$$

We will use the following facts (the first two were proved in lecture 11):

(1) The additive group $(\mathcal{O}_d, +)$ is free of rank 2; similarly, every fractional ideal of $\mathcal{O}_d$ is a free abelian group of rank 2.
(2) The *norm* of a fractional ideal I was defined; it satisfies the property $N(\alpha I) = |N(\alpha)|N(I)$.
(3) If I is a fractional ideal of $\mathcal{O}_d$ and $\alpha \in I \smallsetminus \{0\}$ then $\frac{N(\alpha)}{N(I)} \in \mathbb{Z}$. Indeed, after multiplying by a suitable $k \in \mathbb{Z}$ we may assume $I \subset \mathcal{O}_d$. In this case $\mathcal{O}_d/\alpha\mathcal{O}_d$ surjects onto $\mathcal{O}_d/I$ hence $N(I) = [\mathcal{O}_d : I] \mid [\mathcal{O}_d : \alpha\mathcal{O}_d] = |N(\alpha)|$.

Notation 13.1.1. (1) We will denote by $I(\mathcal{O}_d)$ the set of fractional ideals of $\mathcal{O}_d$, by $P(\mathcal{O}_d)$ the subset of *principal* fractional ideals (i.e. those of the form $\alpha\mathcal{O}_d$ for some $\alpha \in K^\times$) and by $P^+(\mathcal{O}_d)$ the set of fractional ideals of the form $\alpha\mathcal{O}_d$ for some α such that $N(\alpha) > 0$. Notice that $P^+(\mathcal{O}_d)$ is a group, with composition law $(\alpha)(\beta) = (\alpha\beta)$.
(2) An *oriented fractional ideal* of $\mathcal{O}_d$ is a couple $(I, (\alpha_1, \alpha_2))$ where $I \in I(\mathcal{O}_d)$ and (α_1, α_2) is a basis of I as a $\mathbb{Z}$-module which is *positive* when seen as a basis of K. The set of oriented fractional ideals of $\mathcal{O}_d$ will be denoted by $I^+(\mathcal{O}_d)$.
(3) If $d > 0$, the set of integral binary quadratic forms of discriminant d will be denoted by $Q(d)$; for $d < 0$ we will denote by $Q(d)$ the set of *positive definite* integral binary quadratic forms of discriminant d.

13.1.2. *Group actions: general terminology.* Let X be a set and G a group. A *left action* of G on X is a map $G \times X \to X$ such that, denoting the image of (g, x) by $g \cdot x$, the following equalities hold:

(1) $e \cdot x = x$ for every $x \in X$, where e is the identity of G.
(2) $g \cdot (h \cdot x) = (gh) \cdot x$ for every $g, h \in G, x \in X$.

The orbit of an element $x \in X$ is the set $\{g \cdot x, g \in G\} \subset X$; we denote by $G\backslash X$ the set of orbits of elements of X.

The mirror notion is that of a *right action*, i. e. a map $X \leftarrow X \times G$ such that, denoting the image of (x, g) by $x \cdot g$, the following equalities hold:

(1) $x \cdot e = x$ for every $x \in X$, where e is the identity of G.
(2) $(x \cdot g) \cdot h = x \cdot (gh)$ for every $g, h \in G, x \in X$.

Orbits of elements are defined as for left actions, and the set of orbits is denoted by X/G.

Let X be a set with a left (resp. right) G-action and $f : X \to Y$ a surjective map of sets. If the fibres of the map f (i. e. the preimages $f^{-1}(y)$ of elements $y \in Y$) are precisely the orbits in X, then f induces a bijection $G\backslash X \xrightarrow{\sim} Y$ (resp. $X/G \xrightarrow{\sim} Y$).

The group $GL_2(\mathbb{Z})$ acts on the right on the set $Q(d)$ by linear change of variables: precisely, if $f(X, Y) = aX^2 + bXY + cY^2 \in Q(d)$ then we can write $f(X, Y)$ in matrix form as

$$f(X, Y) = (X, Y) \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix};$$

if $\mathbf{A} = \begin{pmatrix} m_1 & m_2 \\ n_1 & n_2 \end{pmatrix} \in GL_2(\mathbb{Z})$, then we define $f_{|\mathbf{A}}(X,Y)$ as

$$f_{|\mathbf{A}}(X,Y) = (X,Y)\mathbf{A}^t \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \mathbf{A} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Definition 13.1.3. *We say that two binary integral quadratic forms $f(X,Y), g(X,Y)$ are* properly equivalent *if there exists $\mathbf{A} \in SL_2(\mathbb{Z})$ such that $g(X,Y) = f_{|\mathbf{A}}(X,Y)$.*

13.2. **From oriented ideals to quadratic forms.**

Definition 13.2.1. *Let $(I,(\alpha_1,\alpha_2)) \in I^+(\mathcal{O}_d)$ be an oriented fractional ideal of $\mathcal{O}_d$. The* quadratic form attached to $(I,(\alpha_1,\alpha_2))$ *is the integral binary quadratic form*

$$f_{I,(\alpha_1,\alpha_2)}(X,Y) = \frac{N(\alpha_1 X + \alpha_2 Y)}{N(I)} = \frac{(\alpha_1 X + \alpha_2 Y)(\alpha_1' X + \alpha_2' Y)}{N(I)}$$

where $\alpha \mapsto \alpha'$ is the involution sending $\sqrt{d}$ to $-\sqrt{d}$.

Example 13.2.2. If $I = \mathcal{O}_d$ and $(\alpha_1,\alpha_2) = (1,\sqrt{d})$ or $(1,\frac{1+\sqrt{d}}{2})$ then we obtain the fundamental form defined in 10.2.2. If $d = -20$ then $\mathcal{O}_d = \mathbb{Z}[i\sqrt{5}]$; if $I = (2, 1+i\sqrt{5})$ and $(\alpha_1,\alpha_2) = (2, 1+i\sqrt{5})$ (check that this is indeed a basis of I) then we obtain the form $2X^2 + 2XY + 3Y^2$ (cf. (10.1.3.1)).

Exercise 13.2.3. (cf. Problem session)

(1) Using (3) at the beginning of the lecture verify that $f_{I,(\alpha_1,\alpha_2)}(X,Y)$ is indeed a polynomial with integer coefficients.
(2) Let $\beta \in K^\times$ such that $N(\beta) > 0$, and let $(I,(\alpha_1,\alpha_2)) \in I^+(\mathcal{O}_d)$. Using (11.2.5.1) prove that $(\beta I,(\beta\alpha_1,\beta\alpha_2)) \in I^+(\mathcal{O}_d)$, then check that
$$f_{\beta I,(\beta\alpha_1,\beta\alpha_2)} = f_{I,(\alpha_1,\alpha_2)}.$$
If $N(\beta) < 0$ and $(I,(\alpha_1,\alpha_2)) \in I^+(\mathcal{O}_d)$, show that $(\beta\alpha_1, -\beta\alpha_2)$ is a positive basis of βI and
$$f_{\beta I,(\beta\alpha_1,-\beta\alpha_2)}(X,Y) = -f_{I,(\alpha_1,\alpha_2)}(X,-Y).$$

By the previous exercise, the multiplicative group $K^\times_{N>0} = \{\beta \in K^\times \mid N(\beta) > 0\}$ acts on the left on $I^+(\mathcal{O}_d)$: the action of $\beta \in K^\times_{N>0}$ sends $(I,(\alpha_1,\alpha_2)) \in I^+(\mathcal{O}_d)$ to $(\beta I,(\beta\alpha_1,\beta\alpha_2)) \in I^+(\mathcal{O}_d)$. We also have a right action of $SL_2(\mathbb{Z})$ on $I^+(\mathcal{O}_d)$ changing the basis of fractional ideals.

$$K^\times_{N>0} \curvearrowright I^+(\mathcal{O}_d) \curvearrowleft SL_2(\mathbb{Z})$$

Theorem 13.2.4. (1) *The map*

$$I^+(\mathcal{O}_d) \to Q(d)$$
$$(I,(\alpha_1,\alpha_2)) \mapsto f_{I,(\alpha_1,\alpha_2)}$$

is surjective, and its fibres are the $K^\times_{N>0}-$orbits in $I^+(\mathcal{O}_d)$. Hence the above map induces a bijection

$$K^\times_{N>0}\backslash I^+(\mathcal{O}_d) \xrightarrow{\sim} Q(d).$$

(2) *The map*

$$I^+(\mathcal{O}_d) \to I(\mathcal{O}_d)$$
$$(I,(\alpha_1,\alpha_2)) \mapsto I$$

is surjective, and its fibres are the $SL_2(\mathbb{Z})$-orbits in $I^+(\mathcal{O}_d)$. Hence the map induces a bijection

$$I^+(\mathcal{O}_d)/SL_2(\mathbb{Z}) \xrightarrow{\sim} I(\mathcal{O}_d).$$

Before proving the theorem, let us notice the following easy but crucial consequence.

Corollary 13.2.5. *The map $I^+(\mathcal{O}_d) \to Q(d)$ defined in the previous theorem induces a bijection*

$$P^+(\mathcal{O}_d)\backslash I(\mathcal{O}_d) \xrightarrow{\sim} Q(d)/SL_2(\mathbb{Z}).$$

Proof. By the above theorem

$$P^+(\mathcal{O}_d)\backslash I(\mathcal{O}_d) \xrightarrow{\sim} K^\times_{N>0}\backslash(I^+(\mathcal{O}_d)/SL_2(\mathbb{Z})) = (K^\times_{N>0}\backslash I^+(\mathcal{O}_d))/SL_2(\mathbb{Z}) \xrightarrow{\sim} Q(d)/SL_2(\mathbb{Z}).$$

□

13.2.6. *Proof of Theorem 13.2.4.* (2) is clear: indeed, two elements in $I^+(\mathcal{O}_d)$ mapping to the same fractional ideal only differ by the basis. On the other hand (α_1, α_2) and (β_1, β_2) are two oriented bases of the same ideal if and only if there exists $\mathbf{A} \in SL_2(\mathbb{Z})$ such that $(\beta_1, \beta_2) = (\alpha_1, \alpha_2)\mathbf{A}$.

It remains to prove (1). First of all we have to check that the map defined in (1) is well defined, i.e. that the discriminant of $f_{I,(\alpha_1,\alpha_2)}$ is d. By point (2) of Exercise 13.2.3 we may replace $(I, (\alpha_1, \alpha_2))$ by $(\alpha_1^{-1}I, (1, \pm\alpha_1^{-1}\alpha_2))$, where the sign equals the sign of $N(\alpha_1)$. Hence we can assume that

$$I = \mathbb{Z} \oplus \mathbb{Z}\alpha \text{ with } (1, \alpha) \text{ positive basis}$$

Then, setting $f = f_{I,(1,\alpha)}$, we have

$$f = \frac{(X + \alpha Y)(X + \alpha' Y)}{N(I)} = \frac{X^2 + Tr(\alpha)XY + N(\alpha)Y^2}{N(I)}.$$

Write $\alpha = a + b\frac{\sqrt{d}}{2}, a, b \in \mathbb{Q}$ if $d \equiv 0 \pmod 4$, and $\alpha = a + b\frac{1+\sqrt{d}}{2}, a, b \in \mathbb{Q}$ if $d \equiv 1 \pmod 4$. We have $Tr(\alpha)^2 - 4N(\alpha) = (\alpha - \alpha')^2 = db^2$. On the other hand, choosing $k \in \mathbb{Z} \smallsetminus \{0\}$ such that $k\alpha \in \mathcal{O}_d$, we find

$$N(I) = \frac{[\mathcal{O}_d : kI]}{k^2} = \frac{[\mathcal{O}_d : (k, k\alpha)]}{k^2} = \frac{|k \cdot kb|}{k^2} = |b|.$$

Hence

$$\Delta(f) = \frac{Tr(\alpha)^2 - 4N(\alpha)}{N(I)^2} = \frac{db^2}{b^2} = d.$$

Secondly, suppose that $f_{I,(\alpha_1,\alpha_2)} = f_{J,(\beta_1,\beta_2)}$. Then we have

$$\frac{N(\alpha_1)X^2 + (\alpha_1\alpha_2' + \alpha_2\alpha_1')XY + N(\alpha_2)Y^2}{N(I)} = \frac{N(\beta_1)X^2 + (\beta_1\beta_2' + \beta_2\beta_1')XY + N(\beta_2)Y^2}{N(J)};$$

Evaluating at $(X, Y) = (1, 0)$ we find

$$\frac{N(\alpha_1)}{N(I)} = \frac{N(\beta_1)}{N(J)}. \tag{13.2.6.1}$$

It follows that, setting $\gamma = \frac{\alpha_2}{\alpha_1}$ and $\delta = \frac{\beta_2}{\beta_1}$:

$$X^2 + (\gamma' + \gamma)XY + N(\gamma)Y^2 = X^2 + (\delta' + \delta)XY + N(\delta)Y^2.$$

Setting $Y = -1$ we see that $\gamma, \delta \in K \smallsetminus \mathbb{Q}$ satisfy the same monic quadratic polynomial with rational coefficients; hence we have either $\gamma = \delta$ or $\gamma = \delta'$. We claim that the second possibility cannot occur: indeed, (α_1, α_2) and (β_1, β_2) are positive bases, and by (13.2.6.1) $N(\alpha_1)$ and $N(\beta_1)$ have the same sign. It follows that $(1, \frac{\alpha_2}{\alpha_1})$ and $(1, \frac{\beta_2}{\beta_1})$ are either both positive bases, or both non-positive bases. But $(1, \frac{\alpha_2}{\alpha_1})$ and $(1, \frac{\alpha_2'}{\alpha_1'})$ have opposite orientations (since the involution of K has determinant -1). Hence we must have

$$\frac{\alpha_2}{\alpha_1} = \frac{\beta_2}{\beta_1} \Leftrightarrow \frac{\alpha_2}{\beta_2} = \frac{\alpha_1}{\beta_1} \Rightarrow (\alpha_1, \alpha_2) = \lambda(\beta_1, \beta_2), \lambda \in K^\times.$$

Since (α_1, α_2) and (β_1, β_2) are positive bases we must have $N(\lambda) > 0$. Hence we find $(J, (\beta_1, \beta_2)) = (\lambda I, (\lambda\alpha_1, \lambda\alpha_2))$; we have proved that the fibres of the map in (1) are precisely the $K^\times_{N>0}$−orbits in $I^+(\mathcal{O}_d)$.

Lastly, we have to prove surjectivity of the map in (1). For this, recall that if $d < 0$ then by definition $Q(d)$ consist of positive definite forms. In particular, if $f(X, Y) = aX^2 + bXY + cY^2 \in Q(d)$ then $a > 0$. If $d > 0$ then $Q(d)$ contains forms $f(X, Y) = aX^2 + bXY + cY^2$ with $a < 0$. However in this case there exists $\beta \in K^\times$ with $N(\beta) < 0$; by Exercise 13.2.3 $f(X, Y)$ is in the image of the map in (1) if and only if $-f(X, -Y)$ is. Hence we may assume that $a > 0$, and the result follows from the next lemma.

Lemma 13.2.7. *Let $f(X, Y) = aX^2 + bXY + cY^2 \in Q(d)$ with $a > 0$. Then*

1. *The abelian group $I = a\mathbb{Z} \oplus \frac{b+\sqrt{d}}{2}\mathbb{Z} \subset K$ is an ideal of $\mathcal{O}_d$.*
2. *The basis $(\alpha_1, \alpha_2) = (a, \frac{b+\sqrt{d}}{2})$ of I is positive.*
3. *The quadratic form associated to $(I, (\alpha_1, \alpha_2))$ is f.*

Proof. Exercise (cf. Problem session). □

14. Lecture 14: Reduction theory and finiteness of $Cl(\mathcal{O}_d)$

In this lecture we will study reduction theory for positive definite integral binary quadratic forms, and deduce a crucial finiteness result for equivalence classes of quadratic forms of given discriminant.

14.1. **Reduction theory.** Fix an integer $d \equiv 0, 1 \pmod 4$ which is not a square. Recall that $Q(d)$ is the set of binary integral quadratic forms of discriminant d - positive definite if $d < 0$. We denote by $Q(d)^{prim}$ the subset of primitive forms. We will denote a quadratic form $f(X,Y) = aX^2 + bXY + cY^2$ by $[a,b,c]$. With this notation we have that $[a,b,c]$ is primitive if and only if $\gcd(a,b,c) = 1$. Furthermore, for every $[a,b,c] \in Q(d)$, there exists $\lambda \in \mathbb{Z}$ such that $\lambda^2 \mid d$ and $[a',b',c'] \in Q(\frac{d}{\lambda^2})^{prim}$ such that $[a,b,c] = [\lambda a', \lambda b', \lambda c']$.

The aim of *reduction theory* is to find a representative in a given proper equivalence class of quadratic forms which is as simple as possible. We will explain how this works in the simplest case of *definite* quadratic forms; in other words, we will assume that $d < 0$ in what follows.

Exercise 14.1.1. (Reduction theory over the reals, I) Let $d < 0$ be any real number and $Q_{\mathbb{R}}(d)$ the set of positive definite binary quadratic forms with discriminant d and real coefficients. Let $f_d = \left[1, 0, \frac{-d}{4}\right]$. Prove that for every $f \in Q_{\mathbb{R}}(d)$ there exists $\mathbf{A} \in SL_2(\mathbb{R})$ such that $f_{|\mathbf{A}} = f_d$. Letting $SO(f_d) = \left\{ \begin{pmatrix} p & \frac{d}{4}r \\ r & p \end{pmatrix} \mid p, r \in \mathbb{R},\ p^2 - \frac{d}{4}r^2 = 1 \right\}$, deduce that the (right) action of $SL_2(\mathbb{R})$ on $Q_{\mathbb{R}}(d)$ induces a bijection

$$SO(f_d)\backslash SL_2(\mathbb{R}) \xrightarrow{\sim} Q_{\mathbb{R}}(d).$$

Definition 14.1.2. *A positive definite binary quadratic form $f = [a,b,c]$ is* reduced *if $|b| \leq a \leq c$ and, if $a = c$ or $|b| = a$, then $b \geq 0$.*

Example 14.1.3. The fundamental form

$$[1, 0, \frac{-d}{4}] \text{ if } d \equiv 0 \pmod 4$$
$$[1, 1, \frac{1-d}{4}] \text{ if } d \equiv 1 \pmod 4$$

is reduced (recall that we are assuming $d < 0$).

The form $[2,2,3] \in Q(-20)^{prim}$ is reduced, whereas the form $[7,6,2] \in Q(-20)^{prim}$ is *not* reduced.

Theorem 14.1.4. *Let $d \equiv 0, 1 \pmod 4, d < 0$.*

1. *Every positive definite integral binary quadratic form of discriminant d is properly equivalent to a reduced form.*
2. *If $f, f' \in Q(d)$ are reduced and properly equivalent, then $f = f'$.*

Proof. We will prove (1); for the proof of (2) we refer the reader to [9, Chapter 1, Theorem 2.8].

Let $f = [a,b,c] \in Q(d)$ (in particular $a, c > 0$). Let us consider the following two elements of $SL_2(\mathbb{Z})$:

$$\mathbf{S} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ \mathbf{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix};$$

then

$$f_{|\mathbf{S}} = [c, -b, a],\ f_{|\mathbf{T}^{\pm 1}} = [a, b \pm 2a, a \pm b + c].$$

To transform f into a reduced form we iterate the following operations:

1. If $a > c$ then apply $\mathbf{S}$ so that $a \leq c$.
2. If $|b| > a$ apply a suitable power of $\mathbf{T}$, so that $|b| \leq a$.

Notice that the process must stop, because every time we apply (1) we are decreasing the value of $a > 0$, and when applying (2) we are leaving a unchanged. The outcome will be a form properly equivalent to f such that $|b| \leq a \leq c$.

Finally, if $a = c$ then $[a,b,a]_{\mathbf{S}} = [a,-b,a]$ hence we can ensure that $b \geq 0$; if $a = -b$ then $[a,-a,c]_{\mathbf{T}} = [a,a,c]$. □

Exercise 14.1.5. Transform the form $[7,6,2]$ in reduced form.

Definition 14.1.6. *We denote*

$$Cl(\mathcal{O}_d) = Q(d)^{prim}/SL_2(\mathbb{Z}).$$

An integer $d \neq 1$ is called a fundamental discriminant *if either $d \equiv 1 \pmod 4$ is square-free, or $d = 4d'$ with $d' \equiv 2, 3 \pmod 4$ square-free.*

A discriminant d is fundamental if and only if $Q(d) = Q(d)^{prim}$. For a fundamental discriminant d, Corollary 13.2.5, plus the fact that $P^+(\mathcal{O}_d) = P(\mathcal{O}_d)$ as $d < 0$, yield a bijection

$$Cl(\mathcal{O}_d) = I(\mathcal{O}_d)/P(\mathcal{O}_d).$$

Remark 14.1.7. There is a slight change of notation with respect to Corollary 13.2.5, where we were working with $P^+(\mathcal{O}_d)\backslash I(\mathcal{O}_d)$. While it was important in the proof ot Theorem 13.2.4 to distinguish between left and right group actions, this is immaterial if we are only interested in the action of $P(\mathcal{O}_d)$ on $I(\mathcal{O}_d)$. Whatever choice one makes, the set of orbits is identified with the set of equivalence classes of fractional ideals of $\mathcal{O}_d$ modulo multiplication by an element of $K^\times$. In what follows we will often write this set as $I(\mathcal{O}_d)/P(\mathcal{O}_d)$.

Example 14.1.8. $d = -3$ is a fundamental discriminant, and $\mathcal{O}_{-3} = \mathbb{Z}[\zeta_3]$. The ring $\mathbb{Z}[i\sqrt{3}]$ is the quadratic ring of discriminant -12, which is not fundamental. For example, the quadratic form $f = [2, 2, 2]$ belongs to $Q(-12)$ but it is not primitive. Notice that f is the quadratic form associated to the oriented ideal $(I, (\alpha_1, \alpha_2))$ with $I = (2, 1 + i\sqrt{3}), \alpha_1 = 2, \alpha_2 = 1 + i\sqrt{3}$, which created trouble in 8.3.5. Using Corollary 13.2.5 we find

$$\{(1), I\} = P(\mathcal{O}_{-12})\backslash I(\mathcal{O}_{-12}) = Q(-12)/SL_2(\mathbb{Z}) = \{[1, 0, 3], f\} = Cl(\mathcal{O}_{-12}) \coprod \{f\}.$$

14.2. **Finiteness of $Cl(\mathcal{O}_d)$.** The fact that every form is properly equivalent to a reduced one has the following very important consequence:

Corollary 14.2.1. *Let $d \equiv 0, 1 \pmod 4$, $d < 0$. Then $Q(d)/SL_2(\mathbb{Z})$ is finite, hence $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ is finite and $Cl(\mathcal{O}_d)$ is finite.*

Proof. Take $f \in Q(d)$. We know from the previous theorem that f is properly equivalent to a form such that $|b| \leq a \leq c$. Hence

$$|d| = 4ac - b^2 \geq 4a^2 - a^2 = 3a^2 \Rightarrow 1 \leq a \leq \sqrt{\frac{-d}{3}}.$$

Hence there are finitely many possibilities for a; since $|b| \leq a$ there are finitely many choices for b as well. Finally, $ac = \frac{b^2-d}{4}$, hence c is determined by a, b.

This shows that $Q(d)/SL_2(\mathbb{Z})$ is finite, hence $Cl(\mathcal{O}_d)$ is finite; by Corollary 13.2.5, $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ is also finite. □

Remark 14.2.2. Corollary 14.2.1 is the first non-trivial *finiteness result* we encounter. As we will soon see, its importance could hardly be overestimated. Notice that the result not only tells us that there are finitely many proper equivalence classes of quadratic forms of a given discriminant (hence, finitely many fractional ideals of $\mathcal{O}_d$ up to multiplication by elements in $K^\times$) but also gives a bound for the coefficients of reduced forms in terms of the discriminant d, making it possible to list all of them.

Example 14.2.3.

(1) Let $d = -4$. Then a reduced form of discriminant d satisfies $1 \leq a \leq \sqrt{\frac{4}{3}}$ hence $a = 1$ and b is either 0 or ± 1. But $b^2 \equiv d \pmod 4$ hence $b = 0$ and $c = 1$. Hence $Cl(\mathbb{Z}[i]) = Q(-4)/SL_2(\mathbb{Z}) = \{[1, 0, 1]\}$. In particular we rediscover that $\mathbb{Z}[i]$ is a principal ideal domain. Furthermore, using Lemma 10.2.5 we find another proof of the first equivalence in Example 10.2.6.

(2) If $d = -3$ we find $a = 1$ hence $b = 1$ and $c = 1$, so $Cl(\mathbb{Z}[\zeta_3]) = Q(-3)/SL_2(\mathbb{Z}) = \{[1, 1, 1]\}$, and using Lemma 10.2.5 we deduce again the second equivalence in Example 10.2.6.

(3) Let $d = -20$, hence $\mathcal{O}_d = \mathbb{Z}[i\sqrt{5}]$. Then for a reduced form of discriminant d we must have $1 \leq a \leq 2\sqrt{\frac{5}{3}}$, hence either $a = 1$ or $a = 2$. In the first case we find $b = 0$ and $c = 5$, giving the form $f = X^2 + 5Y^2$. In the second case we have $b = 2$ and $c = 3$, hence we obtain $g = 2X^2 + 2XY + 3Y^2$. Notice that these two forms are not equivalent since they do not represent the same integers. Using Lemma 10.2.5 and the discussion in 10.1 we can now

prove the last two equivalences in Example 10.2.6. Finally, the calculation (10.1.3.1) and Corollary 13.2.5 show that

$$I(\mathbb{Z}[i\sqrt{5}])/(P(\mathbb{Z}[i\sqrt{5}])) \xrightarrow{\sim} Q(-20)/SL_2(\mathbb{Z})$$
$$(1, i\sqrt{5}) \mapsto [1, 0, 5]$$
$$(2, 1 + i\sqrt{5}) \mapsto [2, 2, 3].$$

Exercise 14.2.4. Find all reduced forms of discriminant $d = -26 \cdot 4 = -104$.

Corollary 14.2.5. *Let $d \equiv 0, 1 \pmod 4$, $d < 0$. Let $I \in I(\mathcal{O}_d)$; then there exist two integers $l > m > 0$ and an element $\alpha \in K^\times$ such that*

$$\alpha I^l = I^m.$$

Proof. Because of Corollary 14.2.1 we know that $Cl(\mathcal{O}_d) = I(\mathcal{O}_d)/P(\mathcal{O}_d)$ is finite. It follows that among the elements $I, I^2, I^3, \ldots \in I(\mathcal{O}_d)$ there are two which have the same image in $Cl(\mathcal{O}_d)$. Call these two elements I^l, I^m with $l > m$. Then the fact that I^l and I^m have the same image in $Cl(\mathcal{O}_d)$ means precisely that, for some $\alpha \in K^\times$, we have $\alpha I^l = I^m$. □

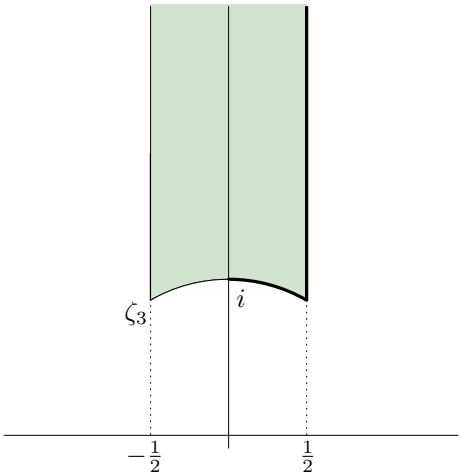
Remark 14.2.6. Reduction theory also works for quadratic forms with positive discriminant d (in fact, a version of the theory for quadratic forms in several variables also exists), and shows that the set $Q(d)/SL_2(\mathbb{Z}) \xrightarrow{\sim} I(\mathcal{O}_d)/P^+(\mathcal{O}_d)$ is finite. Notice that in this case $P^+(\mathcal{O}_d) \subset P(\mathcal{O}_d)$ and the inclusion is *strict* in general (convince yourself of this).

Remark 14.2.7. What about integral binary forms of degree higher than 2? One can still define the discriminant of such a form, and show that the set of $SL_2(\mathbb{Z})-$equivalence classes of forms of given discriminant is finite. See [3].

14.3. ***Quadratic forms, reduction theory and the Poincaré upper half plane.** To a positive definite quadratic form with real coefficients

$$f(X, Y) = aX^2 + bXY + cY^2 = a(X + \tau Y)(X + \bar{\tau} Y), \tau = \frac{b + i\sqrt{-d}}{2a}, \mathrm{Re}(\tau) = \frac{b}{2a}, |\tau| = \frac{c}{a}$$

we can associate the point $\tau \in \mathbf{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$. The fact that f is reduced is then equivalent to the fact that τ belongs to the following region:



This is an extremely important picture, at the intersection of several areas of Mathematics: quadratic forms, elliptic functions and modular forms, hyperbolic geometry.

Exercise 14.3.1. (Reduction theory over the reals, II) Let $\alpha : SO(f_d)\backslash SL_2(\mathbb{R}) \to Q_{\mathbb{R}}(d)$ the bijection constructed in Exercise 14.1.1 (whose notation we will use). In this exercise we will relate α to the above picture, and use this to generalise the first point of Theorem 14.1.4.

(1) Show that the map sending $[a, b, c] \in Q_{\mathbb{R}}(d)$ to $\tau = \frac{b+i\sqrt{-d}}{2a}$ induces a bijection
$$\beta : Q_{\mathbb{R}}(d) \to \mathbf{H}.$$

(2) Let $f = [a, b, c] \in Q_{\mathbb{R}}(d)$ and $\mathbf{A} = \begin{pmatrix} \sqrt{a} & \frac{b}{2\sqrt{a}} \\ 0 & \frac{1}{\sqrt{a}} \end{pmatrix}$. Prove that $f = f_{d|\mathbf{A}}$.

(3) Show that the map

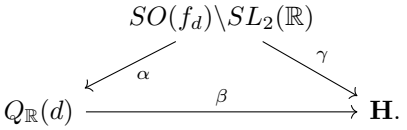
$$\mathbf{H} \times SL_2(\mathbb{R}) \to \mathbf{H}$$
$$(\tau, \begin{pmatrix} p & q \\ r & s \end{pmatrix}) \mapsto \frac{s\tau + q}{r\tau + p}$$

defines a right action[8] of $SL_2(\mathbb{R})$ on $\mathbf{H}$, and the map $SL_2(\mathbb{R}) \to \mathbf{H}$ sending $\mathbf{A}$ to $\frac{i\sqrt{-d}}{2} \cdot \mathbf{A}$ induces a bijection

$$\gamma : SO(f_d)\backslash SL_2(\mathbb{R}) \to \mathbf{H}.$$

Furthermore prove that the bijection β respects the $SL_2(\mathbb{R})$-action on the source and the target, i. e., for every $f \in Q_{\mathbb{R}}(d)$ and $\mathbf{A} \in SL_2(\mathbb{R})$, one has $\beta(f \cdot \mathbf{A}) = \beta(f) \cdot \mathbf{A}$ - a map with this property is called $SL_2(\mathbb{R})$-equivariant.

(4) Show that $\beta \circ \alpha = \gamma$, i. e. the following diagram commutes



(5) Consider the right action of $SL_2(\mathbb{Z}) \subset SL_2(\mathbb{R})$ on $\mathbf{H}$. Show that for every $\tau \in \mathbf{H}$ the orbit $\tau \cdot \mathrm{SL}_2(\mathbb{Z})$ contains an element in the green region in the above picture. (Hint: use carefully the matrices $\mathbf{S}$ and $\mathbf{T}$).

(6) Deduce that for every $f \in Q_{\mathbb{R}}(d)$ there is $\mathbf{A} \in SL_2(\mathbb{Z})$ such that $f_{|\mathbf{A}}$ is reduced.

[8]You probably have seen (or will see) a different formula giving a left action of $SL_2(\mathbb{R})$ on $\mathbf{H}$. The discrepancy is due to the fact that with our conventions $SL_2(\mathbb{R})$ acts on quadratic forms on the right. Making this into a left action as in [33, Section 1] one is led to work with the more common left action on $\mathbf{H}$.

15. Lecture 15: The group law on $Cl(\mathcal{O}_d)$

In this lecture we will show that, for a fundamental discriminant $d < 0$, there is a natural group structure on $Cl(\mathcal{O}_d)$.

15.1. **Rings with fundamental discriminant.** In this section we will investigate the consequences of Corollary 14.2.5 - which in turn follows from finiteness of $Cl(\mathcal{O}_d)$ - for the arithmetic of quadratic rings $\mathcal{O}_d$ when $d < 0$ is a *fundamental discriminant*. Let us point out that one can develop a more general theory, dealing with arbitrary discriminants; however rings associated to fundamental discriminants have particularly pleasant properties which do not hold true in general, as the example of $\mathbb{Z}[\zeta_3]$ compared to $\mathbb{Z}[i\sqrt{3}]$ already shows; see also Remark 15.2.7.

Recall that a discriminant d is fundamental if either $d \equiv 1 \pmod 4$ is squarefree or $d = 4d'$ with $d' \equiv 2, 3 \pmod 4$ squarefree. The associated ring $\mathcal{O}_d$ is $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ in the first case and $\mathbb{Z}[\frac{\sqrt{d}}{2}]$ in the second case. In other words, quadratic rings with fundamental discriminant are those of the form $\mathbb{Z}[\sqrt{k}]$ for $k \equiv 2, 3 \pmod 4$ *square-free* or $\mathbb{Z}[\frac{1+\sqrt{k}}{2}]$ for $k \equiv 1 \pmod 4$ *square-free*. The fundamental property or these rings, which distinguishes them from general quadratic rings, is the following:

Proposition 15.1.1. *Let d be a fundamental discriminant, $\mathcal{O}_d$ the corresponding quadratic ring and $K = \mathbb{Q}(\sqrt{d})$ its fraction field. Let $P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0 \in \mathbb{Z}[X]$ be a* monic *polynomial with* integer *coefficients. Let $\alpha \in K$ be a root of P. Then $\alpha \in \mathcal{O}_d$.*

Proof. If $\alpha \in \mathbb{Q}$ then the result follows from the fact that $\mathbb{Z}$ is a UFD, hence integrally closed (see Proposition 5.2.7). Let $\alpha \in K \smallsetminus \mathbb{Q}$; we know that α is a root of the polynomial $P_\alpha(X) = X^2 - Tr(\alpha)X + N(\alpha) \in \mathbb{Q}[X]$; furthermore, as we already observed in section 11.2.4, if we let $I_\alpha = \{Q \in \mathbb{Q}[X] \mid Q(\alpha) = 0\}$, then $I_\alpha = P_\alpha\mathbb{Q}[X]$. On the other hand by hypothesis $P \in I_\alpha$; it follows that $P_\alpha \mid P$. Gauss' lemma 15.1.2 implies that $P_\alpha \in \mathbb{Z}[X]$, hence $Tr(\alpha) \in \mathbb{Z}, N(\alpha) \in \mathbb{Z}$. Setting $k = d$ if $d \equiv 1 \pmod 4$ and $k = \frac{d}{4}$ if $d \equiv 0 \pmod 4$ we have that k is square-free, and $\alpha = a + b\sqrt{k}$ for some $a, b \in \mathbb{Q}$. Hence we must have

$$Tr(\alpha) = 2a \in \mathbb{Z},\ N(\alpha) = a^2 - kb^2 \in \mathbb{Z}, 4kb^2 = (2a)^2 - 4(a^2 - kb^2) \in \mathbb{Z} \Rightarrow 2b \in \mathbb{Z},$$

where the implication holds true because k is square-free. We now distinguish two cases:

(1) Let $k \equiv 2, 3 \pmod 4$; if $a = \frac{a'}{2}$ for some odd $a' \in \mathbb{Z}$ then, since $a^2 - kb^2 \in \mathbb{Z}$, we must have $b = \frac{b'}{2}$ for some odd $b' \in \mathbb{Z}$. We find
$$a^2 - kb^2 = \frac{(a')^2 - k(b')^2}{4};$$
however $(a')^2 \equiv (b')^2 \equiv 1 \pmod 4$, hence $(a')^2 - k(b')^2 \equiv -1, -2 \pmod 4$, which gives a contradiction. Therefore $a, b \in \mathbb{Z}$ and $\alpha \in \mathbb{Z}[\sqrt{k}] = \mathcal{O}_d$.

(2) If $k \equiv 1 \pmod 4$ then the same reasoning as above shows that either $a, b \in \mathbb{Z}$ or $2a, 2b$ are odd integers. In other words we have $\alpha = \frac{a+b\sqrt{k}}{2}$ with $a \equiv b \pmod 2$. As $\mathcal{O}_d = \left\{\frac{a+b\sqrt{k}}{2}, a \equiv b \pmod 2\right\}$ we find $\alpha \in \mathcal{O}_d$.

□

In the previous proof we made use of the following lemma due to Gauss; in its proof we will employ the following notation: if p is a prime and $\frac{a}{b} = \frac{p^r a'}{p^s b'} \in \mathbb{Q} \smallsetminus \{0\}$ with $\gcd(a', p) = \gcd(b', p) = 1$ then we set $v_p(\frac{a}{b}) = r - s$.

Lemma 15.1.2. *Let $P, Q \in \mathbb{Q}[X]$ be* monic *polynomials. Then*

$$PQ \in \mathbb{Z}[X] \Rightarrow P, Q \in \mathbb{Z}[X].$$

Proof. Assume that one of $P = X^n + a_{n-1}X^{n-1} + \ldots + a_0$ and $Q = X^m + b_{m-1}X^{m-1} + \ldots + b_0$ does not have integer coefficients: say $P \notin \mathbb{Z}[X]$. Let p be a prime such that $v_p(a_i) < 0$ for some $i = 1, \ldots, n-1$. There exists a unique index i_0 such that

$$v_p(a_{i_0}) < v_p(a_i)\ \forall i < i_0$$
$$v_p(a_{i_0}) \leq v_p(a_i)\ \forall i \geq i_0$$

and for such an i_0 we must have $v_p(a_{i_0}) < 0$. Similarly, there is a unique index j_0 such that

$$v_p(b_{j_0}) < v_p(b_j)\ \forall j < j_0$$
$$v_p(b_{j_0}) \leq v_p(b_j)\ \forall j \geq j_0$$

and we have $v_p(b_{j_0}) \leq 0$ (since Q is monic). Write $PQ = X^{m+n} + c_{m+n-1}X^{m+n-1} + \ldots + c_0$; then

$$c_{i_0+j_0} = a_{i_0}b_{j_0} + \sum_{i<i_0} a_i b_{i_0+j_0-i} + \sum_{j<j_0} b_j a_{i_0+j_0-j}. \tag{15.1.2.1}$$

Now, if $i < i_0$ then we have $v_p(a_i) > v_p(a_{i_0})$; on the other hand, $v_p(b_{i_0+j_0-i}) \geq v_p(b_{j_0})$. It follows that $v_p(a_i b_{i_0+j_0-i}) > v_p(a_{i_0}b_{j_0})$. A symmetric argument shows that if $j < j_0$ we have $v_p(b_j a_{i_0+j_0-j}) > v_p(a_{i_0}b_{j_0})$. Finally, since $v_p(a_{i_0}) < 0$ and $v_p(b_{j_0}) \leq 0$, we obtain that $v_p(a_{i_0}b_{j_0}) < 0$.

In other words we have shown that the exponent of the maximum power of p dividing the denominator of $a_{i_0}b_{j_0}$ (written in lowest terms) is positive, and strictly larger than the exponent of the maximum power of p dividing the denominator of any other summand on the right hand side of (15.1.2.1). It follows that $c_{i_0+j_0}$ is not an integer, contradiction. □

Exercise 15.1.3. (1) Convince yourself that you understood the above proof by running it on an example: e.g. you may try $P = X^4 + \frac{1}{3}X^3 + \frac{2}{9}X^2 + \frac{4}{9}X + \frac{4}{3}$, $Q = X^3 + 3X^2 + 9X + 9$, $p = 3$.

(2) The above statement, which is precisely what we used in the proof of Proposition 15.1.1, is the original form of what is now known as Gauss' lemma, as given by Gauss in [11, Article 42]. The proof we gave also follows closely Gauss' argument. Nowadays there are several equivalent formulations of the lemma, usually different from the one given above. Look them up in the literature, and adapt the previous argument to prove them. You may try to use an appropriate reformulation of the lemma to show that $\mathbb{Z}[X]$ and $\mathbb{C}[X, Y]$ are UFD.

(3) The function $v_p : \mathbb{Q} \smallsetminus \{0\} \to \mathbb{Z}$ introduced above is called *p-adic valuation*. Prove that it enjoys the following properties, which are the only properties we used in the previous proof:
   (a) $v_p(ab) = v_p(a) + v_p(b)$
   (b) $v_p(a + b) \geq \min(v_p(a), v_p(b))$. Furthermore equality holds if $v_p(a) \neq v_p(b)$.
   The *p-adic norm* of a rational number $a \neq 0$ is defined as $|a|_p = p^{-v_p(a)}$. Translate (a), (b) into properties of the norm.

15.2. **Group structure on $Cl(\mathcal{O}_d)$.** We will now show that finiteness of $Cl(\mathcal{O}_d)$ implies that this set has a natural *group structure*.

Lemma 15.2.1. *Let $d < 0$ be a fundamental discriminant, $I \in I(\mathcal{O}_d)$ and $l > m > 0$ such that $\alpha I^l = I^m$ for some $\alpha \in K^\times$. Then*

$$\alpha I^{l-m} = \mathcal{O}_d;$$

therefore, for every $I \in I(\mathcal{O}_d)$ there exist a power of I which is principal.

Proof. Let $J = \alpha I^{l-m}$. Then $JI^m = \alpha I^{l-m} I^m = \alpha I^l = I^m$. Write $I^m = \mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2$. Then we have, for $\beta \in J$:

$$\beta \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}a_1 & a_2\\ b_1 & b_2\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix},\ a_1, a_2, b_1, b_2 \in \mathbb{Z} \Rightarrow \begin{pmatrix}\beta - a_1 & -a_2\\ -b_1 & \beta - b_2\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = 0$$
$$\Rightarrow \begin{pmatrix}\beta - b_2 & a_2\\ b_1 & \beta - a_1\end{pmatrix}\begin{pmatrix}\beta - a_1 & -a_2\\ -b_1 & \beta - b_2\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = 0$$
$$\Rightarrow \begin{pmatrix}(\beta - a_1)(\beta - b_2) - a_2 b_1 & 0\\ 0 & (\beta - a_1)(\beta - b_2) - a_2 b_1\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = 0.$$

Hence we find $\beta^2 - (a_1 + b_2)\beta + (a_1 b_2 - a_2 b_1) = 0$, and Proposition 15.1.1 tells us that $\beta \in \mathcal{O}_d$. Hence we have proved that $J \subset \mathcal{O}_d$; notice that to prove this inclusion we only used the fact that $JI^m \subset I^m$; however we also know by hypothesis that $I^m \subset JI^m$. Therefore

$$\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}\gamma_1 & \gamma_2\\ \gamma_3 & \gamma_4\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix},\ \text{ for some } \gamma_1, \gamma_2, \gamma_3, \gamma_4 \in J.$$

The same computation as above shows that

$$\det\begin{pmatrix}1-\gamma_1 & -\gamma_2\\ -\gamma_3 & 1-\gamma_4\end{pmatrix}=0\Rightarrow -\gamma_1\gamma_4+\gamma_2\gamma_3+\gamma_1+\gamma_4=1\Rightarrow 1\in J\Rightarrow \mathcal{O}_d\subset J.$$

We have proved the first assertion of the lemma; the last one follows in view of Corollary 14.2.5. □

Theorem 15.2.2. *Let $d<0$ be a fundamental discriminant. Then the composition law $I\cdot J=IJ$ endows $Cl(\mathcal{O}_d)=I(\mathcal{O}_d)/P(\mathcal{O}_d)$ with the structure of an abelian group, with identity element the class of principal fractional ideals.*

Proof. As $I(JL)=(IJ)L$ and $IJ=JI$ for every $I,J,L\in I(\mathcal{O}_d)$ the composition law is associative and commutative. Furthermore for $\alpha\in K^\times$ the fractional ideals αI and I have the same image in $Cl(\mathcal{O}_d)$, hence the class of principal fractional ideals is the identity element in $Cl(\mathcal{O}_d)$. Finally, we learn from the previous lemma that for every $I\in I(\mathcal{O}_d)$ there exists $k\in\mathbb{Z}_{>0}$ such that I^k is principal, hence $I^{k-1}I=1\in Cl(\mathcal{O}_d)$; in other words every element of $Cl(\mathcal{O}_d)$ has an inverse, and the proof is complete. □

Translating the theorem back to the language of quadratic forms we discover the following remarkable facts, explaining anad generalising (10.1.3.2):

Corollary 15.2.3. (1) *The set $Q(d)/SL_2(\mathbb{Z})$ of* proper *equivalence classes of quadratic forms of (fundamental) discriminant $d<0$ acquires a group structure via the bijection*

$$Q(d)/SL_2(\mathbb{Z})\xrightarrow{\sim} P(\mathcal{O}_d)\backslash I(\mathcal{O}_d)$$

constructed in Corollary 13.2.5.

(2) *The identity element for the group structure in* (1) *is the proper equivalence class of the fundamental form.*

(3) *Let $\mathbf{I}=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\in GL_2(\mathbb{Z})$. The inverse of a form $f=aX^2+bXY+cY^2$ is the form $f_{|\mathbf{I}}=aX^2-bXY+cY^2$.*

Proof. (1), (2) are a restatement of Theorem 15.2.2. To prove (3), in view of Lemma 13.2.7, it suffices to show that $(a,\frac{b+\sqrt{d}}{2})(a,\frac{-b+\sqrt{d}}{2})$ is principal. Using the identity $d=b^2-4ac$ we find

$$\left(a,\frac{b+\sqrt{d}}{2}\right)\left(a,\frac{-b+\sqrt{d}}{2}\right)=\left(a^2,a\frac{b+\sqrt{d}}{2},a\frac{-b+\sqrt{d}}{2},\frac{d-b^2}{4}\right)\supset(a^2,ab,ac).$$

Since d is fundamental every quadratic form in $Q(d)$ is primitive, hence $\gcd(a,b,c)=1$. It follows that $(a,\frac{b+\sqrt{d}}{2})(a,\frac{-b+\sqrt{d}}{2})=(a)$, and we are done. □

Definition 15.2.4. *The group $Cl(\mathcal{O}_d)$ is called the* class group *of $\mathcal{O}_d$; its cardinality is the* class number *of $\mathcal{O}_d$, denoted by $h(\mathcal{O}_d)$.*

Remark 15.2.5. (1) The attentive reader will have noticed that in the proof of (3) we actually showed that every element in $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ has an inverse without relying on finiteness of $Cl(\mathcal{O}_d)$: we constructed such an inverse only using Lemma 13.2.7 and Theorem 13.2.4. In particular you can adapt the argument to prove that $I(\mathcal{O}_d)/P^+(\mathcal{O}_d)$ is also a group for fundamental discriminants $d>0$. It is called the *narrow class group* of $\mathcal{O}_d$. The reason why we deduced Theorem 15.2.2 from finiteness of $Cl(\mathcal{O}_d)$ above is that this deduction only relies on the algebraic fact 15.1.1, hence this approach will work more generally, as we shall soon see.

(2) The fact that integral binary quadratic forms up to *proper* (=$\mathrm{SL}_2(\mathbb{Z})$) equivalence can be composed, and the resulting law gives rise to a group if the discriminant is fundamental, was discovered by Gauss [11] much earlier than ideals were even defined. Notice that it is *crucial* to work with $SL_2(\mathbb{Z})$-equivalence and *not* with $GL_2(\mathbb{Z})$-equivalence: with the insight given by Corollary 15.2.3 we see that, had we made the latter choice, we would have ended up with $(I(\mathcal{O}_d)/P(\mathcal{O}_d))/\sim$, where $\sim$ is the equivalence relation identifying every element of $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ with its inverse. One checks (exercise) that the quotient $(I(\mathcal{O}_d)/P(\mathcal{O}_d))/\sim$ inherits a group structure from $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ if and only if for every $I\in I(\mathcal{O}_d)$ one of I and I^2 is principal. This fails in general, e.g. it is false for $d=-56$ (see Exercise 15.2.8).

Of course it was much harder at Gauss' time to realise that proper equivalence was the right notion: in fact this was a crucial insight due to Gauss, which Legendre had missed.

(3) Gauss' composition of quadratic forms has a nice geometric interpretation in terms of 2×2 cubes called *Bhargava cubes* [36]. In fact, Bhargava used this interpretation to define a composition law for integral binary *cubic* forms.

Example 15.2.6. We have seen in Example 14.2.3 that $Cl(\mathbb{Z}[i\sqrt{5}])$ consists of two elements. As it is a finite abelian group, by 11.1.4 (3) we deduce that $Cl(\mathbb{Z}[i\sqrt{5}]) \simeq \mathbb{Z}/2\mathbb{Z}$.

Similarly, using Exercise 14.2.4 one finds that $Cl(\mathbb{Z}[i\sqrt{26}]) \simeq \mathbb{Z}/6\mathbb{Z}$.

Remark 15.2.7. If $d < 0$ is a discriminant which is *not* fundamental, then there exist quadratic forms in $Q(d)$ which are *not* primitive. The counterpart of this on the ideal-theoretic side is that there exist fractional ideals $I \in I(\mathcal{O}_d)$ which are *not* invertible, i. e. such that there is no fractional ideal I^{-1} satisfying $II^{-1} = \mathcal{O}_d$. For example, let $d = -12$, so that $\mathcal{O}_d = \mathbb{Z}[i\sqrt{3}]$; let $K = \mathbb{Q}(\zeta_3)$. For $I \in I(\mathcal{O}_d)$ define $E(I) = \{\alpha \in K \mid \alpha I \subset I\}$. If I is invertible and $\alpha \in E(I)$ then $\alpha II^{-1} \subset II^{-1}$, hence $\alpha \in \mathcal{O}_d$. Now let us take $I = (2, 1 + i\sqrt{3}) \subset \mathcal{O}_d$. Then $\frac{1+i\sqrt{3}}{2} I = (1 + i\sqrt{3}, -1 + i\sqrt{3}) \subset \mathcal{O}_d$. Hence $\frac{1+i\sqrt{3}}{2} \in E(I) \smallsetminus \mathcal{O}_d$, so I is not invertible. Notice that the quadratic form attached to I with respect to the basis $(2, 1+i\sqrt{3})$ is $2X^2+2XY+2Y^2$, which is not primitive. In fact, the argument in the proof of Corollary 15.2.3 shows that fractional ideals corresponding to *primitive* quadratic forms are always invertible; one can show that the converse also holds, hence $Cl(\mathcal{O}_d) = Q(d)^{prim}/SL_2(\mathbb{Z})$ has the structure of a finite abelian group.

Exercise 15.2.8.

(1) Let $p \neq q$ be two primes congruent to 1 (mod 4). Then $\mathbb{Z}[i\sqrt{pq}] = \mathcal{O}_{-4pq}$ has fundamental discriminant. Show that X^2+pqY^2 and pX^2+qY^2 are two non-equivalent quadratic forms (Hint: find an integer represented by the first but not by the second form). Deduce that $Cl(\mathbb{Z}[i\sqrt{pq}])$ is non trivial. Can you generalise this?

(2) Let $d < 0$ be a fundamental discriminant, and $f = [a, b, c] \in Cl(\mathcal{O}_d)$. Show that f has order ≤ 2 in $Cl(\mathcal{O}_d)$ if and only if $b = 0$ or $a = |b|$ or $a = c$.

(3) Let $d = -56$; find all reduced forms of discriminant d and deduce that $Cl(\mathbb{Z}[i\sqrt{14}])$ has 4 elements. Show that the class of the quadratic form $[3, 2, 5]$ does *not* have order 2 in $Cl(\mathbb{Z}[i\sqrt{14}])$; deduce that $Cl(\mathbb{Z}[i\sqrt{14}]) \simeq \mathbb{Z}/4\mathbb{Z}$.

(4) Let $d < -8$ be a fundamental discriminant which is congruent to zero modulo 4. Prove that $h(\mathcal{O}_d)$ is even.

15.3. ***Genus theory.** Corollary 15.2.3 only scratches the surface of Gauss' achievements in the theory of quadratic forms: in [11], Gauss went much further developing what is known as *genus theory*. When translated in the language of ideals, this gives very detailed information on class groups of quadratic fields.

We only report the main results here for the curious reader; we refer to [9, Chapter 1] - and, of course, [11] - for more details.

Let $d < 0$ be a fundamental discriminant (as usual, a theory for arbitrary discriminants exists; Gauss worked in such generality). We say that two forms in $Q(d)$ belong to the same *genus* if they represent the same values in $(\mathbb{Z}/d\mathbb{Z})^\times$. Clearly any two properly equivalent forms belong to the same genus, hence every genus is partitioned into a finite number of classes of properly equivalent forms. In fact much more is true:

(1) Every genus contains the same number of classes.

(2) The number of genera of forms in $Q(d)$ is a power of 2.

These facts are already nice but are actually not so hard; the deepest result in the theory is the following

Theorem 15.3.1. *Let $d < 0$ be a fundamental discriminant.*

(1) *The number of genera is $2^{\mu-1}$, where μ is the number of prime factors dividing d.*

(2) *The genus containing the principal class consists precisely of* squares *of elements in $Cl(\mathcal{O}_d)$.*

Exercise 15.3.2. In order to appreciate the depth of Gauss' result, use it to prove the following statements:

(1) Let $d < 0$ be a fundamental discriminant. Then

$SL_2(\mathbb{Z}) -$ equivalence $= GL_2(\mathbb{Z}) -$ equivalence in $Q(d)$

$\Leftrightarrow Cl(\mathcal{O}_d) \simeq (\mathbb{Z}/2\mathbb{Z})^m$ for some $m \geq 0$

$\Leftrightarrow$ two forms in $Q(d)$ are equivalent if and only if they represent the same values in $(\mathbb{Z}/d\mathbb{Z})^\times$.

If d satisfies one of the above equivalent conditions it is called a *convenient number*. Exercise 15.2.8 gives a concrete way to check whether $Cl(\mathcal{O}_d) \simeq (\mathbb{Z}/2\mathbb{Z})^m$.

(2) Construct imaginary quadratic rings with arbitrarily high class number.

16. Lecture 16: Unique factorisation of ideals and the separating powers trick (again)

Building on the results of the last lecture, we will now prove that ideals in a quadratic ring $\mathcal{O}_d$ with negative fundamental discriminant factor uniquely as a product of prime ideals. We will then use this result to give a condition on $\mathcal{O}_d$ guaranteeing that the separating powers trick holds, and give applications to Mordell equations.

16.1. **Algebraic recollections and complements.** Recall that an ideal I in a ring A is prime if $I \neq A$ and, for every $a, b \in A$, $ab \in I \Rightarrow a \in I$ or $b \in I$. First of all, let us improve this slightly:

Lemma 16.1.1. *Let A be a ring, $I \subset A$ a prime ideal and $J_1, J_2 \subset A$ arbitrary ideals. Then*

$$J_1 J_2 \subset I \Rightarrow J_1 \subset I \text{ or } J_2 \subset I.$$

Proof. Suppose that $J_1 \not\subset I$ and $J_2 \not\subset I$. Take $j_1 \in J_1 \smallsetminus I$ and $j_2 \in J_2 \smallsetminus I$. As I is prime, we have $j_1 j_2 \notin I$. As $j_1 j_2 \in J_1 J_2$ we deduce that $J_1 J_2 \not\subset I$. □

Definition 16.1.2. *Let A be a ring and I, J ideals of A. We say that I* divides *J, and write $I \mid J$, if there exists an ideal $I' \subset A$ such that $II' = J$.*

Recall that an ideal I is maximal if $I \neq A$ and the only ideal properly containing I is A; we proved in Lemma 7.3.8 that every maximal ideal is prime. Furthermore, we showed in Lemma 11.2.3 that if A is a quadratic ring and $I \subset A$ a *non-zero* ideal then the quotient A/I is finite. This property has the following algebraic consequences:

Proposition 16.1.3. *Let A be a ring such that A/I is finite for every non-zero ideal $I \subset A$ (e. g. take A a quadratic ring). Then*

1. *A is noetherian.*
2. *Every ideal $I \subsetneq A$ is contained in a maximal ideal.*
3. *Every non-zero prime ideal of A is maximal.*

Proof. (1) We generalise the argument given in Remark 7.3.4: assume that $I \neq 0$ and take $\alpha \in I \smallsetminus \{0\}$. Then $A/\alpha A$ is finite by assumption, hence $I/\alpha A$ is also finite. Write $I/\alpha A = \{i_1, \dots, i_r\}$; then every $i \in I$ can be written as $i = i_k + \alpha a$ for some $1 \leq k \leq r$ and $a \in A$. If follows that $I = (\alpha, i_1, \dots, i_r)$, hence A is noetherian.

(2) Let $I \subsetneq A$ be a proper ideal; if I is not maximal, then there exists an ideal I_1 such that $I \subsetneq I_1 \subsetneq A$. If I_1 if maximal we are done. Otherwise we repeat the process; as A does not contain infinite ascending chains of distinct ideals, we must stop at some point, finding a maximal ideal $I_k \supset \dots \supset I$.

(3) Let I be a prime ideal of A. In view of Lemma 7.3.8 we know that A/I is an integral domain; if I is non-zero, then by assumption A/I is finite. Since a finite integral domain is a field (see the next exercise) using Lemma 7.3.8 again we find that I is maximal.

□

Exercise 16.1.4. Prove that a finite integral domain A is a field (Hint: every injective map $f : A \to A$ is surjective).

Remark 16.1.5. In fact, (2) is true for every ring; the general proof requires the axiom of choice.

16.2. **Unique factorisation of ideals.**

Theorem 16.2.1. *Let $d < 0$ be a fundamental discriminant, $A = \mathcal{O}_d$ and $I \subset A$ a non-zero ideal. Then:*

1. *There exists a fractional ideal I^{-1} such that $II^{-1} = A$.*
2. *If $J \subset A$ is an ideal, then*

$$I \mid J \Leftrightarrow I \supset J, \qquad \textit{memento: ``to contain is to divide.''}$$

3. *Let J, J' be ideals such that $JI = J'I$. Then $J = J'$ (cancellation property).*
4. *There exist non-zero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that*

$$I = \mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_r.$$

Furthermore, if $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ *are prime ideals such that*

$$\mathfrak{p}_1\mathfrak{p}_2\cdots\mathfrak{p}_r = I = \mathfrak{q}_1\mathfrak{q}_2\cdots\mathfrak{q}_s \tag{16.2.1.1}$$

then $r = s$ *and, up to reordering the* $\mathfrak{q}_i$*'s, we have* $\mathfrak{q}_i = \mathfrak{p}_i$ *for* $i = 1, \dots, r$.

Proof. (1) We know that $Cl(A)$ is a group, hence there exists a fractional ideal $I' \subset A$ such that $II' = \alpha A$, so $I(\alpha^{-1}I') = A$; hence $I^{-1} = \alpha^{-1}I'$ does the job.

(2) If $J = II'$ for some ideal $I' \subset A$ then $J \subset IA = I$. Conversely, if $J \subset I$ then $JI^{-1} \subset II^{-1} = A$. Hence $I' = JI^{-1}$ is an ideal of A such that $II' = J$.

(3) If $JI = J'I$ then multiplying by I^{-1} we find $J = J'$.

(4) Once the correct notions (ideal, prime ideal) have been set up, and the cancellation property has been established, the proof really is just a variation of the arguments in 5.2.6. Let us give the details; first of all we prove that I can be written as a product of non-zero prime (=maximal) ideals. If $I = A$ or I is maximal, we are done. Otherwise by Proposition 16.1.3 we know that there exists a maximal ideal $\mathfrak{p}_1 \supsetneq I$. By (2) we have $\mathfrak{p}_1 \mid I$, i.e. $I = \mathfrak{p}_1 I_1$ for some ideal $I_1 \subset A$. Notice that by the cancellation property we must have $I \subsetneq I_1$. If I_1 is maximal we are done; otherwise iterate the process. As A is Noetherian we must stop at some point, obtaining the desired factorisation.

To prove uniqueness, assume we are given two factorisations as in (16.2.1.1). Then $\mathfrak{p}_1 \supset \mathfrak{q}_1 \cdots \mathfrak{q}_s$ hence $\mathfrak{p}_1$ contains one of the $\mathfrak{q}_i$'s. We may assume after reordering that $\mathfrak{p}_1 \supset \mathfrak{q}_1$; since both ideals are maximal we must have $\mathfrak{p}_1 = \mathfrak{q}_1$. By (3) we may cancel $\mathfrak{p}_1$ on both sides of (16.2.1.1); repeating the argument we eventually kill all the factors on the left hand side of (16.2.1.1), hence we obtain $(1) = \mathfrak{q}_{r+1} \cdots \mathfrak{q}_s$; this implies that no prime ideal has actually survived on the right hand side.

□

16.3. **Back to the separating powers trick.** The previous theorem realises half of the dream in 7.2.1. In fact, it also allows to answer our old questions 4.3.6, 9.1.4, 9.1.5.

Theorem 16.3.1. *Let* $d < 0$ *be a fundamental discriminant and* $h(\mathcal{O}_d)$ *the class number of* $\mathcal{O}_d$. *Let* $n > 0$ *be an integer. If* $\gcd(n, h(\mathcal{O}_d)) = 1$ *then* $\mathcal{O}_d$ *satisfies property* $SP(n)$, *i.e:*

$$\forall\, \alpha, \beta \in \mathcal{O}_d \smallsetminus \{0\},\ \textit{if } (\alpha, \beta) = \mathcal{O}_d \textit{ and } \alpha\beta = \gamma^n \textit{ for some } \gamma \in \mathcal{O}_d$$
$$\textit{then there exist } \alpha_1, \beta_1 \in \mathcal{O}_d \textit{ and } u, v \in \mathcal{O}_d^\times \textit{ such that } \alpha = u\alpha_1^n, \beta = v\beta_1^n.$$

Proof. Take $\alpha, \beta, \gamma \in \mathcal{O}_d \smallsetminus \{0\}$ such that $(\alpha, \beta) = \mathcal{O}_d$ and assume that $\alpha\beta = \gamma^n$. Write

$$(\alpha) = \mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_r$$
$$(\beta) = \mathfrak{q}_1 \cdot \mathfrak{q}_2 \cdots \mathfrak{q}_s$$

where $\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_r, \mathfrak{q}_1, \dots, \mathfrak{q}_s$ are non zero prime ideals of $\mathcal{O}_d$. As $(\alpha, \beta) = \mathcal{O}_d$ the ideals $(\alpha), (\beta)$ have no common factor. On the other hand $(\alpha\beta) = (\gamma)^n$, hence every non-zero prime ideal $\mathfrak{p}$ appears $n \cdot r_\mathfrak{p}$ times in the factorisation of $(\alpha\beta)$, for some $r_\mathfrak{p} \geq 0$. Hence, if $r_\mathfrak{p} > 0$, then $\mathfrak{p}$ appears $n \cdot r_\mathfrak{p}$ times as a factor of exactly one of $(\alpha), (\beta)$. It follows that there are ideals $\mathfrak{a}, \mathfrak{b} \subset \mathcal{O}_d$ such that

$$(\alpha) = \mathfrak{a}^n,\ (\beta) = \mathfrak{b}^n.$$

In particular $\mathfrak{a}^n$ and $\mathfrak{b}^n$ are principal ideals. Therefore the ideals $\mathfrak{a}, \mathfrak{b}$, when regarded as elements of the abelian group $Cl(\mathcal{O}_d)$, have order dividing n. On the other hand $Cl(\mathcal{O}_d)$ is a finite abelian group of order $h(\mathcal{O}_d)$. As we are assuming that $\gcd(n, h(\mathcal{O}_d)) = 1$, we deduce that $\mathfrak{a}, \mathfrak{b}$ are themselves principal: $\mathfrak{a} = (\alpha_1), \mathfrak{b} = (\beta_1)$. Hence $(\alpha) = (\alpha_1)^n, (\beta) = (\beta_1)^n$, which implies the result. □

Remark 16.3.2. Notice that to formulate $SP(n)$ it is crucial to use the condition $(\alpha, \beta) = \mathcal{O}_d$ in place of the (weaker) condition $\delta \in \mathcal{O}_d, \delta \mid \alpha, \beta \Rightarrow \delta \in \mathcal{O}_d^\times$. As seen in the proof of the previous theorem, this is because what is lurking behind property $SP(n)$ is unique factorisation of *ideals*, and not of numbers. It could be instructive to go through the main steps which led us to Theorem 16.3.1 once more, starting from the example in 7.2, where all the story has begun.

Using Proposition 9.1.2 we deduce

Corollary 16.3.3. *Let* $k \in \mathbb{Z}_{>0}$ *be squarefree and congruent to 1 or 2 modulo 4. Assume that the ring* $\mathbb{Z}[i\sqrt{k}]$ *has class number coprime to* 3. *Then:*

(1) *if there exists an integer a such that $k = 3a^2 \pm 1$ then the only integer solutions of the equation $Y^2 + k = X^3$ are*

$$(x, y) = (a^2 + k, \pm a(a^2 - 3k)).$$

(2) *If there is no integer a such that $k = 3a^2 \pm 1$ then the equation $Y^2 + k = X^3$ has no integer solution.*

Example 16.3.4. (1) We know (e. g. see Example 15.2.6) that $h(\mathbb{Z}[i\sqrt{5}]) = 2$, hence $\mathbb{Z}[i\sqrt{5}]$ satisfies property $SP(3)$. This finally explains why the calculations in 4.3.1 worked. On the other hand, we saw in 4.3.3 that assuming that $SP(3)$ holds in $\mathbb{Z}[i\sqrt{26}]$ leads to a contradiction. It follows that $3 \mid h(\mathbb{Z}[i\sqrt{26}])$. In fact, you (should) have checked by other means in Exercise 14.2.4 that $h(\mathbb{Z}[i\sqrt{26}]) = 6$.

(2) Example 9.1.3 gives quadratic rings whose class number is divisible by 3.

(3) Let us determine $Cl(\mathbb{Z}[i\sqrt{61}])$ with minimal effort, using what we have learned.

(a) As $\mathbb{Z}[i\sqrt{61}] = \mathcal{O}_{-61\cdot 4}$, ideal classes are in bijection with reduced quadratic forms of discriminant -244. We know that every such form $[a, b, c]$ satisfies $|b| \leq a \leq c$ and $b > 0$ if any of the two inequalities is an equality.

(b) We compute $9 < \sqrt{\frac{244}{3}} < 10$ hence $1 \leq a \leq 9$.

(c) Switching to the language of ideals: the norm of the ideal $(a, \frac{b+i\sqrt{244}}{2})$ is a.

(d) We have $\mathbb{Z}[i\sqrt{61}] = \mathbb{Z}[X]/(X^2 + 61)$. If p is a prime and there is an ideal $\mathfrak{p}$ of norm p then $\mathfrak{p} \supsetneq (p)$, so (p) is not a prime ideal. It follows that $\mathbb{F}_p[X]/(X^2 + 61)$ is not a domain, hence if p is odd we find $\left(\frac{-61}{p}\right) = 1$.

(e) We know by Example 9.1.3 that $3 \mid h(\mathbb{Z}[i\sqrt{61}])$.

Let's go: for $a = 1$ we find the fundamental form. A form with $a = 2$ would give rise to an ideal of norm 2, hence dividing 2. As $\mathbb{F}_2[X]/(X^2 + 1) = \mathbb{F}_2[X]/(X + 1)^2$ we find that (2) is indeed the square of a prime ideal $\mathfrak{p}_2$, which is the unique ideal of norm 2 and gives an element of order 2 in $Cl(\mathbb{Z}[i\sqrt{61}])$.

For $a = 3$ we would obtain a form associated to an ideal of norm 3. But $\left(\frac{-61}{3}\right) = -1$, hence such a form cannot exist. Hence for $a = 9$ we can have only the form associated to the principal ideal (3), which gives nothing new. This also implies that there is no ideal of norm 6 (if there was one, it should be product of prime ideals dividing (2) and (3). But (3) itself is prime and has norm 9).

An ideal of norm 4 or 8 must divide (4) or (8). As $(2) = \mathfrak{p}_2^2$ we obtain nothing new.

It remains to check $a = 5$ and $a = 7$, i.e. we have to look for ideals of norm $5, 7$, hence dividing 5 and 7. We have $\left(\frac{-61}{5}\right) = \left(\frac{-61}{7}\right) = 1$, hence $X^2 + 61$ splits into linear factors over $\mathbb{F}_5$ and $\mathbb{F}_7$, giving 4 ideals. Hence $Cl(\mathbb{Z}[i\sqrt{61}])$ has order at most 6. On the other hand by (e) the group $Cl(\mathbb{Z}[i\sqrt{61}])$ contains an element of order 3; since it also has an element of order 2, we must have $Cl(\mathbb{Z}[i\sqrt{61}]) \simeq \mathbb{Z}/6\mathbb{Z}$.

Exercise 16.3.5. Let $\mathcal{O}_d$ be an imaginary quadratic ring with fundamental discriminant. Prove that every non-zero ideal $I \subset \mathcal{O}_d$ divides $(N(I))$.

17. Lecture 17: The Mordell equation $Y^2 = X^3 + 2$, the ring $\mathbb{Z}[\sqrt{2}]$ and Thue equations

In this lecture we study the equation $Y^2 = X^3 + 2$ via the "separating powers trick", as usual. In doing so we will discover that *units* in the ring $\mathbb{Z}[\sqrt{2}]$ have a more interesting structure than the imaginary quadratic counterpart $\mathbb{Z}[i\sqrt{2}]$.

17.1. **The equation $Y^2 = X^3 + 2$.** So far we have mainly studied the class group of imaginary quadratic rings, and we have used our results to solve Mordell equations of the form $Y^2 = X^3 + k$ with $k < 0$. What happens in the case $k > 0$? For example, let us look for integers $x, y \in \mathbb{Z}$ such that

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

We know that $\mathbb{Z}[\sqrt{2}]$ is a Euclidean domain, hence a unique factorisation domain. Furthermore reducing modulo 8 we see that x must be odd; on the other hand, letting $\delta = \gcd(y + \sqrt{2}, y - \sqrt{2})$, we find $N(\delta) \mid \gcd(8, x^6) = 1$, hence the separating powers trick tells us that there exists a unit $u \in \mathbb{Z}[\sqrt{2}]^\times$ and an element $\alpha \in \mathbb{Z}[\sqrt{2}]$ such that

$$y + \sqrt{2} = u\alpha^3. \tag{17.1.0.1}$$

17.1.1. *The group $\mathbb{Z}[\sqrt{2}]^\times$.* In order to proceed further, we need to study the group $\mathbb{Z}[\sqrt{2}]^\times$. In the imaginary quadratic case this was no big deal: the relevant group of units is finite and almost always equal to $\{\pm 1\}$. However for real quadratic rings the situation is very different! For example, the element

$$\varepsilon = 1 + \sqrt{2}$$

is a unit, and it is *not* a root of unity - the only roots of unity contained in $\mathbb{R}$ are ± 1. Hence for every $k \neq l \in \mathbb{Z}$ we must have $\varepsilon^k \neq \varepsilon^l$. In other words, ε generates a free abelian subgroup of rank one

$$\{\varepsilon^k, k \in \mathbb{Z}\} \subset \mathbb{Z}[\sqrt{2}]^\times.$$

We claim that $\mathbb{Z}[\sqrt{2}]^\times = \{\pm\varepsilon^k, k \in \mathbb{Z}\}$. To show this, let us first notice that $\varepsilon > 1$ and

$$\forall\, u \in \mathbb{Z}[\sqrt{2}]^\times, u > 1 \Rightarrow \varepsilon \leq u; \tag{17.1.1.1}$$

indeed, let $U_{[1,3]} = \{\alpha \in \mathbb{Z}[\sqrt{2}]^\times \mid 1 \leq |\alpha| \leq 3\}$. If $\alpha \in U_{[1,3]}$ then $\alpha\alpha' = \pm 1$ hence $|\alpha'| \leq 1$; it follows that the trace of an element in $U_{[1,3]}$ has absolute value at most 4. Hence such an element is a root of a polynomial of the form $X^2 + aX \pm 1$ with $|a| \leq 4$. Using this one finds that $U_{[1,3]} = \{\pm 1, \pm\varepsilon\}$, from which (17.1.1.1) follows.

Now let $u \in \mathbb{Z}[\sqrt{2}]^\times$. Up to changing the sign of u and replacing it with its inverse we may assume that $1 < u$. Let k be the largest nonnegative integer such that $\varepsilon^k < u \leq \varepsilon^{k+1}$. Then we find

$$1 < \varepsilon^{-k}u \leq \varepsilon \Rightarrow \varepsilon^{-k}u = \varepsilon \Rightarrow u = \varepsilon^{k+1}.$$

17.1.2. We can now come back to equation (17.1.0.1). As every element in $\mathbb{Z}[\sqrt{2}]^\times$ can be uniquely written as $\pm\varepsilon^k$ with $k \in \mathbb{Z}$ we deduce that $\mathbb{Z}[\sqrt{2}]^\times/(\mathbb{Z}[\sqrt{2}]^\times)^3 = \{1, \varepsilon, \varepsilon^{-1}\} = \{1, 1+\sqrt{2}, -1+\sqrt{2}\}$. Hence (17.1.0.1) tells us that there exist $a, b \in \mathbb{Z}$ such that one of the following three equations is satisfied:

$$y + \sqrt{2} = (a + b\sqrt{2})^3 = (a^3 + 6ab^2) + \sqrt{2}(3a^2b + 2b^3);$$
$$y + \sqrt{2} = (1 + \sqrt{2})(a + b\sqrt{2})^3 = (a^3 + 6a^2b + 6ab^2 + 4b^3) + \sqrt{2}(a^3 + 3a^2b + 6ab^2 + 2b^3);$$
$$y + \sqrt{2} = (-1 + \sqrt{2})(a + b\sqrt{2})^3 = (-a^3 + 6a^2b - 6ab^2 + 4b^3) + \sqrt{2}(a^3 - 3a^2b + 6ab^2 - 2b^3).$$

The first equation gives $b(3a^2 + 2b^2) = 1$, which has no integral solution. The second equation yields

$$a^3 + 3a^2b + 6ab^2 + 2b^3 = 1 \tag{17.1.2.1}$$

and the third equation gives $a^3 - 3a^2b + 6ab^2 - 2b^3 = 1$, which gives back the previous equation after changing sign to b.

How do we find the integral solutions of (17.1.2.1)? This would be easy if we could factor the polynomial $P(X, Y) = X^3 + 3X^2Y + 6XY^2 + 2Y^3$ as a product $P(X, Y) = L(X, Y)Q(X, Y)$ with $L(X, Y) \in \mathbb{Z}[X, Y]$ homogeneous of degree one and $Q(X, Y) \in \mathbb{Z}[X, Y]$ homogeneous of degree

two. Unfortunately this is not possible: indeed, assume that a such a factorisation exists, and write $L(X,Y) = rX + sY$. Then $r \neq 0$ hence, letting $t = \frac{-s}{r}$, we find

$$L(t,1) = 0 \Rightarrow P(t,1) = 0.$$

But the polynomial $X^3 + 3X^2 + 6X + 2$ has no rational root. Notice that equation (17.1.2.1) has the integral solution $(1,0)$, which yields the solution $(x,y) = (-1,1)$ of the equation $Y^2 - 2 = X^3$.

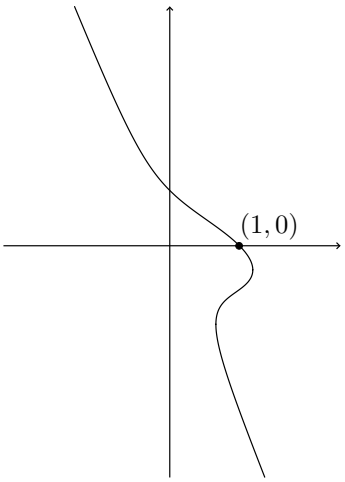


FIGURE 3. The curve $X^3 + 3X^2Y + 6XY^2 + 2Y^3 = 1$.

Similarly, the solution $(1,0)$ of the equation $a^3 - 3a^2b + 6ab^2 - 2b^3 = 1$ gives the solution $(x,y) = (-1,-1)$. However it is not a priori clear whether there are other integral solutions; we are led to the following question.

Question 17.1.3. Does the equation $Y^2 - 2 = X^3$ have finitely many integral solutions? If so, how can we find them? What about general Mordell equations?

Remark 17.1.4. It turns out that there is a trick deal with the equation $Y^2 - 2 = X^3$, working in a different ring from $\mathbb{Z}[\sqrt{2}]$, which allows to find all its integral solutions. The interested reader can look for such an argument; however, no trick will work in general, and we will need to do some hard work to answer the third question.

17.2. **Diophantine approximation and Thue equations.** So far we have discovered two new phenomena in this lecture:

(1) There are infinitely many units in the ring $\mathbb{Z}[\sqrt{2}]$. Consider the units $\varepsilon^k = x_k + \sqrt{2}y_k, k > 0$: then $x_k, y_k > 0$, hence

$$|x_k^2 - 2y_k^2| = 1 \Rightarrow \left|\left(\frac{x_k}{y_k}\right)^2 - 2\right| = \frac{1}{|y_k^2|} \Rightarrow \left|\frac{x_k}{y_k} - \sqrt{2}\right| \leq \frac{1}{\sqrt{2}|y_k^2|}.$$

In particular there are *infinitely many* rational numbers $\frac{x}{y}$ satisfying the inequality

$$\left|\frac{x}{y} - \sqrt{2}\right| < \frac{1}{|y|^2}.$$

(2) Finding the integral solutions of the Mordell equation $Y^2 - 2 = X^3$ boils down to finding the integers x, y such that $x^3 + 3x^2y + 6xy^2 + 2y^3 = 1$; if we discard the solution $(1,0)$ we can write this equation in the form

$$\left(\frac{x}{y}\right)^3 + 3\left(\frac{x}{y}\right)^2 + 6\left(\frac{x}{y}\right) + 2 = \frac{1}{y^3}.$$

The polynomial X^3+3X^2+6X+2 has one real irrational root α and two complex conjugate roots $\beta, \bar{\beta}$. Denoting by C the absolute value of the imaginary part of β, for every rational number $\frac{x}{y}$ we have $\left|\frac{x}{y}-\beta\right| = \left|\frac{x}{y}-\bar{\beta}\right| \geq C$, hence the above equation yields the inequality

$$\left|\frac{x}{y}-\alpha\right| \leq \frac{1}{C^2|y|^3}.$$

Therefore, *if* the answer to Question 17.1.3 is negative, then there are infinitely many rational approximations of α up to an error of at most $\frac{1}{C^2|y|^3}$.

Definition 17.2.1. *A* Thue equation *is a Diophantine equation of the form*

$$P(X,Y) = m$$

where $P(X,Y) \in \mathbb{Z}[X,Y]$ is homogeneous and $m \in \mathbb{Z} \smallsetminus \{0\}$.

Exercise 17.2.2.
(1) Let $P(X,Y) \in \mathbb{Z}[X,Y]$ be a homogeneous polynomial such that $P(X,1)$ is non-constant and has no real root. Prove that for every integer $m \neq 0$ the equation $P(X,Y) = m$ has finitely many integral solutions.
(2) Let K be a field.
   (a) Let $P(X,Y) \in K[X,Y]$ be a homogeneous polynomial. Show that every factor of P is homogeneous.
   (b) Let $P(X,Y) \in K[X,Y]$ be a homogeneous polynomial which is not divisible by Y. Show that P is irreducible if and only if $P(X,1) \in K[X]$ is irreducible.
   (c) Deduce that every homogeneous polynomial $P(X,Y) \in K[X,Y] \smallsetminus K$ factors as a product of irreducible homogeneous polynomials.

Notice that the study of general Thue equations reduces to that of equations such that $P(X,Y)$ is irreducible in $\mathbb{Q}[X,Y]$. Indeed if this is not the case then we can factor an integer multiple of P as a product of homogeneous polynomials with integer coefficients which are irreducible in $\mathbb{Q}[X,Y]$, and our original equation breaks up into a finite number of Thue equations involving irreducible polynomials. In 17.2 we saw that integral solutions (x,y) of certain Thue equations $P(X,Y) = \pm 1$, with $P(X,Y)$ irreducible and homogeneous of degree $d = 2, 3$, gave rise to approximations $\frac{x}{y}$ of a root of $P(X,1)$ with error bounded by $\frac{B}{|y|^d}$, where B is a constant only depending on P. Let us generalise this.

Proposition 17.2.3. *Let $P(X,Y) \in \mathbb{Z}[X,Y]$ be a homogeneous polynomial of degree $d > 1$ and $m \in \mathbb{Z} \smallsetminus \{0\}$. Assume that P is irreducible in $\mathbb{Q}[X,Y]$. There exists a real number $B > 0$ (depending on P and m) such that the following assertion holds: for every $(x,y) \in \mathbb{Z}^2$ satisfying $P(x,y) = m$ and $y \neq 0$ there exists a root α of $P(X,1)$ such that*

$$\left|\frac{x}{y}-\alpha\right| \leq \frac{B}{|y|^d}.$$

Proof. Let $\alpha_1, \ldots, \alpha_d \in \mathbb{C}$ be the roots of $P(X,1)$. The idea is to bound the distance between $\frac{x}{y}$ and any of the roots which is as close as possible to $\frac{x}{y}$. Let us give the details: we can write

$$P(X,Y) = a_0X^d + a_1X^{d-1}Y + \ldots + a_dY^d = a_0\prod_{i=1}^{d}(X - \alpha_i Y) \in \mathbb{C}[X,Y].$$

Furthermore $P(X,1) \in \mathbb{Q}[X]$ is irreducible hence its complex roots are distinct. Set $\eta = \min_{i\neq j}|\alpha_i - \alpha_j| > 0$ and $B = |m|\left(\frac{2}{\eta}\right)^{d-1}$.

Given x, y as in the statement of the proposition, let α be a root of $P(X,1)$ such that

$$|x - \alpha y| = \min_{1\leq i\leq d}|x - \alpha_i y|.$$

Letting $\mu = |x - \alpha y|$, we will show that $\mu \leq \frac{B}{|y|^{d-1}}$. Notice that $|a_0|\mu^d \leq |P(x,y)| = |m|$ hence $\mu^d \leq |m|$. We will distiguish two cases.

(1) If $|y| \leq \frac{2\mu}{\eta}$ then

$$\mu|y|^{d-1} \leq \mu^d(2/\eta)^{d-1} \leq |m|(2/\eta)^{d-1} = B \Rightarrow \mu \leq \frac{B}{|y|^{d-1}}.$$

(2) If $|y| > \frac{2\mu}{\eta}$ then, if $\alpha_i \neq \alpha$ we find

$$|x - \alpha_i y| = |x - \alpha y + (\alpha - \alpha_i)y| \geq |\alpha - \alpha_i||y| - \mu \geq \eta|y| - \mu \geq \eta\frac{|y|}{2}$$

$$\Rightarrow |m| = |a_0| \prod_{i=1}^{d} |x - \alpha_i y| \geq \mu \left(\eta \frac{|y|}{2}\right)^{d-1} \Rightarrow \mu \leq \frac{B}{|y|^{d-1}}.$$

□

The above proposition gives a bridge between Diophantine equations and Diophantine approximation: the problem of solving Thue equations is related to that of approximating certain irrational numbers. In turn, in all the examples of Mordell equations $Y^2 = X^3 + k$ we have considered so far the problem of finding their integral solutions has been reduced to that of solving a finite number of Thue equations of degree 3. It turns out that this is always the case.

Theorem 17.2.4. *(Mordell) Let k be a non-zero integer. There is a finite number of Thue equations*

$$P_i(X, Y) = m_i, i = 1, \ldots, r \tag{17.2.4.1}$$

of degree three such that every integral solution of the equation $Y^2 + k = X^3$ can be obtained (in an explicit way) from an integral solution of one of the equations (17.2.4.1).

Remark 17.2.5. The previous theorem was originally proved by Mordell, relating integral solutions of the equation $Y^2 + k = X^3$ to certain *binary integral cubic forms* of given discriminant. The desired finiteness then follows as a consequence of reduction theory for binary integral cubic forms. See [23, Chapter 24, 26] and [2]. A different proof, which reduces a more general class of Diophantine equations to Thue equations, was found by Landau and Ostrowski in 1919. Their argument is close in spirit to the methods we employed so far, relying on the arithmetic properties of quadratic rings. You will be able to build the proof by the end of this course.

18. Lecture 18: Problem session III

(1) Solve Exercise 13.2.3.
(2) Prove Lemma 13.2.7.
(3) Let $I = (3, 1 + i\sqrt{5}) \subset \mathbb{Z}[i\sqrt{5}]$. Compute the integral binary quadratic form associated to the oriented ideal $(I, (3, 1+i\sqrt{5}))$. Compare the result with (10.1.3.1): what does Theorem 13.2.4 tell you?

Let $I = (7, 3 + i\sqrt{5})$; compute the quadratic form attached to $(I, (7, 3 + i\sqrt{5}))$ and show that it is properly equivalent to the form attached to $(3, 1 + i\sqrt{5})$.
(4) Compute all the reduced forms of discriminant $-26 \cdot 4 = -104$; deduce that $Cl(\mathbb{Z}[i\sqrt{26}]) \simeq \mathbb{Z}/6\mathbb{Z}$. Explain how this is related to the equality $1 + 26 = 27 = 3^3$.
(5) Let $d < 0$ be an integer congruent to 0 or 1 modulo 4 and let $\mathcal{O}_d$ be the quadratic ring with discriminant d.
   (a) Prove that every $I \in I(\mathcal{O}_d)$ has a representative in $I(\mathcal{O}_d)/P(\mathcal{O}_d)$ which is of the form $(a, \frac{b+\sqrt{d}}{2})$ for some $a > 0$ and $b \in \mathbb{Z}$ such that $4a \mid b^2 - d$. Write $d = b^2 - 4ac$.
   (b) Show that a, b in the previous point can be chosen such that $|b| \leq a \leq \sqrt{\frac{|d|}{3}}$ and $a \leq (b^2 - d)/4a$.
   (c) Assume that d is fundamental and take a, b as in the previous point. Show that $(a, \frac{b+\sqrt{d}}{2})$ has order ≤ 2 in $Cl(\mathcal{O}_d)$ (i. e. either it is principal or its square is principal) if and only if $b = 0$ or $|b| = a$ or $a = c$.
   (d) Show that $Cl(\mathbb{Z}[i\sqrt{14}]) \simeq \mathbb{Z}/4\mathbb{Z}$.
(6) Prove that $Cl(\mathbb{Z}[i\sqrt{53}]) \simeq \mathbb{Z}/6\mathbb{Z}$. (Hint: you could start by just finding all reduced quadratic forms of discriminant -212; however it is better to use the dictionary ideals-quadratic forms during the process in order to avoid unnecessary computations; if you try to do this you may find some results in the next exercise useful).

What is the relation between the result you found and the equality $26^2 + 53 = 9^3$?
(7) Let $\mathcal{O}_d$ be an imaginary quadratic ring with fundamental discriminant.
   (a) Prove that every non-zero ideal $I \subset \mathcal{O}_d$ divides the ideal $(N(I))$.
   (b) Let p be a prime number, and assume that there is a form $f(X, Y) \in Q(d)$ representing p. Show that f is properly equivalent to $pX^2 + rXY + sY^2$ for some $r, s \in \mathbb{Z}$. Deduce that there is an ideal $I \subset \mathcal{O}_d$ of norm p, hence that (p) is not a prime ideal.
   (c) Conversely, assume that (p) is not a prime ideal. Show that there exists an ideal $I \subsetneq \mathcal{O}_d$ of norm p, and deduce that p is represented by a form in $Q(d)$.
(8) Let $f \in Q(d)$; let $Aut(f) = \{\mathbf{A} \in SL_2(\mathbb{Z}) \mid f_{|\mathbf{A}} = f\}$. If $d \equiv 1 \pmod 4$ is squarefree or $d = 4d'$ with $d' \equiv 2, 3 \pmod 4$ squarefree then show that $Aut(f)$ is isomorphic to $\mathcal{O}_d^{\times,+}$, where $\mathcal{O}_d^{\times,+} \subset \mathcal{O}_d^{\times}$ is the subset of elements with positive norm.

Bonus exercise: composition of quadratic forms and quadratic reciprocity. The aim of this exercise is to use the group law on the set of binary integral positive definite quadratic forms of negative fundamental discriminant to prove certain cases of quadratic reciprocity (the general case can be dealt with similarly, but it requires to work with quadratic forms with positive discriminant). If $d < 0$ is a fundamental discriminant and $f, g \in Q(d)/SL_2(\mathbb{Z})$, we will denote by $f \star g \in Q(d)/SL_2(\mathbb{Z})$ the composition of f, g.

(1) Let $d < 0$ be a fundamental discriminant, and $I, J \subset \mathcal{O}_d$ non-zero fractional ideals. Prove that $N(IJ) = N(I)N(J)$.
(2) Let $f, g \in Q(d)$ and $n, m \in \mathbb{Z}_{>0}$. Show that if f (resp. g) represents n (resp. m) then $f \star g$ represents nm.
(3) Let $p \equiv 3 \pmod 4$. Prove that the order of $Cl(\mathcal{O}_{-p})$ is odd.
(4) Let q be an odd prime such that $\left(\frac{-p}{q}\right) = 1$. Show that an odd power of q is represented by the fundamental form of discriminant $-p$.
(5) Deduce that $\left(\frac{q}{p}\right) = 1$.
(6) Conclude that, if $p \neq q$ are primes congruent to 3 modulo 4, then $\left(\frac{q}{p}\right) = -\left(\frac{p}{q}\right)$.

19. Lecture 19: Diophantine approximation: the theorems of Dirichlet and Liouville

The outcome of Lecture 17 is that we have transformed Question 17.1.3 into a problem in Diophantine approximation; we have also seen that non-trivial units in $\mathbb{Z}[\sqrt{2}]$ give rise to good approximations of $\sqrt{2}$. In this lecture we will study two classical results on Diophantine approximation, due to Dirichlet and Liouville, and we will use Dirichlet's theorem to determine the structure of units in real quadratic rings.

19.1. **Units in real quadratic rings.** We have seen that every unit in $\mathbb{Z}[\sqrt{2}]$ can be written in the form $\pm(1+\sqrt{2})^k$ for a unique $k \in \mathbb{Z}$. The first aim of this lecture is to prove that this phenomenon occurs for every *real* quadratic ring.

Theorem 19.1.1. *Let $d > 0$ be be an integer congruent to $0, 1$ modulo 4 and which is not a square. Then there exists a unit $\varepsilon \in \mathcal{O}_d^\times$ such that every unit in $\mathcal{O}_d^\times$ can be written uniquely as $\pm\varepsilon^k$ with $k \in \mathbb{Z}$. In other words there is an isomorphism of groups*

$$\mathcal{O}_d^\times \xrightarrow{\sim} \{\pm 1\} \times \mathbb{Z}$$
$$\varepsilon \mapsto (1, 1).$$

Let us start by giving an upper bound for the "size" of the group $\mathcal{O}_d^\times$.

Lemma 19.1.2. *Let $\mathcal{O}_d$ be a real quadratic ring. If there exists a unit $u \in \mathcal{O}_d^\times \smallsetminus \{\pm 1\}$ then there is a unit $\varepsilon \in \mathcal{O}_d^\times$ such that every unit in $\mathcal{O}_d^\times$ can be written uniquely as $\pm\varepsilon^k$ with $k \in \mathbb{Z}$.*

Proof. We use the same argument as in 17.1.1. Take u as in the statement of the lemma; up to changing sign to it and passing to the inverse we may assume that $u > 1$. Pick an integer $M > u$, and let $U_{[1,M]} = \{\alpha \in \mathcal{O}_d^\times \mid 1 \leq |\alpha| \leq M\}$. For every $\alpha \in U_{[1,M]}$ we have $|\alpha| \leq M$, $|\alpha\alpha'| = 1$ and $|\alpha'| \leq 1$. Hence the polynomial $P_\alpha(X) = (X-\alpha)(X-\alpha')$ has bounded integer coefficients, so there are finitely many possibilities for P_α. It follows that $U_{[1,M]}$ is finite; in particular the set $\{\alpha \in \mathcal{O}_d^\times \mid \alpha > 1\}$ has a minimum ε. Now let $v \in \mathbb{Z}[\sqrt{2}]^\times$; up to changing the sign of v and replacing it with its inverse we may assume that $1 < v$. Let k be the largest nonnegative integer such that $\varepsilon^k < v \leq \varepsilon^{k+1}$. Then

$$1 < \varepsilon^{-k} v \leq \varepsilon \Rightarrow \varepsilon^{-k} v = \varepsilon \Rightarrow v = \varepsilon^{k+1}.$$

□

19.2. **Approximation of quadratic irrational numbers and the Pell equation.** In view of the previous lemma, in order to complete the proof of Theorem 19.1.1 it suffices to show that $\mathcal{O}_d^\times$ contains one unit other than ± 1 - equivalently, we must show that it contains infinitely many units. As $\mathbb{Z}[\sqrt{d}] \subset \mathcal{O}_d$, it is enough to prove the following

Theorem 19.2.1. *Let $d > 0$ be an integer which is not a square. Then the equation $X^2 - dY^2 = 1$ has infinitely many integer solutions.*

The equation $X^2 - dY^2 = 1$ is called the *Pell equation*. We are looking for integers $x, y \in \mathbb{Z}$ such that $x \neq \pm 1$ and $x^2 - dy^2 = 1$; in particular $y \neq 0$, hence we can write our equation in the form $\left(\frac{x}{y}\right)^2 - d = \frac{1}{y^2}$. Restricting to x, y with the same sign - which we may always do - we have $\frac{x}{y} + \sqrt{d} > 1$, hence we find

$$\left|\frac{x}{y} - \sqrt{d}\right| < \frac{1}{y^2}.$$

In other words, we need to find a good rational approximation of $\sqrt{d}$. Notice that of course there are rational numbers as close as we wish to $\sqrt{d}$, but what we want is that the distance between $\frac{x}{y}$ and $\sqrt{d}$ is bounded in terms of (the square of) the denominator y.

Proposition 19.2.2. *(Dirichlet)*

1. *Let $\alpha \in \mathbb{R}$ and let $M > 0$ be an integer. There exist $p, q \in \mathbb{Z}$ such that $0 < q \leq M$ and $|q\alpha - p| < \frac{1}{M}$.*
2. *For every $\alpha \in \mathbb{R} \smallsetminus \mathbb{Q}$ there exist infinitely many $\frac{p}{q} \in \mathbb{Q}$ such that $|\alpha - \frac{p}{q}| < \frac{1}{q^2}$.*

Proof. The second assertion follows from the first (exercise). To prove the first claim, write

$$[0,1) = \coprod_{i=0}^{M-1} \left[\frac{i}{M}, \frac{i+1}{M}\right).$$

For $0 \leq j \leq M$ let $\lfloor j\alpha \rfloor$ be the integral part of $j\alpha$ (the largest integer less than or equal to $j\alpha$) and $\{j\alpha\} = j\alpha - \lfloor j\alpha \rfloor$ the fractional part of $j\alpha$. By the pigeonhole principle there exist $0 \leq j < j' \leq M$ and $0 \leq i < M$ such that $\{j'\alpha\}, \{j\alpha\} \in [\frac{i}{M}, \frac{i+1}{M})$. Hence $|(j'-j)\alpha - (\lfloor j'\alpha \rfloor - \lfloor j\alpha \rfloor)| = |\{j'\alpha\} - \{j\alpha\}| < \frac{1}{M}$. Therefore $q = j' - j$ and $p = \lfloor j'\alpha \rfloor - \lfloor j\alpha \rfloor$ satisfy the requirement of (1). □

Exercise 19.2.3.
(1) Show that the converse of (2) in the above proposition holds: if α is a rational number then there are finitely many $\frac{p}{q} \in \mathbb{Q}$ such that $|\alpha - \frac{p}{q}| < \frac{1}{q^2}$. Better, show that the same conclusion holds replacing the exponent 2 by any exponent $\delta > 1$.
(2) Prove the implication (1) $\Rightarrow$ (2) above - notice that, while (1) holds for any real number α, in view of the previous point you *have to* use the assumption that α is irrational in the proof of (1) $\Rightarrow$ (2).

19.2.4. *Proof of Theorem 19.2.1.* We know thanks to Dirichlet that there exists a sequence of distinct couples $(p_n, q_n) \in \mathbb{Z}^2$, $n \geq 1$, such that $\left|\frac{p_n}{q_n} - \sqrt{d}\right| < \frac{1}{q_n^2}$, hence $\left|\frac{p_n}{q_n} + \sqrt{d}\right| < \frac{1}{q_n^2} + 2\sqrt{d} \leq 2\sqrt{d} + 1$. It follows that

$$|p_n^2 - dq_n^2| < 2\sqrt{d} + 1.$$

Hence there is an integer m such that $|m| < 2\sqrt{d} + 1$ and a sequence of distinct numbers $\alpha_k = p_k + \sqrt{d}q_k, k \geq 1$ satisfying
(1) For every k, $N(\alpha_k) = m$.
(2) For every k, $p_k \equiv p_{k+1} \pmod{m}$ and $q_k \equiv q_{k+1} \pmod{m}$.

Hence $\alpha_k \equiv \alpha_{k+1} \pmod{m\mathbb{Z}[\sqrt{d}]}$ for $k \geq 1$, which implies:

$$\alpha_k'\alpha_1 \equiv \alpha_1'\alpha_1 \equiv N(\alpha_1) = m \equiv 0 \pmod{m\mathbb{Z}[\sqrt{d}]} \Rightarrow \exists\, \beta_k \in \mathbb{Z}[\sqrt{d}] : \alpha_k'\alpha_1 = m\beta_k.$$

Write $\beta_k = u_k + \sqrt{d}v_k$; then $N(\beta_k) = u_k^2 - dv_k^2$ and $N(\beta_k) = \frac{N(\alpha_k')N(\alpha_1)}{m^2} = \frac{m \cdot m}{m^2} = 1$, proving Theorem 19.2.1.

19.3. **Approximation of algebraic numbers: Liouville's theorem.**

19.3.1. *The irrationality exponent of quadratic irrationals.* We have seen that the existence of non-trivial units in a real quadratic ring $\mathcal{O}_d$ is intimately related to the fact that there are *infinitely many* rationals $\frac{p}{q}$ such that $|\frac{p}{q} - \sqrt{d}| < \frac{1}{q^2}$. Can we improve the exponent on the right hand side? Observe that, for every rational number $\frac{p}{q}$, we have

$$\left(\frac{p}{q}\right)^2 - d \neq 0 \Rightarrow |p^2 - dq^2| \geq 1.$$

On the other hand if $|\frac{p}{q} - \sqrt{d}| < 1$ then $|\frac{p}{q} + \sqrt{d}| < 1 + 2\sqrt{d}$. We deduce that

$$\left|\frac{p}{q} - \sqrt{d}\right| < 1 \Rightarrow \left|\frac{p}{q} - \sqrt{d}\right| \geq \frac{1}{(1+2\sqrt{d})q^2}.$$

We claim that this implies that, for every $\delta > 2$, there are *finitely many* rational numbers such that

$$\left|\frac{p}{q} - \sqrt{d}\right| < \frac{1}{|q|^\delta}. \tag{19.3.1.1}$$

In other words, the exponent in Dirichlet's theorem is the best possible for $\sqrt{d}$. Indeed, assume by contradition that there is a $\delta > 2$ such that (19.3.1.1) is satisfied by infinitely many rational numbers $x_1 = \frac{p_1}{q_1}, x_2 = \frac{p_2}{q_2}, \ldots$. In particular $|\frac{p_i}{q_i} - \sqrt{d}| < 1$, and we find

$$\frac{1}{(1+2\sqrt{d})q_i^2} \leq \left|\frac{p_i}{q_i} - \sqrt{d}\right| < \frac{1}{|q_i|^\delta}.$$

Observe that for a fixed q there are finitely many integers p such that (19.3.1.1) holds. Hence the absolute values of the denominators q_i must go to infinity. Since $\delta > 2$, we obtain a contradition.

19.3.2. *Algebraic numbers.* A close examination of the previous argument shows that the key input is the fact that $\sqrt{d}$ is a root of the polynomial with rational coefficients $X^2 - d$. In lecture 17 we were interested in approximations of the real root of the polynomial $X^3 + 3X^2 + 6X + 2$, while studying the Fermat equation with exponent a prime p one is led to consider the roots of the polynomial $1 + X + \ldots + X^{p-1}$. These are all examples of *algebraic numbers*, the central object of study of algebraic number theory.

Definition 19.3.3. *A complex number α is called an* algebraic number *if there exists a non-zero polynomial $P(X) = a_n X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in \mathbb{Q}[X]$ such that $P(\alpha) = 0$.*

Lemma 19.3.4. *Let α be an algebraic number. Then there exists a unique irreducible monic polynomial $P_\alpha^{\min}(X) \in \mathbb{Q}[X]$ such that $P_\alpha^{\min}(\alpha) = 0$. The ring $\mathbb{Q}[X]/(P_\alpha^{\min})$ is a field, and its image via the field morphism $\mathbb{Q}[X]/(P_\alpha^{\min}) \to \mathbb{C}$ sending X to α is the smallest subfield of $\mathbb{C}$ containing α.*

Proof. Consider the evaluation map

$$\begin{aligned} ev_\alpha : \mathbb{Q}[X] &\to \mathbb{C} \\ P &\to P(\alpha). \end{aligned}$$

Its kernel, denoted by I_α, is an ideal of $\mathbb{Q}[X]$, and it is non-zero by assumption. Let $P_\alpha^{\min}$ be the unique monic generator of I_α. As the image of ev_α is contained in $\mathbb{C}$ it is an integral domain, hence $P_\alpha^{\min}(X)$ is irreducible. Any other polynomial vanishing at α must be a multiple of $P_\alpha^{\min}(X)$, hence equal to it if it is required to be irreducible and monic. Finally, as $P_\alpha^{\min}(X)$ is irreducible the ring $\mathbb{Q}[X]/(P_\alpha^{\min})$ is a field; any subfield of $\mathbb{C}$ containing α must contain the image of ev_α, hence the last assertion of the lemma follows. □

Definition 19.3.5. *Let α be an algebraic number. The unique irreducible monic polynomial $P_\alpha^{\min}(X) \in \mathbb{Q}[X]$ vanishing at α is called the* minimal polynomial *of α. Its degree is called the* degree *of α. The smallest subfield of $\mathbb{C}$ containing α is denoted by $\mathbb{Q}(\alpha)$.*

Example 19.3.6. (1) Algebraic numbers of degree one are rational numbers.

(2) If $K = \mathbb{Q}(\sqrt{k})$ is a quadratic field then $\alpha = a + b\sqrt{k} \in K \smallsetminus \mathbb{Q}$ then α is an algebraic number of degree two, with minimal polynomial $X^2 - Tr(\alpha)X + N(\alpha)$.

(3) The number $\sqrt[3]{2}$ is an algebraic number of degree 3, with minimal polynomial $X^3 - 2$ (which is irreducible in $\mathbb{Q}[X]$ as it has no rational root).

(4) The numbers in the following list are all algebraic (check it), and they all belong to the field $\mathbb{Q}(\sqrt[4]{5})$:

$$1, 1 + \sqrt[4]{5}, 1 + \sqrt[4]{5} + \frac{\sqrt[4]{25}}{2}, 1 + \sqrt[4]{5} + \frac{\sqrt[4]{25}}{2} + \frac{\sqrt[4]{125}}{6}.$$

What about the infinite sum $\sum_{i \geq 0} \frac{\sqrt[4]{5^i}}{i!}$?

Theorem 19.3.7. *(Liouville, 1844) Let $\alpha \in \mathbb{R}$ be an algebraic number of degree d. For every $\delta > d$ there are finitely many $\frac{p}{q} \in \mathbb{Q}$ satisfying the inequality*

$$\left| \frac{p}{q} - \alpha \right| < \frac{1}{|q|^\delta}. \tag{19.3.7.1}$$

Proof. The case $d = 1$ was the subject of Exercise 19.2.3; let us restrict to the case $d > 1$. We will break the proof in five steps, which appear in several other arguments proving Diophantine approximation and transcendence results. Take $(p, q) \in \mathbb{Z}^2$ with $q \neq 0$.

Step 1: construction of a polynomial vanishing at α: we know that α is algebraic of degree d, hence we can pick a polynomial $P(X) \in \mathbb{Z}[X]$ of degree d such that $P(\alpha) = 0$.

Step 2: upper bound of $|P(\frac{p}{q})|$: let $C = \max\{|P'(\xi)| \mid \xi \in \mathbb{R}, |\xi - \alpha| \leq 1\}$. If $\frac{p}{q}$ is a rational number such that $\left|\frac{p}{q} - \alpha\right| \leq 1$ then by Step 1 and the mean value theorem there is $\xi \in \mathbb{R}$ such that $|\xi - \alpha| \leq 1$ and

$$\left| P\left(\frac{p}{q}\right) \right| = \left| P\left(\frac{p}{q}\right) - P(\alpha) \right| = \left| \frac{p}{q} - \alpha \right| |P'(\xi)| \Rightarrow \left| P\left(\frac{p}{q}\right) \right| \leq C \left| \frac{p}{q} - \alpha \right|.$$

Step 3: $P(\frac{p}{q})$ does not vanish: the number α is algebraic of degree $d > 1$ and it is a root of $P(X) \in \mathbb{Z}[X]$ of degree d. Hence $P(X)$ is irreducible in $\mathbb{Q}[X]$, and in particular it has no rational root, so $P(\frac{p}{q}) \neq 0$.

Step 4: $|P(\frac{p}{q})|$ cannot be too small: given the previous step, this is another manifestation of the key fact 1.2.1: we can write $P(\frac{p}{q}) = \frac{r}{q^d}$ for some integer r. Furthermore

$$P\left(\frac{p}{q}\right) \neq 0 \Rightarrow r \neq 0 \Rightarrow |r| \geq 1.$$

It follows that $P(\frac{p}{q}) \geq \frac{1}{|q|^d}$.

Step 5: comparing upper bound and lower bound: assume that there is an infinite sequence of distinct couples $(p_n, q_n) \in \mathbb{Z}^2, n \geq 0$, such that $q_n \neq 0$ and $\left|\frac{p_n}{q_n} - \alpha\right| < \frac{1}{|q_n|^\delta}$ for some $\delta > d$. In particular we have $\left|\frac{p_n}{q_n} - \alpha\right| \leq 1$, hence combining steps 2 and 4 we find

$$\frac{1}{|q_n|^d} \leq \left|P\left(\frac{p_n}{q_n}\right)\right| \leq C \cdot \left|\frac{p_n}{q_n} - \alpha\right| \leq \frac{C}{|q_n|^\delta}.$$

As for any given integer q the inequality (19.3.7.1) has finitely many solutions in p, the numbers $|q_n|$ must go to infinity as n grows. Since $d < \delta$ we obtain a contradiction.

□

19.3.8. *Transcendental numbers.* A complex number which is not algebraic is called *transcendental*. Nowadays one can easily argue that there exist transcendental numbers, and in fact "most" complex numbers are transcendental: indeed $\mathbb{C}$ is uncountable whereas the set of algebraic numbers is countable. Such an argument was not available at Liouville's time, and in any case it does not provide any concrete example of a transcendental number. Liouville was the first to construct such an example in 1851 [20], as an application of the previous theorem. For example, the following number is transcendental (exercise):

$$1 + \frac{1}{10} + \ldots + \frac{1}{10^{n!}} + \ldots.$$

Transcendental numbers are quite hard to handle; in particular it is often far from trivial to prove that a given number is transcendental. A landmark result in this area was proven by Lindemann in 1882; it asserts that

$$\alpha \in \mathbb{C}^\times \text{ algebraic } \Rightarrow e^\alpha \text{ transcendental.}$$

In particular, this says that the numbers e, π are transcendental.

19.3.9. *The irrationality exponent.* For a real number α, we define its *irrationality exponent* $e(\alpha)$ as

$$e(\alpha) = \sup\left\{\delta > 0 \mid \text{ there are infinitely many } \frac{p}{q} \in \mathbb{Q} \text{ such that } \left|\frac{p}{q} - \alpha\right| < \frac{1}{|q|^\delta}\right\}.$$

Then, for $\alpha \in \mathbb{R}$, we have:

(1) $\alpha \in \mathbb{Q} \Leftrightarrow e(\alpha) = 1$ (this follows from Proposition 19.2.2 and Exercise 19.2.3).
(2) If α is algebraic of degree $d \geq 2$ then $2 \leq e(\alpha) \leq d$ (this follows from Proposition 19.2.2 and Theorem 19.3.7).
(3) Liouville's theorem tells us that if $e(\alpha) = \infty$ then α is transcendental. Such a number is called a *Liouville number*; an example was given above. Liouville numbers are uncountable, and they form a dense subset of the real line; however they have Lebesgue measure zero.

A natural question (among many) at this point is the following:

Question 19.3.10. Let α be an algebraic number of degree $d \geq 3$. Can one improve Liouville's upper bound $e(\alpha) \leq d$? Equivalently, is there $\delta < d$ such that the inequality

$$\left|\frac{p}{q} - \alpha\right| < \frac{1}{|q|^\delta}$$

has finitely many solutions $\frac{p}{q} \in \mathbb{Q}$?

We conclude this lecture by observing that a positive aswer to the above question would have deep diophantine applications: for α as in Example 17.2(2) it would imply that the equation $Y^2 - 2 = X^3$ has finitely many integral solutions, answering part of question 17.1.3. Indeed if this was not the case we have seen in Example 17.2 that there would exist infinitely many rational numbers satisfying $\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{C^2|q|^3}$. But, as in the proof of Liouville's theorem, this would imply that $|q|$ gets arbitrarily large; hence for any $\delta < 3$ we would find infinitely many rational numbers such that $\left|\frac{p}{q} - \alpha\right| < \frac{1}{|q|^\delta}$. In the same way, using Proposition 17.2.3 and Theorem 17.2.4, one obtains the following result, whose proof is left as an exercise.

Lemma 19.3.11. *If for every algebraic number α of degree 3 the inequality $e(\alpha) < 3$ holds then every Mordell equation $Y^2 = X^3 + k$ has finitely many integral solutions.*

At the end of this course we will study Thue's groundbreaking achievement regarding question 19.3.10.

20. Lecture 20: Algebraic integers and cyclotomic fields

We now start our study of number fields and their rings of integers, generalising quadratic rings. Our main examples will be cyclotomic fields $\mathbb{Q}(\zeta_p)$; we will come back to quadratic rings as well as to Diophantine approximation at the end of the course.

20.1. **Number fields and algebraic integers.** While up to now we have mainly studied quadratic rings and fields, more general fields show up naturally when trying to solve Diophantine equations. Often one is lead to adjoin to the field $\mathbb{Q}$ of rational numbers roots of suitable polynomials with rational coefficients. Trying to solve Mordell equations $Y^2 = X^3 + k$ brought us to look at quadratic fields $\mathbb{Q}(\sqrt{k})$. If we want to study the Fermat equation $X^n + Y^n = Z^n$ it seems natural to look at fields $\mathbb{Q}(\zeta_n)$ obtained adjoining to $\mathbb{Q}$ the n-th root of unity $\zeta_n = e^{\frac{2\pi i}{n}}$.

Definition 20.1.1. *A* number field *is a field K which is* finite *over $\mathbb{Q}$, i.e. which is finite dimensional as a $\mathbb{Q}$-vector space. The dimension of K as a $\mathbb{Q}$-vector space is called the* degree *of the number field and is denoted by $[K : \mathbb{Q}]$.*

Remark 20.1.2. More generally, if K is a field and $F \subset K$ is a subfield then the dimension of K as an F-vector space is denoted by $[K : F]$ and called the *degree of* K over F. Beware this is *not* the index of the abelian group $(F, +)$ in $(K, +)$; hopefully this conflict of notation with 11.1.3 will not cause any confusion.

Example 20.1.3.
(1) Quadratic fields are number fields of degree 2.
(2) The field $\mathbb{Q}(\sqrt[3]{2}) = \mathbb{Q}[X]/(X^3 - 2)$ is a number field of degree 3. More generally, if $\alpha \in \mathbb{C}$ is an algebraic number of degree d, then $\mathbb{Q}(\alpha)$ is a number field of degree d.
(3) The field $\mathbb{Q}(\zeta_\infty) = \cup_{n\geq 1}\mathbb{Q}(\zeta_n) \subset \mathbb{C}$ is *not* a number field.

Exercise 20.1.4. Let $K \subset \mathbb{C}$ be a number field of degree d. Prove that every element of K is an algebraic number of degree dividing d.

20.1.5. *Quadratic rings and fields: summary of results.* We will generalise several properties of quadratic rings with fundamental discriminant to suitable subrings of number fields. It may be helpful to summarise our main results for quadratic rings. Let $d < 0$ be a *fundamental* discriminant, $\mathcal{O}_d$ the quadratic ring of discriminant d and $K = \mathbb{Q}(\sqrt{d})$ the fraction field of $\mathcal{O}_d$. We proved:

(1) The abelian group $(\mathcal{O}_d, +)$ is free. Moreover for every non-zero ideal $I \subset \mathcal{O}_d$ the quotient $\mathcal{O}_d/I$ is finite.
(2) The set $Cl(\mathcal{O}_d) = I(\mathcal{O}_d)/P(\mathcal{O}_d)$ is *finite*.
(3) Let $P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0 \in \mathbb{Z}[X]$ be a *monic* polynomial with *integer* coefficients. Let $\alpha \in K$ be a root of $P(X)$. Then $\alpha \in \mathcal{O}_d$.
(4) Multiplication of fractional ideals turns the set $Cl(\mathcal{O}_d)$ into an abelian group.
(5) Let $I \subset \mathcal{O}_d$ be a non-zero ideal. Then I can be written as a product of (non-zero) prime ideals of $\mathcal{O}_d$, which are unique up to reordering.

Recall that we established (1), (2) for *arbitrary* quadratic rings with negative discriminant. The key feature of rings with fundamental discriminant is (3) (= Proposition 15.1.1), which allowed us to deduce (4) from (2). Property (5) was then an easy consequence of (4).

20.1.6. *Algebraic integers: definition.* It is sometimes a good idea to turn an important property proved in a given setting into a definition in a more general context. This philosophy, applied to property (3) above, suggests the following definition.

Definition 20.1.7. *Let $\alpha \in \mathbb{C}$ be a complex number. We say that α is an* algebraic integer *if there exists a* monic *polynomial $P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0 \in \mathbb{Z}[X]$ such that $P(\alpha) = 0$.*

The set of algebraic integers in a number field K is denoted by $\mathcal{O}_K$.

Remark 20.1.8.
(1) Every algebraic integer is an algebraic number, but the converse is not true, as the next example shows.
(2) To be precise, in the above definition of $\mathcal{O}_K$ we are tacitly using the fact (which we will prove later) that every number field can be seen as a subfield of the complex numbers; furthermore whether or not an element $\alpha \in K \subset \mathbb{C}$ is an algebraic integer does not depend on the chosen embedding of K in $\mathbb{C}$. A more intrinsic way of defining $\mathcal{O}_K$ is to say that

it consists of elements of K which are *integral over* $\mathbb{Z}$. We will explain below what this terminology means.

Example 20.1.9. If $K = \mathbb{Q}$ then $\mathcal{O}_K = \mathbb{Z}$. If $K = \mathbb{Q}(\sqrt{k})$ is a quadratic field, with $k \in \mathbb{Z}$ squarefree, then Proposition 15.1.1 shows that

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

In particular, $\mathcal{O}_K$ is a ring.

20.1.10. *Plan of action (warning: spoiler).* The aim of the next few lectures is to study the properties of the set $\mathcal{O}_K$ for an arbitrary number field K. First of all we will prove that $\mathcal{O}_K$ is always a ring; then we will show that properties (1), (2) of quadratic rings still hold true for $\mathcal{O}_K$. Since (3) is true by definition, we will then deduce that (4), (5) also hold true in general.

Remark 20.1.11. Notice that in the previous lectures we *defined* quadratic rings with fundamental discriminant and *proved* that they satisfy property (3); we are now reversing the process, using (3) as a *defining property* of the set $\mathcal{O}_K$; however the fact that $\mathcal{O}_K$ is a ring now requires an argument. So, while it is wise to isolate a good definition capturing an important property like (3), this does *not*, of course, give us any theorem for free, but only moves the difficulty elsewhere.

Exercise 20.1.12. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}) = \{a_0 + a_1\sqrt{2} + a_2\sqrt{3} + a_3\sqrt{6}, a_0, \dots, a_3 \in \mathbb{Q}\}$. Then $\sqrt{2}, \sqrt{3} \in \mathcal{O}_K$. Find a monic polynomial in $\mathbb{Z}[X]$ vanishing at $\sqrt{2} + \sqrt{3}$.

20.2. **Cyclotomic polynomials and cyclotomic fields.** Following our general policy, before starting to develop the general theory outlined above we will look at a concrete, interesting example to keep in mind.

Definition 20.2.1. *Let $n \geq 1$. The n-th cyclotomic field, denoted by $\mathbb{Q}(\zeta_n)$, is the smallest subfield of $\mathbb{C}$ containing the n-th root of unity $\zeta_n = e^{\frac{2\pi i}{n}}$.*

Example 20.2.2. We have $\mathbb{Q}(\zeta_4) = \mathbb{Q}(i) = \mathbb{Q}[X]/(X^2+1)$, and $\mathbb{Q}(\zeta_6) = \mathbb{Q}(\zeta_3) = \mathbb{Q}[X]/(X^2+X+1)$. Moreover, for every odd n, we have $\mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{2n})$ (since $-\zeta_n$ is a primitive $2n$-th root of unity).

To keep things as simple as possible, we will only consider the case of $n = p$ an odd prime number. Notice that, if one is interested in showing that the Fermat equation $X^n + Y^n = Z^n$ has no non trivial solution for $n > 2$, this is a harmless restriction: if $x^n + y^n = z^n$ with $xyz \neq 0$ and $n = pm$, then (x^m, y^m, z^m) is a non trivial solution of the equation $X^p + Y^p = Z^p$. So it suffices to consider the case $n = 4$ - in which case Fermat's own proof is actually known - and $n = p$ an odd prime.

Fix an odd prime number p and let $\zeta_p = e^{\frac{2\pi i}{p}} \in \mathbb{C}$. The complex number ζ_p is a root of the *p-th cyclotomic polynomial*

$$\Phi_p(X) = \frac{X^p - 1}{X - 1} = 1 + X + X^2 + \cdots + X^{p-1} \in \mathbb{Z}[X].$$

Proposition 20.2.3. (1) *The polynomial $\Phi_p(X) \in \mathbb{Q}[X]$ is irreducible.*

(2) *The map $\mathbb{Q}[X] \to \mathbb{Q}(\zeta_p)$, $X \mapsto \zeta_p$ induces an isomorphism $\mathbb{Q}[X]/(\Phi_p(X)) \xrightarrow{\sim} \mathbb{Q}(\zeta_p)$. In particular, $\mathbb{Q}(\zeta_p)$ is a number field of degree $p-1$.*

Proof. Notice that (2) follows from (1). Indeed, (1) tells us that $\Phi_p(X) \in \mathbb{Q}[X]$ is a monic irreducible polynomial vanishing at ζ_p, hence it is the minimal polinomial of ζ_p. It follows that the map $\mathbb{Q}[X] \to \mathbb{Q}(\zeta_p)$ sending X to ζ_p factors through ann isomorphism $\iota : \mathbb{Q}[X]/(\Phi_p(X)) \hookrightarrow \mathbb{Q}(\zeta_p)$. Furthermore a basis of $\mathbb{Q}[X]/(\Phi_p(X))$ as a $\mathbb{Q}$-vector space is $1, X, \dots, X^{p-2}$, hence the $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = p - 1$.

It remains to prove (1). As $X \mapsto X + 1$ is an automorphism of $\mathbb{Q}[X]$ we may as well show that $\Phi_p(X+1) = \frac{(X+1)^p - 1}{(X+1)-1} = X^{p-1} + pX^{p-2} + \cdots + p$ is irreducible in $\mathbb{Q}[X]$. Assume that $\Phi_p(X+1) = PQ$ for some $P, Q \in \mathbb{Q}[X]$, and let λ the leading coefficient of P. Then $\Phi_p(X+1) = \frac{1}{\lambda}P \cdot (\lambda Q)$, and both $\frac{1}{\lambda}P$ and λQ are monic. It follows from Gauss' Lemma 15.1.2 that $\frac{1}{\lambda}P, \lambda Q \in \mathbb{Z}[X]$. Hence it suffices to show that $\Phi_p(X+1)$ is irreducible in $\mathbb{Z}[X]$. This follows from Eisenstein's irreducibility criterion, which is recalled below. □

Proposition 20.2.4 (Eisenstein's irreducibility criterion)**.** *Let $P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in \mathbb{Z}[X]$ be a monic polynomial with integer coefficients. Assume that*

$$p \mid a_i,\ i = 0, 1, \ldots, n-1$$
$$p^2 \nmid a_0;$$

then $P(X)$ is irreducible in $\mathbb{Z}[X]$.

Proof. Suppose that $P = QR$ for some $Q, R \in \mathbb{Z}[X] \smallsetminus \mathbb{Z}$. As P is monic, up to changing signs to Q and R we may (and will) assume that Q, R are monic. Write $Q(X) = X^m + b_{m-1}X^{m-1} + \ldots + b_0$ and $R(X) = X^{n-m} + c_{n-m-1}X^{n-m-1} + \ldots + c_0$ Let $\bar{P} \in \mathbb{F}_p[X]$ (resp. $\bar{Q} \in \mathbb{F}_p[X]$, resp. $\bar{R} \in \mathbb{F}_p[X]$) be the reduction modulo p of P (resp. Q, resp. R). Then we obtain

$$X^n = \bar{P} = \bar{Q}\bar{R} \in \mathbb{F}_p[X] \Rightarrow \bar{Q} = X^m, \bar{R} = X^{n-m};$$

hence we must have $p \mid b_i$ for $0 \leq i \leq m-1$ and $p \mid c_j$ for $0 \leq j \leq n-m-1$. This implies that $p^2 \mid b_0 c_0 = a_0$, contradiction. □

20.2.5. *Kummer's approach to the Fermat equation.* In order to study integer solutions of the Fermat equation $X^p + Y^p = Z^p$, one idea is to write it in the form

$$(X+Y)(X+\zeta_p Y)(X+\zeta_p^2 Y)\cdots(X+\zeta_p^{p-1}Y) = Z^p$$

and to work in the ring $\mathbb{Z}[\zeta_p] = \{a_0 + a_1\zeta_p + \ldots + a_{p-2}\zeta_p^{p-2}, a_0, \ldots, a_{p-2} \in \mathbb{Z}\}$. In the best of all possible worlds, one may hope that unique factorisation holds in these rings and the separating powers trick can be used to show that the equation has no non trivial integer solutions. Kummer soon noticed that life is not so easy [17], [18]; however he also observed that introducing suitable *ideal numbers* (later replaced by *ideals* by Dedekind) one could circumvent the problem of failure of unique factorisation in some cases. We have already seen this phenomenon when we tried to solve Mordell equations. In the next lectures we will extend our previous analysis: we will prove that, for $K = \mathbb{Q}(\zeta_p)$, we have $\mathcal{O}_K = \mathbb{Z}[\zeta_p]$; the general results mentioned above, when applied to this example, can then be used to obtain information on the Fermat equation.

Remark 20.2.6. It is perhaps worth pointing out that the Fermat equation does not seem to have been the main motivation behind Kummer's work on the arithmetic properties of cyclotomic fields. Instead, Kummer's main desire was to generalise the quadratic reciprocity law - similarly, Gauss introduced Gaussian integers in order to study biquadratic reciprocity. Kummer's work was later refined by Dedekind and Hilbert (among others) who investigated the properties of general number fields. Dedekind's (resp. Hilbert's) own exposition of the subject can be found in [10] (resp. [14]).

20.3. **Basic properties of algebraic integers.** We will now study some general properties of algebraic integers. It is actually useful to work with the following more general definition.

Definition 20.3.1. *Let B be a ring and $A \subset B$ a subring. We say that $b \in B$ is* integral over A *if there exists a* monic *polynomial*[9] *$P(X) \in A[X]$ such that $P(b) = 0$.*

We say that $A \subset B$ is an integral extension*, or that B is* integral over A*, if every element of B is integral over A.*

Example 20.3.2. If $A = \mathbb{Z}$ and $B = \mathbb{C}$ then we recover algebraic integers as defined in 20.1.7. If $A = \mathbb{Q}$ and $B = \mathbb{C}$ we find the definition of an algebraic number.

Lemma 20.3.3. *Let A be an integral domain, F its fraction field and K/F a field extension. Let $\mathbf{M}$ be a $n \times n$ matrix with coefficients in A. Let $x \in K$ be an eigenvalue of $\mathbf{M}$ for some non-zero eigenvector $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in K^n$, i.e. assume that we have*

$$\mathbf{M} \cdot \mathbf{v}^t = x \cdot \mathbf{v}^t.$$

Then x is a root of the characteristic polynomial of $\mathbf{M}$. In particular, x is integral over A.

Proof. As $(\mathbf{M} - x \cdot Id)\mathbf{v}^t = 0$ and $K^n \ni \mathbf{v} \neq 0$ we must have $\det(\mathbf{M} - x \cdot Id) = 0$, i.e. x is a root of the characteristic polynomial $P_{\mathbf{M}}(X) = \det(X \cdot Id - \mathbf{M})$. As M has coefficients in A we have $\det(X \cdot Id - \mathbf{M}) \in A[X]$; since this polynomial is monic we deduce that x is integral over A. □

[9] N. B. By definition a monic polynomial is non-zero.

Theorem 20.3.4. (1) *Let B be an integral domain and $A \subset B$ be a subring. Assume that $b_1, b_2 \in B$ are integral over A. Then $b_1 + b_2, b_1 - b_2, b_1 \cdot b_2 \in B$ are all integral over A.*

(2) *Let A be an integral domain with fraction field F. Let K/F be a field extension. Then the set*

$$\{x \in K \mid x \text{ is integral over } A\}$$

is a subring of K.

Proof. Let us prove (1). Let F (resp. K) be the fraction field of A (resp. B). Take $P(X) \in A[X]$ monic of degree n (resp. $Q(X) \in A[X]$ monic of degree m) such that $P(b_1) = 0$ (resp. $Q(b_2) = 0$). Let $\mathbf{w} = (1, b_2, b_2^2, \ldots, b_2^{m-1}) \in K^m$ and

$$\mathbf{v} = (\mathbf{w}, b_1\mathbf{w}, b_1^2\mathbf{w}, \ldots, b_1^{n-1}\mathbf{w}) \in K^{mn}.$$

Let us prove that $b_1 + b_2$ is integral over A. For every $0 \leq i \leq n-1, 0 \leq j \leq m-1$, the elements $b_1 \cdot b_1^i b_2^j$ and $b_2 \cdot b_1^i b_2^j$ are linear combinations with A-coefficients of the coordinates of $\mathbf{v}$. Hence the same is true for the element $(b_1 + b_2)b_1^i b_2^j$; it follows that there exists a matrix $\mathbf{M}_{b_1+b_2}$ with A-coefficients such that

$$(b_1 + b_2)\mathbf{v}^t = \mathbf{M}_{b_1+b_2} \cdot \mathbf{v}^t;$$

Lemma 20.3.3 implies that $b_1 + b_2$ is integral over A. As $-b$ is integral over A whenever b is (check this) it follows that $b_1 - b_2$ is also integral over A. Finally, integrality of $b_1 \cdot b_2$ is proved in a similar way: for $0 \leq i \leq n-1, 0 \leq j \leq m-1$, the element $(b_1 b_2) \cdot b_1^i b_2^j$ is a linear combination with A-coefficients of the coordinates of $\mathbf{v}$. Hence there exists a matrix $\mathbf{M}_{b_1 b_2}$ with A-coefficients such that $(b_1 b_2)\mathbf{v}^t = \mathbf{M}_{b_1 b_2} \cdot \mathbf{v}^t$, so $b_1 b_2$ is integral over A in view of Lemma 20.3.3. This proves assertion (1), from which (2) follows. □

Remark 20.3.5. Notice that the technique of proof of the previous result, based on Lemma 20.3.3, can also be adapted to prove that every element of a number field is an algebraic number.

Taking $A = \mathbb{Z}$ and K a number field in point (2) of the previous theorem we obtain

Corollary 20.3.6. *Let K be a number field. The set $\mathcal{O}_K = \{x \in K \mid \exists P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in \mathbb{Z}[X], P(x) = 0\}$ is a subring of K.*

Definition 20.3.7. *The ring $\mathcal{O}_K$ is called the* ring of integers *of the number field K.*

Example 20.3.8. Taking $A = \mathbb{Z}$ and $K = \mathbb{C}$ in (2) of the above theorem we deduce that the set of *all* algebraic integers is a ring. It is not noetherian, but it is a Bézout domain (we will be able to show this later).

Definition 20.3.9. *Let A be an integral domain with fraction field F. Let K/F be a field extension. The ring*

$$B = \{x \in K \mid x \text{ is integral over } A\}$$

is called the integral closure *of A in K.*

If $A = F$ (i. e. A is a field) then B is a field (exercise) and is most often called the algebraic closure *of F in K.*

Exercise 20.3.10. Let $\bar{\mathbb{Q}}$ be the algebraic closure the $\mathbb{Q}$ in $\mathbb{C}$. Prove that $\bar{\mathbb{Q}}$ is an *algebraically closed field*, i. e. every polynomial with coefficients in $\bar{\mathbb{Q}}$ has all its roots in $\bar{\mathbb{Q}}$.

21. Lecture 21: Norm and trace

In this lecture we generalise the definition of trace and norm of an element in a quadratic field to arbitrary number fields.

21.1. **Recollection on quadratic fields and rings.** Let us first recall and slightly reformulate the definitions of trace and norm given in 5.1. Let $1 \neq k \in \mathbb{Z}$ be a squarefree integer. Then the quadratic field $K = \mathbb{Q}(\sqrt{k}) = \{a + b\sqrt{k} \mid a, b \in \mathbb{Q}\} \subset \mathbb{C}$ has two embeddings in the field of complex numbers:

$$\begin{aligned} \tau_1 : K &\to \mathbb{C}, & \tau_2 : K &\to \mathbb{C} \\ a + b\sqrt{k} &\mapsto a + b\sqrt{k} & a + b\sqrt{k} &\mapsto a - b\sqrt{k}. \end{aligned}$$

The trace and norm of $\alpha = a + b\sqrt{k} \in K$ are given by $Tr(\alpha) = \tau_1(\alpha) + \tau_2(\alpha)$, $N(\alpha) = \tau_1(\alpha)\tau_2(\alpha)$. The polynomial $P_\alpha(X) = X^2 - Tr(\alpha)X + N(\alpha) = (X - \tau_1(\alpha))(X - \tau_2(\alpha)) \in \mathbb{Q}[X]$ coincides with the characteristic polynomial of the multiplication by α map $m_\alpha : K \to K$. If α belongs to the ring of integers of K then $P_\alpha(X) \in \mathbb{Z}[X]$.

21.2. **Trace, norm and characteristic polynomial.** Let K be a number field. For every $\alpha \in K$ the map

$$\begin{aligned} m_\alpha : K &\to K \\ x &\mapsto \alpha x \end{aligned}$$

is a $\mathbb{Q}$-linear endomorphism of K. Let $P_\alpha = \det(X \cdot \mathrm{Id} - m_\alpha) \in \mathbb{Q}[X]$ be the characteristic polynomial of m_α. We will call it the *characteristic polynomial of* α.

Definition 21.2.1. *The* trace *(resp.* norm*) of α is the trace (resp. determinant) of m_α; it is denoted by $Tr(\alpha)$ (resp. $N(\alpha)$).*

Remark 21.2.2. As the trace is a $\mathbb{Q}$-linear map and the determinant is a multiplicative map the maps

$$\begin{aligned} Tr : K &\to \mathbb{Q} & N : K &\to \mathbb{Q} \\ \alpha &\mapsto Tr(\alpha) & \alpha &\mapsto N(\alpha) \end{aligned}$$

are respectively $\mathbb{Q}$-linear and multiplicative.

We wish to express trace and norm of an element α in a number field K in terms of its images via the complex embeddings of K. This is easy in the case $K = \mathbb{Q}(\alpha)$; to deal with the general case we will need to work in a slightly more general setting. The reader may benefit from looking at Example 21.2.7 while reading the proofs of the next two statements.

Lemma 21.2.3. *Let $F \subset K$ be number fields, and let $n = [K : F]$. For every field embedding $\tau : F \to \mathbb{C}$, the set $\{\sigma : K \to \mathbb{C}$ field morphism, $\sigma_{|F} = \tau\}$ consists of n elements.*

Proof. Fix $\tau : F \to \mathbb{C}$; let $\Sigma^\tau_{K/F} = \{\sigma : K \to \mathbb{C}$ field morphism, $\sigma_{|F} = \tau\}$ and $|\Sigma^\tau_{K/F}|$ its cardinality. We want to show that $|\Sigma^\tau_{K/F}| = [K : F]$.

Let us first deal with the case $K = F(\alpha)$ for some $\alpha \in K$. Let $P(X) \in F[X]$ be the monic polynomial of least degree such that $P(\alpha) = 0$. Then we have $K = F[X]/(P)$, hence field embeddings $\sigma : K \to \mathbb{C}$ such that $\sigma_{|F} = \tau$ are in bijection with roots of $P(X)$ in $\mathbb{C}$ (which are distinct, as P is irreducible): if β is such a root, then there is a unique σ extending τ which sends X to β. Hence in this case we find that $[K : F] = \deg(P(X)) = |\Sigma^\tau_{K/F}|$.

To prove the general case we argue by induction on $[K : F]$. If $[K : F] = 1$ there is nothing to prove; otherwise let $\alpha \in K \smallsetminus F$ and $L = F(\alpha)$. The result holds for L/F by the previous argument, and it does for K/L by the induction hypothesis. To complete the proof it suffices to prove the following equalities:

(1) $[K : F] = [K : L][L : F]$;
(2) For every $\tau : F \to \mathbb{C}$ and $\iota : L \to \mathbb{C}$ extending τ, we have $|\Sigma^\tau_{K/F}| = |\Sigma^\iota_{K/L}||\Sigma^\tau_{L/F}|$.

The first one is a linear algebra exercise left to the reader. To prove the second equality, let $r : \Sigma^\tau_{K/F} \to \Sigma^\tau_{L/F}$ be the restriction map. Take $\sigma, \sigma' \in \Sigma^\tau_{K/F}$ such that $r(\sigma) = r(\sigma')$. Letting $\iota = r(\sigma) = r(\sigma')$ we find that σ, σ' both restrict to $\iota : L \to \mathbb{C}$, hence $\sigma, \sigma' \in \Sigma^\iota_{K/L}$. It remains to show that the restriction map r is surjective: the case $K = L(\beta)$ for some β follows from the first part of the proof; iterating it one deduces the general case (fill in the details). □

Remark 21.2.4. In fact, by the primitive element theorem, in the situation of the above lemma it is always true that $K = F(\alpha)$ for some $\alpha \in K$.

Proposition 21.2.5. *Let $F \subset K \subset \mathbb{C}$ be number fields; let us denote by $\iota : F \to \mathbb{C}$ the inclusion and let $\alpha \in K$. Let $\Sigma_{K/F} = \{\sigma : K \to \mathbb{C}$ field morphism, $\sigma_{|F} = \iota\}$. Let $P^{K/F}_\alpha(X) \in F[X]$ be the characteristic polynomial of the F-linear map $K \to K$, $x \mapsto \alpha x$. Then*

$$P^{K/F}_\alpha(X) = \prod_{\tau \in \Sigma_{K/F}} (X - \tau(\alpha)).$$

Proof. Let us first suppose that $K = F(\alpha) = F[X]/(P)$ for $P \in F[X]$ monic of minimal degree such that $P(\alpha) = 0$. Then, as seen in the proof of the previous lemma, roots of P in $\mathbb{C}$ are distinct and of the form $\tau(\alpha)$ for $\tau \in \Sigma_{K/F}$, hence $P(X) = \prod_{\tau \in \Sigma_{K/F}}(X - \tau(\alpha))$. Furthermore using the Cayley-Hamilton theorem we see that $P^{K/F}_\alpha(\alpha) = 0$; it follows that $P^{K/F}_\alpha(\tau(\alpha)) = 0$ for every $\tau \in \Sigma_{K/F}$. Therefore P divides $P^{K/F}_\alpha$ and, as both polynomials are monic of degree $[K : F]$, they must be equal. Hence $P^{K/F}_\alpha(X) = \prod_{\tau \in \Sigma_{K/F}}(X - \tau(\alpha))$.

Let us now deal with the general case. Let $\alpha \in K$ and set $L = F(\alpha)$. Let $e_1, \dots, e_m \in K$ be a basis of K as an L-vector space. Then we have

$$K = \oplus_{i=1}^m L e_i;$$

for $1 \leq i \leq m$ multiplication by α sends the F-subspace Le_i to itself, and the characteristic polynomial of the induced map on Le_i is $P^{L/F}_\alpha(X)$. It follows that $P^{K/F}_\alpha(X) = (P^{L/F}_\alpha(X))^m$.

On the other hand, by Lemma 21.2.3 for every embedding $\sigma \in \Sigma_{L/F}$ there are $m = [K : L]$ embeddings $\tau \in \Sigma_{K/F}$ such that $\tau_{|L} = \sigma$. It follows that

$$\prod_{\tau \in \Sigma_{K/F}} (X - \tau(\alpha)) = \left(\prod_{\sigma \in \Sigma_{L/F}} (X - \sigma(\alpha)) \right)^m.$$

We know from the first part of the proof that $P^{L/F}_\alpha(X) = \prod_{\sigma \in \Sigma_{L/F}}(X - \sigma(\alpha))$, hence the proposition follows. □

Notation 21.2.6. For a number field K, we let $\Sigma_K = \{\sigma : K \to \mathbb{C}$ field morphism$\}$.

Example 21.2.7. Let $F = \mathbb{Q}$ and let K be any number field. Then the previous results tell us that

(1) Σ_K has cardinality $[K : \mathbb{Q}]$.
(2) For every $\alpha \in K$ we have $P_\alpha = \prod_{\tau \in \Sigma_K}(X - \tau(\alpha))$. In particular $Tr(\alpha) = \sum_{\tau \in \Sigma_K} \tau(\alpha)$ and $N(\alpha) = \prod_{\tau \in \Sigma_K} \tau(\alpha)$.

Let us see the above proofs in action in one concrete example: let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}) \subset \mathbb{C}$; letting $L = \mathbb{Q}(\sqrt{2})$ we have a tower of fields

$$F = \mathbb{Q} \subset L \subset K$$

such that $[L : F] = [K : L] = 2$. Furthermore $\Sigma_L = \{\sigma_1, \sigma_2\}$ where

$$\begin{aligned} \sigma_1 : L &\to \mathbb{C}, & \sigma_2 : L &\to \mathbb{C} \\ a + b\sqrt{2} &\mapsto a + b\sqrt{2} & a + b\sqrt{2} &\mapsto a - b\sqrt{2}. \end{aligned}$$

With the notation in the proof of Lemma 21.2.3, we have $\Sigma^{\sigma_1}_{K/L} = \{\tau^1_1, \tau^2_1\}$ where

$$\begin{aligned} \tau^1_1 : K &\to \mathbb{C}, & \tau^2_1 : K &\to \mathbb{C} \\ \sqrt{2} &\mapsto \sqrt{2} & \sqrt{2} &\mapsto \sqrt{2} \\ \sqrt{3} &\mapsto \sqrt{3} & \sqrt{3} &\mapsto -\sqrt{3} \end{aligned}$$

and likewise $\Sigma_{K/L}^{\sigma_2} = \{\tau_2^1, \tau_2^2\}$ where

$$\begin{aligned} \tau_2^1 : K &\to \mathbb{C}, & \tau_2^2 : K &\to \mathbb{C} \\ \sqrt{2} &\mapsto -\sqrt{2} & \sqrt{2} &\mapsto -\sqrt{2} \\ \sqrt{3} &\mapsto \sqrt{3} & \sqrt{3} &\mapsto -\sqrt{3}. \end{aligned}$$

It follows that, letting $\alpha = \sqrt{2}$, we have

$$\begin{aligned} \prod_{\tau \in \Sigma_K} (X - \tau(\alpha)) &= (X - \tau_1^1(\alpha))(X - \tau_1^2(\alpha))(X - \tau_2^1(\alpha))(X - \tau_2^2(\alpha)) \\ &= (X - \sqrt{2})^2 (X + \sqrt{2})^2 = (X^2 - 2)^2. \end{aligned}$$

On the other hand, let us compute the characteristic polynomial P_α of the multiplication by α map on K. Let us choose the $\mathbb{Q}$-basis $(1, \sqrt{2}, \sqrt{3}, \sqrt{3}\sqrt{2})$ of K. Notice that the first couple of numbers is a $\mathbb{Q}$-basis of L, and the second couple is a $\mathbb{Q}$-basis of $\sqrt{3}L \subset K$. With respect to this basis, multiplication by α is represented by the matrix

$$\mathbf{M}_\alpha = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

hence we find $P_\alpha(X) = (X^2 - 2)^2$.

Exercise 21.2.8. Compute trace, norm and characteristic polynomial of $\sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Corollary 21.2.9. *Let K be a number field and $\mathcal{O}_K$ its ring of integers. For every $\alpha \in \mathcal{O}_K$, the polynomial $P_\alpha(X)$ belongs to $\mathbb{Z}[X]$. In particular $N(\alpha)$ and $Tr(\alpha)$ belong to $\mathbb{Z}$.*

Proof. We have $P_\alpha(X) = \prod_{\tau \in \Sigma_K} (X - \tau(\alpha)) \in \mathbb{Q}[X]$ by the previous proposition. On the other hand, as α is an algebraic integer, the number $\tau(\alpha)$ is also an algebraic integer for every $\tau \in \Sigma_K$. Let a be a coefficient of $P_\alpha(X)$: as a is a sum of products of numbers $\tau(\alpha)$ for $\tau \in \Sigma_K$, Theorem 20.3.4 implies that a is an algebraic integer. As $a \in \mathbb{Q}$, we deduce that $a \in \mathbb{Z}$. □

Exercise 21.2.10. Let $f(X) = X^3 + 5X^2 + 10X + 5 \in \mathbb{Q}[X]$; let $\alpha, \beta, \gamma \in \mathbb{C}$ be the roots of $f(X)$.

(1) Prove that $\alpha^{100} + \beta^{100} + \gamma^{100} \in \mathbb{Z}$.
(2) Show that $f : \mathbb{R} \to \mathbb{R}$ is an increasing function. Deduce that two of the roots of f belong to $\mathbb{C} \smallsetminus \mathbb{R}$.
(3) Say $\alpha \in \mathbb{C} \smallsetminus \mathbb{R}$. Let $K = \mathbb{Q}(\alpha)$. Prove that the equation $X^2 + Y^2 = -1$ has no solution $(x, y) \in K^2$.

Example 21.2.11. Let p be an odd prime and $K = \mathbb{Q}(\zeta_p) = \mathbb{Q}[X]/(\Phi_p(X))$. The set $\Sigma_K = \{\tau : K \to \mathbb{C}\}$ can be identified with $(\mathbb{Z}/p\mathbb{Z})^\times$, sending $j \in (\mathbb{Z}/p\mathbb{Z})^\times$ to the map induced by $X \mapsto \zeta_p^j$. We have

$$P_{\zeta_p - 1} = \prod_{j \in (\mathbb{Z}/p\mathbb{Z})^\times} (X - (\zeta_p^j - 1)) = \Phi_p(1 + X)$$

hence, for any $j \in (\mathbb{Z}/p\mathbb{Z})^\times$, we find $Tr(1 - \zeta_p^j) = p$ and $N(1 - \zeta_p^j) = p$.

Exercise 21.2.12. Let $P(X) = X^4 + 4X^2 + 2 \in \mathbb{Q}[X]$ and $K = \mathbb{Q}[X]/(P(X))$. Prove that K is a number field of degree 4. Compute the trace and norm of $X, X^2 \in K$.

22. Lecture 22: Algebraic properties of $\mathcal{O}_K$ & cyclotomic fields bis

In this lecture we prove that for every number field K the ring $\mathcal{O}_K$ is a free abelian group of rank $[K:\mathbb{Q}]$; then we study the main algebraic properties of $\mathcal{O}_K$. We will also determine the ring of integers of the cyclotomic field $\mathbb{Q}(\zeta_p)$.

22.1. **The trace form.** Let K be a number field and $\mathcal{O}_K$ its ring of integers. Our aim is to show that $(\mathcal{O}_K,+)$ is a free abelian group of rank $[K:\mathbb{Q}]$. It is easy to show that $\mathcal{O}_K$ contains a free abelian group of rank $[K:\mathbb{Q}]$ (see Lemma 22.2.1 below) hence, by Proposition 11.1.2, it suffices to prove that $\mathcal{O}_K$ is contained in a free abelian group of the same rank. This rests on the properties of the *trace form* on K, defined as follows:

$$\begin{aligned}\langle\cdot,\cdot\rangle : K\times K &\to \mathbb{Q}\\ (\alpha,\beta) &\mapsto \langle\alpha,\beta\rangle = Tr(\alpha\beta).\end{aligned}$$

The trace form enjoys the following properties:

Bilinearity: for $\alpha,\beta,\gamma\in K$, the equalities $\langle\alpha+\beta,\gamma\rangle = \langle\alpha,\gamma\rangle+\langle\beta,\gamma\rangle$ and $\langle\alpha,\beta+\gamma\rangle = \langle\alpha,\beta\rangle+\langle\alpha,\gamma\rangle$ hold.
Symmetry: for $\alpha,\beta\in K$, we have $\langle\alpha,\beta\rangle=\langle\beta,\alpha\rangle$.
Non-degeneracy: for every $\alpha\in K\smallsetminus\{0\}$ there exists $\beta\in K\smallsetminus\{0\}$ such that $\langle\alpha,\beta\rangle\neq 0$.

Proof. Bilinearity follows from linearity of $Tr: K\to\mathbb{Q}$. Symmetry holds since $K^\times$ is commutative (or because the trace is independent of the order of composition of two endomorphisms); finally, for $\alpha\in K\smallsetminus\{0\}$ we have $\langle\alpha,\alpha^{-1}\rangle = Tr(\alpha\alpha^{-1}) = Tr(1) = [K:\mathbb{Q}]\neq 0$, proving that the trace form is non-degenerate. □

22.2. **Freeness of $\mathcal{O}_K$.**

Lemma 22.2.1. *For every $\alpha\in K$ there exists $d\in\mathbb{Z}\smallsetminus\{0\}$ such that $d\alpha\in\mathcal{O}_K$. In particular there exists a basis $(e_1,\dots,e_{[K:\mathbb{Q}]})$ of K as a $\mathbb{Q}$-vector space such that $e_i\in\mathcal{O}_K$ for $1\leq i\leq[K:\mathbb{Q}]$.*

Proof. Let $\alpha\in K$. Then α is an algebraic number, i.e. there exists a (monic) polynomial $P(X) = X^n+a_{n-1}X^{n-1}+\ldots+a_0\in\mathbb{Q}[X]$ such that $P(\alpha)=0$ (see Remark 20.3.5). Write $a_i=\frac{b_i}{d}$ with $b_i,d\in\mathbb{Z}$. Then

$$d^n\alpha^n + b_{n-1}(d^{n-1}\alpha^{n-1}) + b_{n-2}d(d^{n-2}\alpha^{n-2})+\ldots+d^{n-1}b_0 = 0.$$

As $X^n+b_{n-1}X^{n-1}+\ldots+d^{n-1}b_0\in\mathbb{Z}[X]$ is monic, the element $d\alpha$ belongs to $\mathcal{O}_K$. This proves the first assertion of the lemma. Applying it to each element of a basis of K we deduce the second assertion. □

22.2.2. *Cramer's rule.* The previous lemma implies that $\mathcal{O}_K$ contains a free abelian group of rank $[K:\mathbb{Q}]$; we now want to prove that $\mathcal{O}_K$ is contained in a free abelian group of the same rank. This rests on the *non-degeneracy* of the trace form and makes use of Cramer's rule, which we recall: let F be a field, $\mathbf{A}\in GL_n(F)$ and $\mathbf{b}=(b_1,\dots,b_n)\in F^n$. Then the system of linear equations $\mathbf{A}\mathbf{x}^t=\mathbf{b}^t$ has a unique solution $(x_1,\dots,x_n)\in F^n$, given by the formula $x_i=\frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})}$, where, for $1\leq i\leq n$, $\mathbf{A}_i$ is the matrix obtained replacing the i-th column of $\mathbf{A}$ with $\mathbf{b}^t$.

Theorem 22.2.3. *Let K be a number field. Then $(\mathcal{O}_K,+)$ is a free abelian group of rank $[K:\mathbb{Q}]$.*

Proof. Let $n=[K:\mathbb{Q}]$ and let $(e_1,\dots,e_n)$ be a basis of K with $e_i\in\mathcal{O}_K$, whose existence is guaranteed by the previous lemma. Let $\mathbf{M}$ be the matrix representing the trace form in this basis: concretely, for $1\leq i,j\leq n$, the $(i,j)-$th entry of $\mathbf{M}$ is $Tr(e_ie_j)$. It follows from Corollary 21.2.9 that $\mathbf{M}$ has integer coefficients, hence $\det(\mathbf{M})\in\mathbb{Z}$. Better, as the trace form is non-degenerate we must have $\det(\mathbf{M})\in\mathbb{Z}\smallsetminus\{0\}$.

Let $\alpha\in\mathcal{O}_K$; write $\alpha=\sum_{i=1}^n\lambda_ie_i$ with $\lambda_i\in\mathbb{Q}$. We wish to bound the denominators of the coefficients λ_i. We know that for $1\leq j\leq n$ we have $\alpha e_j\in\mathcal{O}_K$, hence

$$Tr(\alpha e_j)=\sum_{i=1}^n\lambda_iTr(e_ie_j)\in\mathbb{Z}\Rightarrow(\lambda_1,\dots,\lambda_n)\mathbf{M}\in\mathbb{Z}^n.$$

Cramer's rule implies that $\det(\mathbf{M})(\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$. Hence we have proved that

$$\mathbb{Z}e_1 \oplus \cdots \oplus \mathbb{Z}e_n \subset \mathcal{O}_K \subset \mathbb{Z}\frac{1}{\det(\mathbf{M})}e_1 \oplus \ldots \oplus \mathbb{Z}\frac{1}{\det(\mathbf{M})}e_n$$

i. e. the abelian group $\mathcal{O}_K$ is sandwiched between two free abelian groups of rank n. Proposition 11.1.2 implies that $\mathcal{O}_K$ is itself free of rank n. □

Remark 22.2.4. The previous theorem rests on the following facts:

(1) The trace form is non-degenerate.
(2) The trace of an element of $\mathcal{O}_K$ is an integer.

We established the second property in the previous lecture as a consequence of the equality $Tr(\alpha) = \sum_{\sigma \in \Sigma_K} \sigma(\alpha)$. An alternative argument is given in the next exercise.

Exercise 22.2.5. Let K be a number field and $\alpha \in K$.

(1) Show that the characteristic polynomial P_α is a power of the minimimal polynomial P_α^{min}.
(2) Prove that $\alpha \in \mathcal{O}_K$ if and only if $P_\alpha^{min} \in \mathbb{Z}[X]$.
(3) Deduce that if $\alpha \in \mathcal{O}_K$ then $P_\alpha \in \mathbb{Z}[X]$, hence the trace of α is an integer.

22.2.6. *The discriminant of $\mathcal{O}_K$.* By the above theorem we can pick a $\mathbb{Z}$-basis $(e_1, \dots, e_n)$ of $\mathcal{O}_K$ and look at the matrix $\mathbf{M}$ representing the trace form with respect to this basis, i. e. satisfying

$$\forall\, \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n,\ \ \langle e_1x_1 + \ldots + e_nx_n, e_1x_1 + \ldots + e_nx_n \rangle = \mathbf{x}\mathbf{M}\mathbf{x}^t.$$

Any two bases of $\mathcal{O}_K$ as an abelian group differ by multiplication by a matrix in $\mathbf{U} \in GL_{[K:\mathbb{Q}]}(\mathbb{Z})$, which has determinant ± 1. If we replace the basis $(e_1, \dots, e_n)$ by $(e_1, \dots, e_n)\mathbf{U}$ then $\mathbf{M}$ gets replaced by $\mathbf{U}^t\mathbf{M}\mathbf{U}$. As $\det(\mathbf{U}^t\mathbf{M}\mathbf{U}) = \det(\mathbf{M})$ we can give the following definition.

Definition 22.2.7. *Let K be a number field. The* discriminant *of $\mathcal{O}_K$, denoted by $\Delta(\mathcal{O}_K)$, is the determinant of the matrix representing the trace form in a chosen basis of $\mathcal{O}_K$.*

Remark 22.2.8. Notice that the discriminant of $\mathcal{O}_K$ is a *non-zero* integer, because the trace form is non-degenerate.

Example 22.2.9. Let $k \neq 1$ be a squarefree integer and $K = \mathbb{Q}(\sqrt{k})$. We know from Example 20.1.9 that

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

If $k \equiv 2, 3 \pmod 4$ a $\mathbb{Z}$-basis of $\mathcal{O}_K$ is $(1, \sqrt{k})$ and the trace form in this basis is represented by $\begin{pmatrix} 2 & 0 \\ 0 & 2k \end{pmatrix}$ hence $\Delta(\mathcal{O}_k) = 4k$. If $k \equiv 1 \pmod 4$ then a $\mathbb{Z}$-basis of $\mathcal{O}_K$ is $(1, \frac{1+\sqrt{k}}{2})$ and the trace form in this basis is represented by $\begin{pmatrix} 2 & 1 \\ 1 & \frac{1+k}{2} \end{pmatrix}$ hence $\Delta(\mathcal{O}_K) = k$. Hence the notion of discriminant we introduced agrees with the terminology coming from quadratic forms.

22.3. **The ring of integers of $\mathbb{Q}(\zeta_p)$.** In this section we will prove that the ring of integers of $K = \mathbb{Q}(\zeta_p)$ is $\mathbb{Z}[\zeta_p]$. Notice that ζ_p is an algebraic integer, hence we certainly have $\mathbb{Z}[\zeta_p] \subset \mathcal{O}_K$. We have to prove the opposite inclusion.

Lemma 22.3.1. *Let $K = \mathbb{Q}(\zeta_p)$.*

(1) $(1 - \zeta_p)\mathcal{O}_K \cap \mathbb{Z} = p\mathbb{Z}$.
(2) *Let $j \in (\mathbb{Z}/p\mathbb{Z})^\times$. Then*
$$\frac{1 - \zeta_p^j}{1 - \zeta_p} \in \mathcal{O}_K^\times.$$
(3) *For any $\alpha \in \mathcal{O}_K$, we have $Tr(\alpha(1 - \zeta_p)) \in p\mathbb{Z}$.*

Proof. We have seen in Example 21.2.11 that $N(1 - \zeta_p) = p$, hence $p\mathbb{Z} \subset (1 - \zeta_p)\mathcal{O}_K \cap \mathbb{Z}$. As $(1 - \zeta_p)\mathcal{O}_K \cap \mathbb{Z}$ is an ideal of $\mathbb{Z}$, if the inclusion is proper then we must have $(1 - \zeta_p)\mathcal{O}_K \cap \mathbb{Z} = \mathbb{Z}$. Hence, if this is the case then there exists $\alpha \in \mathcal{O}_K$ such that $\alpha(1 - \zeta_p) = 1$. But $N(1 - \zeta_p) = p$ and $N(\alpha) \in \mathbb{Z}$, hence we obtain a contradiction. This proves (1).

To prove (2), it suffices to show that $\frac{1-\zeta_p}{1-\zeta_p^j} \in \mathcal{O}_K$. Take $t \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that $tj \equiv 1 \pmod{p}$. Then

$$(1-\zeta_p) = (1-\zeta_p^{tj}) = (1-\zeta_p^j)(1+\zeta_p^j+\zeta_p^{2j}+\cdots+\zeta_p^{(t-1)j}) \Rightarrow \frac{1-\zeta_p}{1-\zeta_p^j} = (1+\zeta_p^j+\zeta_p^{2j}+\cdots+\zeta_p^{(t-1)j}) \in \mathcal{O}_K.$$

To show (3), take $\alpha \in \mathcal{O}_K$. For every $j \in (\mathbb{Z}/p\mathbb{Z})^\times$ let $\tau_j : K \to \mathbb{C}$ be the embedding sending ζ_p to ζ_p^j. Then

$$Tr(\alpha(1-\zeta_p)) = \sum_{j\in(\mathbb{Z}/p\mathbb{Z})^\times} (1-\zeta_p^j)\tau_j(\alpha) \in (1-\zeta_p)\mathcal{O}_K \overset{(1)}{\Longrightarrow} Tr(\alpha(1-\zeta_p)) \in p\mathbb{Z}.$$

□

Proposition 22.3.2. *Let p be an odd prime number and $K = \mathbb{Q}(\zeta_p)$. Then $\mathcal{O}_K = \mathbb{Z}[\zeta_p]$.*

Proof. Let $\alpha = \lambda_0 + \lambda_1\zeta_p + \cdots + \lambda_{p-2}\zeta_p^{p-2} \in \mathcal{O}_K$, with $\lambda_i \in \mathbb{Q}$. We want to show that $\lambda_i \in \mathbb{Z}$ for $0 \le i \le p-2$. We have

$$Tr((1-\zeta_p)\alpha) = Tr(1-\zeta_p)\lambda_0 + (Tr(\lambda_1\zeta_p) - Tr(\lambda_1\zeta_p^2)) + \ldots + (Tr(\lambda_{p-2}\zeta_p^{p-2}) - Tr(\lambda_{p-2}\zeta_p^{p-1}));$$

For $1 \le i \le p-2$ the elements ζ_p^i, ζ_p^{i+1} have the same characteristic polynomial $\Phi_p(X)$ hence they have the same trace. Using Example 21.2.11 we find $Tr((1-\zeta_p)\alpha) = p\lambda_0$. The previous lemma tells us that $Tr((1-\zeta_p)\alpha) \in p\mathbb{Z}$, hence $\lambda_0 \in \mathbb{Z}$. It follows that $\alpha - \lambda_0 \in \mathcal{O}_K$, hence $\lambda_1 + \lambda_2\zeta_p + \cdots + \lambda_{p-2}\zeta_p^{p-3} \in \mathcal{O}_K$. Repeating the argument we find that every λ_i belongs to $\mathbb{Z}$. □

Exercise 22.3.3.
(1) Let K be a number field of degree n and $\Sigma_K = \{\sigma_1, \ldots, \sigma_n\}$ be the set of embeddings of K in $\mathbb{C}$. Let $e_1, \ldots, e_n$ be a basis of the abelian group $\mathcal{O}_K$ and $\mathbf{M}$ the $n \times n$ matrix whose $(i,j)-$th entry is $\sigma_i(e_j)$. Prove that $\Delta(\mathcal{O}_K) = (\det \mathbf{M})^2$.
(2) Prove that $\Delta(\mathbb{Z}[\zeta_p]) = (-1)^{\frac{p-1}{2}}p^{p-2}$.
(3) (Stickelberger) Prove that for every number field K with ring of integers $\mathcal{O}_K$ the congruence

$$\Delta(\mathcal{O}_K) \equiv 0, 1 \pmod{4}$$

holds.

22.4. **Algebraic properties of $\mathcal{O}_K$.** Let us come back to the general theory: let K be a number field and $\mathcal{O}_K$ its ring of integers. The following proposition generalises what we have learned about quadratic rings (cf. Lemma 11.2.3, Proposition 16.1.3).

Proposition 22.4.1.
(1) *Let $I \subset \mathcal{O}_K$ be a non-zero ideal. Then the quotient $\mathcal{O}_K/I$ is finite.*
(2) *Every ideal $I \subsetneq \mathcal{O}_K$ is contained in a maximal ideal.*
(3) *$\mathcal{O}_K$ is noetherian.*
(4) *Every non-zero prime ideal of $\mathcal{O}_K$ is maximal.*
(5) *$\mathcal{O}_K$ is integrally closed.*

Proof. To prove (1), let $\alpha \in I$ be a non-zero element. Then $\frac{N(\alpha)}{\alpha} \in K$ is a product of algebraic integers, hence $\frac{N(\alpha)}{\alpha} \in \mathcal{O}_K$. It follows that $N(\alpha) \in I$, so $\mathcal{O}_K/(N(\alpha)) \twoheadrightarrow \mathcal{O}_K/I$. As $\mathcal{O}_K \simeq \mathbb{Z}^n$ the quotient $\mathcal{O}_K/(N(\alpha))$ is finite, hence $\mathcal{O}_K/I$ is finite.

Given (1), points (2), (3), (4) follow in view of Proposition 16.1.3. It remains to prove (5). Let $\alpha \in K$ and suppose that α is integral over $\mathcal{O}_K$, i. e. there exists $P(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in \mathcal{O}_K[X]$ with $P(\alpha) = 0$. We need to prove that α belongs to $\mathcal{O}_K$, i.e. that α is integral over $\mathbb{Z}$. We have $\mathcal{O}_K[\alpha] = \sum_{i=0}^{n-1} \alpha^i\mathcal{O}_K$ as subgroups of K. We also know that $\mathcal{O}_K$ is a finitely generated abelian group; it follows that $\mathcal{O}_K[\alpha]$ is a finitely generated abelian group. Choose generators $\mathbf{v} = (v_1, \ldots, v_k)$ of $\mathcal{O}_K[\alpha]$. Then there exists a matrix $\mathbf{M}$ with $\mathbb{Z}$-coefficients such that

$$\alpha\mathbf{v} = \mathbf{M} \cdot \mathbf{v};$$

Lemma 20.3.3 implies that α is integral over $\mathbb{Z}$, hence α belongs to $\mathcal{O}_K$. □

Properties (3), (4) and (5) are the crucial algebraic properties of $\mathcal{O}_K$. Rings enjoying these properties are called Dedekind domains.

Definition 22.4.2. *An integral domain A is called a* Dedekind domain *if*
(1) *A is noetherian*

(2) *A is integrally closed.*
(3) *Every non-zero prime ideal of A is maximal.*

Example 22.4.3. A great supply of Dedekind domains which are not rings of integers of number fields comes from geometry: for example, let $A = \mathbb{C}[X,Y]/(Y^2 - X^3 + 2)$. This is the ring of "polynomial functions" on the Mordell curve $Y^2 = X^3 - 2$ (cf. (9.2.5.1)). One can prove that A is a Dedekind domain. On the other hand, the ring $B = \mathbb{C}[X,Y]/(Y^2 - X^3)$ of polynomial functions on the curve $Y^2 = X^3$ is *not* a Dedekind domain, because it is not integrally closed. Indeed, $T = \frac{Y}{X}$ satisfies $T^2 = X$, hence it is integral over B; however $T \notin B$. The issue is that the curve $Y^2 = X^3$ has a *singularity* at the origin $(0,0)$: in other words, the derivatives of $Y^2 - X^3$ with respect to both X and Y vanish at the origin, hence the tangent to the curve is not defined there.

In general, one has the following parallelism between number theory and geometry:

$$\begin{aligned} \text{Ring of integers } \mathcal{O}_K \text{ of a number field } K &\leftrightsquigarrow \text{ Smooth curve } C : P(X,Y) = 0, P \in \mathbb{C}[X,Y] \\ \text{Non-zero prime ideals of } \mathcal{O}_K &\leftrightsquigarrow \text{Points } (x,y) \in \mathbb{C}^2 \mid P(x,y) = 0 \\ \text{Inclusion } \mathbb{Z} \subset \mathcal{O}_K &\leftrightsquigarrow \text{ Projection map } C \to \mathbb{A}_1, (x,y) \mapsto x. \end{aligned}$$

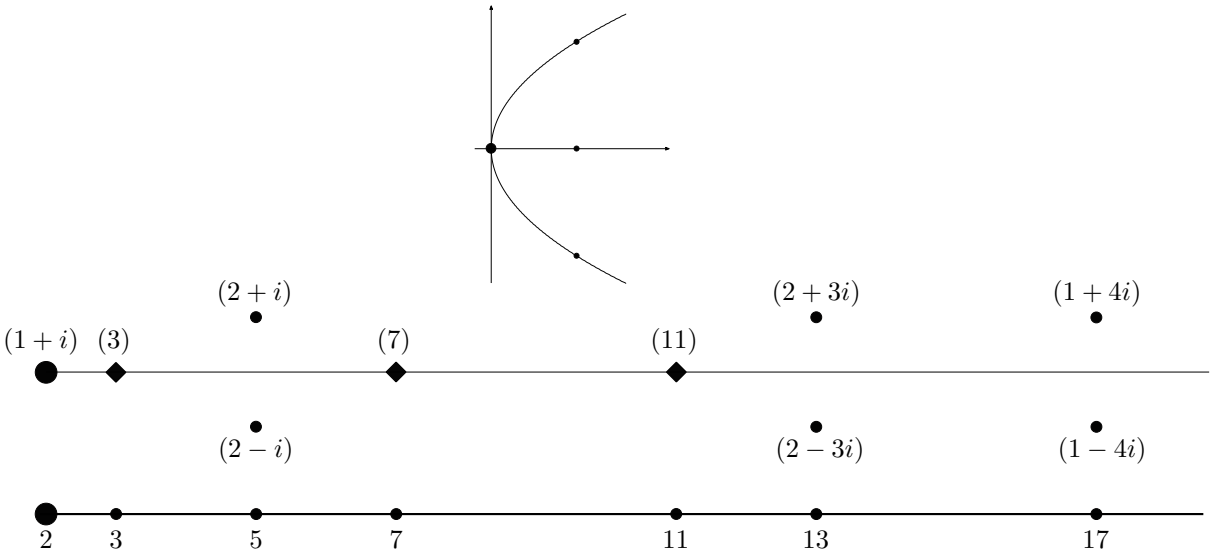


FIGURE 4. $\mathbb{C}[X,Y]/(Y^2 - X)$ vs $\mathbb{Z}[X]/(X^2 + 1)$.

This analogy attracted the attention of countless mathematicians; among several sources, we strongly recommend André Weil's letter to his sister Simone [38]. The above dictionary can nowadays be made quite precise, although some aspects of it remain mysterious.

23. Lecture 23: Finiteness of $Cl(\mathcal{O}_K)$

In this lecture we will prove that the set $Cl(\mathcal{O}_K)$ of fractional ideals of $\mathcal{O}_K$ modulo principal fractional ideals is finite.

23.1. **Properties of fractional ideals.** Let K be a number field with ring of integers $\mathcal{O}_K$. Recall that a *fractional ideal* of $\mathcal{O}_K$ is an additive subgroup I of K such that $\alpha I \subset \mathcal{O}_K$ is a non-zero ideal for some element $\alpha \in K^\times$. Lemma 22.2.1 guarantees that after multiplying α by a non-zero integer we can ensure that $\alpha \in \mathcal{O}_K \smallsetminus \{0\}$ and $\alpha I \subset \mathcal{O}_K$ is a non-zero ideal; this imples that $N(\alpha)I \subset \mathcal{O}_K$ is also a non-zero ideal. Hence if I is a fractional ideal then there is $k \in \mathbb{Z}$ such that $kI \subset \mathcal{O}_K$ is a non-zero ideal.

Every non-zero ideal of $\mathcal{O}_K$ is a fractional ideal, and for every $\alpha \in K^\times$ the group $\alpha\mathcal{O}_K \subset K$ is a fractional ideal (check this, using Lemma 22.2.1), called a *principal* fractional ideal. We denote by $I(\mathcal{O}_K)$ the set of fractional ideals of $\mathcal{O}_K$ and by $P(\mathcal{O}_K)$ the subset of principal fractional ideals. The set $P(\mathcal{O}_K)$ is actually a group, acting on $I(\mathcal{O}_K)$ by the rule $(\alpha) \cdot I = \alpha I$. We set

$$Cl(\mathcal{O}_K) = I(\mathcal{O}_K)/P(\mathcal{O}_K);$$

hence two fractional ideals have the same image in $Cl(\mathcal{O}_K)$ if and only if they differ by multiplication by an element $\alpha \in K^\times$.

Proposition 23.1.1. *Let $I \subset K$ be a fractional ideal of $\mathcal{O}_K$.*

(1) *I is a free abelian group of rank $[K : \mathbb{Q}]$.*
(2) *Let $k \in \mathbb{Z}$ be such that $kI \subset \mathcal{O}_K$ is a non-zero ideal. The quantity*

$$\frac{[\mathcal{O}_K : kI]}{k^{[K:\mathbb{Q}]}}$$

does not depend on the choice of k.

Proof. We know by Proposition 22.4.1 that an ideal $I \subset \mathcal{O}_K$ has finite index in $\mathcal{O}_K$, hence it is a free abelian group of the same rank as $\mathcal{O}_K$. Given an arbitrary fractional ideal I, there exists $\alpha \in K^\times$ such that $\alpha I \subset \mathcal{O}_K$ is a non-zero ideal, hence a free abelian group of rank $[K : \mathbb{Q}]$. Multiplication by α induces an isomorphism of abelian groups $I \simeq \alpha I$, hence $I \simeq \mathbb{Z}^{[K:\mathbb{Q}]}$. This proves (1). Point (2) is proved in the same way as point (3) in Lemma 11.2.3. □

23.1.2. *Norm of a fractional ideal.* Let $I \in I(\mathcal{O}_K)$. The *norm* of I is defined as the quotient

$$N(I) = \frac{[\mathcal{O}_K : kI]}{k^{[K:\mathbb{Q}]}}$$

for any integer k such that $kI \subset \mathcal{O}_K$ is a non-zero ideal. By the previous lemma this is a meaningful definition. If $I \subset \mathcal{O}_K$ is a non-zero ideal we can take $k = 1$ and we find $N(I) = [\mathcal{O}_K : I]$. As in the case of quadratic rings (cf. 11.2.4), the norm function enjoys the following properties:

(1) If $\alpha \in K^\times$ then $N(\alpha\mathcal{O}_K) = |N(\alpha)|$.
(2) For every $\alpha \in K^\times$ and every fractional ideal I,

$$N(\alpha I) = |N(\alpha)|N(I).$$

As in 11.2.4, (2) follows from (1). To show (1), after multiplying by a suitable integer $k \in \mathbb{Z} \smallsetminus \{0\}$ we may assume that $\alpha \in \mathcal{O}_K$, hence we have $N(\alpha\mathcal{O}_K) = [\mathcal{O}_K : \alpha\mathcal{O}_K]$. Because of (2) in 11.1.4 the index $[\mathcal{O}_K : \alpha\mathcal{O}_K]$ is the absolute value of the determinant of the map

$$\begin{aligned} m_\alpha : \mathcal{O}_K &\to \mathcal{O}_K \\ x &\mapsto \alpha x \end{aligned}$$

which is by definition the norm of α.

23.2. **Finiteness of $Cl(\mathcal{O}_K)$.** In this section we will prove the following important result, generalising finiteness of $Cl(\mathcal{O}_d)$ for $d < 0$ a fundamental discriminant.

Theorem 23.2.1. *Let K be a number field. Then $Cl(\mathcal{O}_K) = I(\mathcal{O}_K)/P(\mathcal{O}_K)$ is finite.*

Definition 23.2.2. *The cardinality of $Cl(\mathcal{O}_K)$ is denoted by $h(\mathcal{O}_K)$ (or $h(K)$) and is called the* class number *of $\mathcal{O}_K$ (or of K).*

The proof of the above theorem is the only point where our arguments differ from the case of imaginary quadratic fields: in that setting our proof relied on the relation between ideals and binary quadratic forms and on reduction theory. This does not work in general, so we will give a weaker argument, which however has the benefit of working for arbitrary number fields.

Proposition 23.2.3. *There exists a constant $C > 0$ (depending only on K) with the following property: for every non-zero ideal $I \subset \mathcal{O}_K$ there exists $\alpha \in I \smallsetminus \{0\}$ such that*

$$|N(\alpha)| \leq C \cdot N(I).$$

23.2.4. *Proposition 23.2.3 $\Rightarrow$ Theorem 23.2.1.* Assume Proposition 23.2.3. We want to show that $Cl(\mathcal{O}_K)$ is finite. Let $I \in I(\mathcal{O}_K)$; after multiplying I by a suitable $k \in \mathbb{Z}_{>0}$ we may (and will) assume that $I \subset \mathcal{O}_K$. Let $\alpha \in I \smallsetminus \{0\}$ such that $|N(\alpha)| \leq C \cdot N(I)$. We find

$$\alpha \in I \Rightarrow \mathcal{O}_K \subset \alpha^{-1} I;$$

Furthermore

$$[\alpha^{-1} I : \mathcal{O}_K] = [I : \alpha \mathcal{O}_K] = \frac{[\mathcal{O}_K : \alpha \mathcal{O}_K]}{[\mathcal{O}_K : I]} = |N(\alpha)| N(I)^{-1} \leq C.$$

Hence, letting $d = [\alpha^{-1} I : \mathcal{O}_K]$, we have $d \leq C$ and $d\alpha^{-1} I \subset \mathcal{O}_K$. Letting M be the least common multiple of all the positive integers not larger than C, we deduce that $M(\alpha^{-1} I) \subset \mathcal{O}_K$. Hence we have inclusions

$$\mathcal{O}_K \subset \alpha^{-1} I \subset \frac{1}{M} \mathcal{O}_K.$$

We know that $[\frac{1}{M}\mathcal{O}_K : \mathcal{O}_K] = [\mathcal{O}_K : M\mathcal{O}_K]$ is finite, hence there are finitely many possibilities for $\alpha^{-1} I$. As $\alpha^{-1} I$ and I give the same element in $Cl(\mathcal{O}_K)$ and the number C (hence M) does *not* depend on I, we deduce that $Cl(\mathcal{O}_K)$ is finite.

It remains to prove Proposition 23.2.3; we will make use of the fact that the norm map is a homogeneous polynomial of degree $[K : \mathbb{Q}]$ in $[K : \mathbb{Q}]$ variables. Precisely:

Lemma 23.2.5. *Let $n = [K : \mathbb{Q}]$ and let $(e_1, \ldots, e_n)$ be a $\mathbb{Z}$-basis of $\mathcal{O}_K$. There exists a homogeneous polynomial $P(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$ of degree n such that, for every $(x_1, \ldots, x_n) \in \mathbb{Z}^n$, we have*

$$N(x_1 e_1 + \ldots + x_n e_n) = P(x_1, \ldots, x_n).$$

Proof. For $1 \leq i, j \leq n$ write

$$e_i e_j = \sum_{k=1}^{n} a_{i,j}^k e_k \text{ with } a_{i,j}^k \in \mathbb{Z}.$$

If $\alpha = x_1 e_1 + \ldots + x_n e_n \in \mathcal{O}_K$ and $1 \leq j \leq n$ we have

$$\alpha e_j = x_1 e_1 e_j + \ldots + x_n e_n e_j = x_1 (\sum_{k=1}^{n} a_{1,j}^k e_k) + \ldots + x_n (\sum_{k=1}^{n} a_{n,j}^k e_k) = \sum_{k=1}^{n} (\sum_{i=1}^{n} a_{i,j}^k x_i) e_k.$$

Hence the matrix $\mathbf{M}_\alpha$ of the multiplication by α map in the basis $(e_1, \ldots, e_n)$ has (k, j)-th entry the element $\sum_{i=1}^{n} a_{i,j}^k x_i$. Let $\mathbf{M}$ be the matrix whose (k, j)-th entry is $\sum_{i=1}^{n} a_{i,j}^k X_i$. Then $P(X_1, \ldots, X_n) = \det(\mathbf{M}) \in \mathbb{Z}[X_1, \ldots, X_n]$ satisfies the requirements of the lemma. □

23.2.6. *Proof of Proposition 23.2.3.* Let $I \subset \mathcal{O}_K$ be a non-zero ideal and $n = [K : \mathbb{Q}]$. Fix a basis $e_1, \ldots, e_n$ of $\mathcal{O}_K$ and take $P(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$ as in the previous lemma. Let M be the maximum of the absolute values of the coefficients of P and c the number of non-zero coefficients of P. As P is homogeneous of degree n, for every $(x_1, \ldots, x_n) \in \mathbb{Z}^n$ we have the (naive) estimate

$$|N(x_1 e_1 + \ldots + x_n e_n)| = |P(X_1, \ldots, X_n)| \leq cM \max(|x_i|)^n. \tag{23.2.6.1}$$

Let $m \in \mathbb{Z}_{>0}$ be the unique positive integer such that $m^n \leq N(I) < (m+1)^n$. The set

$$S_m = \{x_1 e_1 + \ldots + x_n e_n, x_1, \ldots, x_n \in \mathbb{Z}, |x_i| \leq m \text{ for } 1 \leq i \leq n\}$$

has at least $(m+1)^n$ distinct elements (as for every $1 \leq i \leq n$ we can choose any $0 \leq x_i \leq m$). As $N(I) = [\mathcal{O}_K : I] < (m+1)^n$ it follows that there are two distinct elements $\beta = b_1 e_1 + \ldots + b_n e_n \in S_m$, $\gamma = c_1 e_1 + \ldots + c_n e_n \in S_m$ such that $\beta \equiv \gamma \pmod{I}$. Therefore $\alpha = \beta - \gamma$ belongs to $I \smallsetminus \{0\}$.

Write $\alpha = a_1e_1 + \ldots + a_ne_n$, where $a_i = b_i - c_i$. As $|b_i|, |c_i| \leq m$ we have $|a_i| \leq 2m$ for $i = 1, \ldots, n$. It follows from (23.2.6.1) that

$$|N(\alpha)| \leq cM(2m)^n \Rightarrow |N(\alpha)| \leq cM2^nN(I).$$

Hence $C = cM2^n$ satisfies the requirement of Proposition 23.2.3.

Remark 23.2.7. The previous argument proves finiteness of $Cl(\mathcal{O}_K)$ using few simple ingredients:

(1) The basic finiteness input: for every non-zero ideal $I \subset \mathcal{O}_K$ the quotient $\mathcal{O}_K/I$ is finite - which in turn follows in the end from finiteness of $\mathbb{Z}/n\mathbb{Z}$ for $n \geq 1$, another manifestation of the key fact 1.2.1.
(2) The fact that the norm map in a chosen basis is a homogeneous polynomial of the coordinates.
(3) The fact that $\mathbb{Z}$ contains *strictly* more than m elements of absolute value at most $m \geq 0$.
(4) The pigeonhole principle.

In fact, such a simple argument can be shown to work for a wider class of rings: see [35], from which we extracted the proof presented above; a similar idea is also used in [21]. The issue with this approach is that it does not give a sharp constant C in Proposition 23.2.3; this is due to our brutal estimate (23.2.6.1). While this does not matter if we only want to prove finiteness of $Cl(\mathcal{O}_K)$, it is important to find a good constant C for computational purposes. Indeed, in 23.2.4 we showed that every fractional ideal I has the same image in $Cl(\mathcal{O}_K)$ as a fractional ideal $J \supset \mathcal{O}_K$ such that $[J : \mathcal{O}_K] \leq C$: so in order to determine $Cl(\mathcal{O}_K)$ it suffices to look for fractional ideals such that $[J : \mathcal{O}_K] \leq C$, and check which of them differ by multiplication by a non-zero constant. In particular, the lower C is the fewer ideals we have to check.

For imaginary quadratic fields we obtained very good estimates by means of reduction theory for quadratic forms (see the computations in Example 16.3.4). For a general number field K, the problem is to minimise the value taken by the homogeneous form expressing the norm in a chosen basis on the set of non-zero elements of a given ideal. One way to do this relies on Minkowski's *geometry of numbers*, which allows to obtain a pretty good constant for arbitrary number fields.

Remark 23.2.8. The main arithmetic properties of number fields we proved - finiteness of the class number and structure of the group of units of real quadratic rings - rely on two important *existence statements*:

(1) Given a number field K, there exists a constant $C > 0$ such that every non-zero ideal $I \subset \mathcal{O}_K$ contains a non-zero element α such that $|N(\alpha)| \leq C \cdot N(I)$.
(2) Given a real quadratic ring $\mathcal{O}_d$ there exists a constant $C > 0$ such that $\mathcal{O}_d$ contains infinitely many elements α with $|N(\alpha)| < C$.

Somewhat curiously, at the end of the day the proof of both statements is an application of the pigeonhole principle (+ the fact that there are $M + 1$ integers between 0 and M). Dirichlet is believed to have been the first to explicitly state the pigeonhole principle in 1834 - though he called it *Schubfachprinzip* (= drawer principle). However at least one concrete example was stated much earlier: in the 1622 book "Selectae Propositiones in Tota Sparsim Mathematica Pulcherrimae" one finds the claim that "Necesse est, duos hominum, habere totidem numero pilos, aureos, et similia".[10]

Exercise 23.2.9. The pigeonhole principle is very basic and easy to understand; much as the key fact 1.2.1, it can be applied in countless contexts.

(1) There are n people in a room ($n > 1$), some of which shake hands to some others (you can't shake your own hand). Then there are at least two people who shake the same number of hands.
(2) Pick $n + 1$ integers between 1 and $2n$; show that you have picked at least two such that one divides the other.
(3) Pick a set S of n positive integers ($n > 1$); show that there is a non-empty subset $T \subset S$ such that n divides the sum of the elements in T.

[10] "It is necessary that two men have the same number of hairs, écus, or other things." This is also assuming that, say, no man has several million hairs. In fact the average number of hairs is around 150 000, so this is a reasonable assumption, and the same claim could be made nowadays about people living in London. This example should convince you of how non-constructive the pigeonhole principle is; the same issue appears in Mathematics, e. g. in the proofs we gave.

23.2.10. **FAQs on class numbers.* Questions on class numbers of number fields are usually very difficult and often very interesting. Here are some examples.

Q: How many imaginary quadratic number fields have class number 1?

A: Those of the form $\mathbb{Q}(i\sqrt{k})$ for k belonging to the following finite list:

$$1, 2, 3, 7, 11, 19, 43, 67, 163.$$

This was conjectured by Gauss and proved by Heegner (a 59 years old high school teacher) in a groundbreaking work in 1952. Heegner's proof was believed to be incorrect, and the result was re-proved in 1966 by Baker and Stark (independently). It was later recognised - after Heegner's death - that Heegner's original proof was essentially correct. You can find a proof of this result in [9].

Q: How does $h(\mathbb{Q}(i\sqrt{k}))$ vary as $k > 0$ grows?

A: It was conjectured by Gauss, and proved by Heilbronn in 1934, that $h(\mathbb{Q}(i\sqrt{k})) \to \infty$ as $k \to \infty$.

Q: What about class numbers of $\mathbb{Q}(\sqrt{k})$ for $k > 0$?

A: Good question. Gauss conjectured that infinitely many real quadratic fields have class number one. Computational evidence suggests that around 70% of real quadratic fields have this property, but Gauss' conjecture is wide open.

Q: How many cyclotomic fields $\mathbb{Q}(\zeta_n)$ have class number 1?

A: Here's the complete list of values of n (see [37, Chapter 11]):

1 to $22, 24, 25, 26, 27, 28, 30, 32, 33, 34, 35, 36, 38, 40, 42, 44, 45, 48, 50, 54, 60, 66, 70, 84, 90$.

Q: How does $h(\mathbb{Q}(\zeta_n))$ vary with n?

A: As $n \to \infty$, the class number of $\mathbb{Q}(\zeta_n)$ goes to infinity (see [37, Chapter 4]). Restricting to $n = p^k$ for p an odd prime, we have the following remarkable theorem due to Iwasawa (see [37, Chapter 13]): for $k \geq 1$ let p^{e_k} be the exact power of p dividing $h(\mathbb{Q}(\zeta_{p^k}))$. There exist integers $\lambda, \mu \in \mathbb{Z}_{\geq 0}$ and $\nu \in \mathbb{Z}$, and a positive integer k_0, such that

$$\forall\, k \geq k_0, \;\; e_k = \lambda k + \mu p^k + \nu.$$

Furthermore, Ferrero and Washington proved that $\mu = 0$ [37, Chapter 7].

Q: Are there infinitely many number fields with class number 1?

A: Nobody knows.

24. Lecture 24: Problem session IV

(1) Let $d > 0$ be an integer which is not a square and is congruent to $0, 1 \pmod 4$; let $\mathcal{O}_d \subset \mathbb{R}$ be the quadratic ring of discriminant d. Prove that there exists a *unique* unit $\varepsilon_d \in \mathcal{O}_d^\times$ such that $\varepsilon_d > 1$ and every element of $\mathcal{O}_d^\times$ can be uniquely written as $\pm\varepsilon_d^n$ with $n \in \mathbb{Z}$. The unit ε_d is called the *fundamental unit* of $\mathcal{O}_d$. The aim of this exercise is to prove the following

Proposition. *The fundamental unit is of the form* $\frac{a+b\sqrt{d}}{2}$ *where* (a, b) *is the couple of positive integers with smallest possible first element solving one of the equations* $X^2 - dY^2 = \pm 4$.

(a) Show that ε_d is the largest of the numbers $\pm\varepsilon_d$, $\pm(\varepsilon_d)^{-1}$.

(b) Assume that $d \equiv 0 \pmod 4$. Write $\varepsilon_d = \frac{a+b\sqrt{d}}{2}$ with $a, b \in \mathbb{Z}$ and a even. Show that a, b are positive integers solving the equation $X^2 - dY^2 = \pm 4$.

(c) For $k > 1$ write $\varepsilon_d^k = \frac{a_k+b_k\sqrt{d}}{2}$; prove that $a_k > a, b_k > b$. Deduce that (a, b) is the couple of positive integers with smallest possible first element solving one of the equations $X^2 - dY^2 = \pm 4$.

(d) Now suppose that $d \equiv 1 \pmod 4$. Write $\varepsilon_d = \frac{a+b\sqrt{d}}{2}$ with $a, b \in \mathbb{Z}$ and $a \equiv b \pmod 2$. Show that (a, b) is the couple of positive integers with smallest possible first coordinate solving one of the equations $X^2 - dY^2 = \pm 4$.

(2) The criterion found in the previous exercise allows in principle to find the fundamental unit of a quadratic ring $\mathcal{O}_d$. Use it to determine the fundamental units of the following rings: $\mathbb{Z}[\sqrt{2}], \mathbb{Z}[\sqrt{3}], \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$ and $\mathbb{Z}[\sqrt{5}]$.

(3) Let K be a number field with ring of integers $\mathcal{O}_K$. You may assume in this exercise the following facts:

- Multiplication of fractional ideals turns $Cl(\mathcal{O}_K)$ into an abelian group.
- Non-zero ideals of $\mathcal{O}_K$ factor uniquely as products of prime ideals.
- For $I, J \subset \mathcal{O}_K$ non-zero ideals we have $N(IJ) = N(I)N(J)$.

The first two facts follow from finiteness of $Cl(\mathcal{O}_K)$ as in the case of rings of integers of quadratic fields, and will be proved in the next lecture; for the third fact, see Exercise 26.1.2.

(a) Show that there exists a constant $C_K > 0$ only depending on K such that every element in $Cl(\mathcal{O}_K) = I(\mathcal{O}_K)/P(\mathcal{O}_K)$ has a representative in $I(\mathcal{O}_K)$ which is an ideal of $\mathcal{O}_K$ whose norm is at most C_K.

(b) Deduce that $Cl(\mathcal{O}_K)$ is generated by principal ideals and prime ideals of norm at most C_K.

(c) Show that if K is imaginary quadratic then we can take $C_K = \sqrt{\frac{|\Delta(\mathcal{O}_K)|}{3}}$; prove that this value is optimal, i. e. there exists an imaginary quadratic number field K for which no smaller value of C_K works.

(d) Let K be a number field of degree n. Let r_2 be the number of pairs of complex conjugate embeddings $\tau : K \hookrightarrow \mathbb{C}$ whose image is not contained in $\mathbb{R}$. Using geometry of numbers one can prove that the following constant works:

$$C_K = \sqrt{|\Delta(\mathcal{O}_K)|}\left(\frac{4}{\pi}\right)^{r_2}\frac{n!}{n^n} \text{ (Minkowski's constant).}$$

Using this fact prove that $\frac{2}{\pi} \geq \frac{1}{\sqrt{3}}$.

(e) Show that the Minkowski constant must be at least one. Deduce that for every number field $\mathcal{O}_K$ we have $|\Delta(\mathcal{O}_K)| > 1$. (you can use the following estimate: for $n \geq 2$ we have $\left(\frac{\pi}{4}\right)^{\frac{n}{2}}\frac{n^n}{n!} > 1$).

(4) Let $A = \mathbb{Z}[\sqrt{82}]$. In this exercise we will determine the group of units and the class group of A.

(a) Show that $9 + \sqrt{82}$ is a unit in A and that every unit in A is of the form $\pm(9 + \sqrt{82})^k$ for some $k \in \mathbb{Z}$.

(b) Show that $Cl(A)$ is generated by prime ideals of norm at most 9 (you can use Minkowski's constant).

(c) Show that $(2) = (2, \sqrt{82})^2$ and $(3) = (3, 1 + \sqrt{82})(3, 1 - \sqrt{82})$. Let $\mathfrak{p}_2 = (2, \sqrt{82})$ and $\mathfrak{p}_3 = (3, 1 + \sqrt{82})$.
(d) Show that (5) and (7) are prime ideals of A. Deduce that $Cl(A)$ is generated by $\mathfrak{p}_2$ and $\mathfrak{p}_3$.
(e) Verify that $\mathfrak{p}_2\mathfrak{p}_3^2 = (10 + \sqrt{82})$. Deduce that $Cl(A)$ is a cyclic group generated by $\mathfrak{p}_3$, which has order dividing 4.
(f) Finally, we wish to show that $\mathfrak{p}_3$ has order precisely 4 in $Cl(A)$, so that $Cl(A) \simeq \mathbb{Z}/4\mathbb{Z}$. Equivalently, we have to prove that $\mathfrak{p}_2$ is not principal. Assume by contradiction that $\mathfrak{p}_2$ is principal. Show that there exist $a, b \in \mathbb{Z}$ such that $2 = (a + b\sqrt{82})^2$, and find a contradiction.
(g) Show that the Diophantine equation $X^2 - 82Y^2 = \pm 2$ has no integral solution.[11]

(5) Let K be a number field and $\alpha \in \mathcal{O}_K$ such that $|\tau(\alpha)| \leq 1$ for every $\tau \in \Sigma_K$. Prove that α is a root of unity.

(6) Let K and L be two number fields embedded in $\mathbb{C}$; we will denote by KL the smallest subfield of $\mathbb{C}$ containing both K and L.
(a) Show that KL is a number field; if K has degree m and L has degree n prove that KL has degree at most mn.
(b) Let $\mathcal{O}_K$ be the ring of integers of K and $\mathcal{O}_L$ the ring of integers of L. Let

$$\mathcal{O}_K\mathcal{O}_L = \{\alpha_1\beta_1 + \ldots + \alpha_r\beta_r \mid r \geq 1, \alpha_i \in \mathcal{O}_K, \beta_i \in \mathcal{O}_L\} \subset KL.$$

Prove that $\mathcal{O}_K\mathcal{O}_L \subset \mathcal{O}_{KL}$. We will now prove the following result.
Proposition. *With the above notations, assume that KL has degree mn and that the discriminants $\Delta(\mathcal{O}_K)$ and $\Delta(\mathcal{O}_L)$ are coprime. Then $\mathcal{O}_K\mathcal{O}_L = \mathcal{O}_{KL}$.*
(c) Using the previous proposition find the ring of integers of $\mathbb{Q}(i, i\sqrt{5})$.
(d) We will now prove the proposition; from now on all the assumptions in the proposition are in force. Prove that the natural map $\Sigma_{KL} \to \Sigma_K \times \Sigma_L$ induced by restriction of embeddings is a bijection.
(e) Let $(e_1, \ldots, e_m)$ be a $\mathbb{Z}$-basis of $\mathcal{O}_K$ and $(f_1, \ldots, f_n)$ a $\mathbb{Z}$-basis of $\mathcal{O}_L$. Prove that $(e_if_j)_{1\leq i\leq m, 1\leq j\leq n}$ is a $\mathbb{Z}$-basis of $\mathcal{O}_K\mathcal{O}_L$.
(f) Let $\alpha = \sum_{1\leq i\leq m, 1\leq j\leq n} \frac{a_{i,j}}{d} e_if_j$ with $a_{i,j}, d \in \mathbb{Z}$ and no factor of d divides all the integers $a_{i,j}$. We wish to show that $d = 1$. Take $\tau \in \Sigma_K$. Show that τ can be extended to an embedding $\sigma \in \Sigma_{KL}$ such that $\Sigma_{|L} = Id$. Deduce that

$$\sigma(\alpha) = \sum_{1\leq i\leq m, 1\leq j\leq n} \frac{a_{i,j}}{d} \sigma(e_i) f_j.$$

(g) Let $x_i = \sum_{1\leq j\leq n} \frac{a_{i,j}}{d} f_j$. Write $\Sigma_K = \{\tau_1, \ldots, \tau_m\}$; choose a lift $\sigma_l \in \Sigma_{KL}$ of each τ_l which restricts to the identity on L and let $\mathbf{M}$ be the matrix whose $(l, i)-$the entry is $\sigma_l(e_i)$. Show that

$$\begin{pmatrix} \sigma_1(\alpha) \\ \vdots \\ \sigma_m(\alpha) \end{pmatrix} = \mathbf{M} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

and that $\det(\mathbf{M})^2 = \Delta(\mathcal{O}_K)$.
(h) Show that $\Delta(\mathcal{O}_K)x_i$ is an algebraic integer for every i (hint: Cramer's rule). Deduce that $\Delta(\mathcal{O}_K)x_i \in \mathcal{O}_L$.
(i) Deduce that $d \mid \Delta(\mathcal{O}_K)$. Using a symmetric argument conclude that $d \mid \Delta(\mathcal{O}_L)$ hence that $d = 1$.
(j) Show that the ring of integers of $\mathbb{Q}(\zeta_{15})$ is $\mathbb{Z}[\zeta_{15}]$. Prove that $1 - \zeta_{15}$ is a unit in $\mathbb{Z}[\zeta_{15}]$.

[11] However, this equation has a solution modulo n for every integer $n \geq 1$ (try to prove it!).

25. Lecture 25: Group structure on $Cl(\mathcal{O}_K)$ and unique factorisation of ideals

In this lecture we will use the fact that $Cl(\mathcal{O}_K)$ is finite to show that this set has a group structure. We will then deduce that ideals in $\mathcal{O}_K$ can be factored uniquely as a product of prime ideals. Nothing new under the sun here: the arguments are borrowed from our previous discussion for quadratic rings with fundamental discriminant.

25.1. Group structure on $Cl(\mathcal{O}_K)$.

Theorem 25.1.1. *Let K be a number field and $\mathcal{O}_K$ its ring of integers. The composition law $I \cdot J = IJ$ endows $Cl(\mathcal{O}_K) = I(\mathcal{O}_K)/P(\mathcal{O}_K)$ with the structure of an abelian group, with identity element the class of principal fractional ideals.*

Proof. We copy the proof of Theorem 15.2.2. As $I(JL) = (IJ)L$ and $IJ = JI$ for every $I, J, L \in I(\mathcal{O}_K)$ the composition law is associative and commutative. For $\alpha \in K^\times$ and $I \in I(\mathcal{O}_K)$ the fractional ideals αI and I have the same image in $Cl(\mathcal{O}_K)$, hence the class of principal fractional ideals is the identity element in $Cl(\mathcal{O}_K)$.

It remains to show that every fractional ideal I has an inverse in $Cl(\mathcal{O}_K)$. Because $Cl(\mathcal{O}_K)$ is finite we know that there exist integers $b > a > 0$ and an element $\alpha \in K^\times$ such that $\alpha I^b = I^a$. Let $J = \alpha I^{b-a}$; it satisfies $JI^a = I^a$. We claim that $J = \mathcal{O}_K$.

Choose a basis $(i_1, \dots, i_n)$ of I^a as a $\mathbb{Z}$-module. Then the inclusion $JI^a \subset I^a$ implies that, for every $\beta \in J$, there exist an $n \times n$-matrix $\mathbf{M}_\beta$ with $\mathbb{Z}$-coefficients such that

$$\beta(i_1, \dots, i_n)^t = \mathbf{M}_\beta (i_1, \dots, i_n)^t.$$

Lemma 20.3.3 implies that $\beta \in \mathcal{O}_K$; hence $J \subset \mathcal{O}_K$.

Conversely, as $I^a \subset JI^a$ there exists a $n \times n$ matrix $\mathbf{M}_J$ with coefficients in J such that

$$(i_1, \dots, i_n)^t = \mathbf{M}_J (i_1, \dots, i_n)^t;$$

It follows that $0 = \det(\mathbf{Id} - \mathbf{M}_J) \in 1 + J$, hence $1 \in J$ and $\mathcal{O}_K \subset J$.

We have proved that $\alpha I^{b-a} = \mathcal{O}_K$; it follows that

$$I^{b-a-1} \cdot I = (\alpha^{-1})$$

hence (the class of) I^{b-a-1} is the inverse of (the class of) I in $Cl(\mathcal{O}_K)$. □

Definition 25.1.2. *The group $Cl(\mathcal{O}_K)$ is called the* ideal class group *of $\mathcal{O}_K$ (or of K).*

25.2. Unique factorisation of ideals.

Theorem 25.2.1. *Let $\mathcal{O}_K$ be the ring of integers of a number field K and $I \subset \mathcal{O}_K$ a non-zero ideal. Then:*

1. *There exists a fractional ideal I^{-1} such that $II^{-1} = \mathcal{O}_K$.*
2. *If $J \subset \mathcal{O}_K$ is an ideal, then*
$$I \mid J \Leftrightarrow I \supset J \qquad \textit{Memento: "to contain is to divide"}$$
3. *Let J, J' be ideals such that $JI = J'I$. Then $J = J'$ (cancellation property).*
4. *There exist non-zero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that*
$$I = \mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_r.$$
Furthermore, if $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ are prime ideals such that
$$\mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_r = I = \mathfrak{q}_1 \mathfrak{q}_2 \cdots \mathfrak{q}_s$$
then $r = s$ and, up to reordering the $\mathfrak{q}_i$'s, we have $\mathfrak{q}_i = \mathfrak{p}_i$ for $i = 1, \dots, r$.

Proof. Same as the proof of Theorem 16.2.1. Let us quickly recall the argument for the reader's convenience.

1. As $Cl(\mathcal{O}_K)$ is a group, there exists a fractional ideal $I' \subset \mathcal{O}_K$ such that $II' = \alpha \mathcal{O}_K$, hence $I^{-1} = \alpha^{-1} I'$ does the job.
2. $J = II'$ for some ideal $I' \subset \mathcal{O}_K \Rightarrow J \subset I\mathcal{O}_K = I$;
$J \subset I \Rightarrow JI^{-1} \subset II^{-1} = \mathcal{O}_K \Rightarrow I' = JI^{-1}$ is an ideal of $\mathcal{O}_K$ s. t. $I'I = J$.

(3) If $JI = J'I$ then multiplying by I^{-1} we find $J = J'$.
(4) Let us prove that non-zero ideal $I \subset \mathcal{O}_K$ can be written as a product of non-zero prime ideals. If $I = \mathcal{O}_K$ or I is maximal, we are done. Otherwise Proposition 22.4.1 tells us that there exists a maximal ideal $\mathfrak{p}_1 \supsetneq I$. By (2) we have $\mathfrak{p}_1 \mid I$, i. e. $I = \mathfrak{p}_1 I_1$ for some ideal $I_1 \subsetneq \mathcal{O}_K$. By the cancellation property we must have $I \subsetneq I_1$. If I_1 is maximal we are done; otherwise iterate the process. As $\mathcal{O}_K$ is Noetherian we must stop at some point, obtaining the sought-for factorisation.

Let us now prove uniqueness of factorisation; assume that

$$\mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_r = I = \mathfrak{q}_1 \cdot \mathfrak{q}_2 \cdots \mathfrak{q}_s \tag{25.2.1.1}$$

Then $\mathfrak{p}_1 \supset \mathfrak{q}_1 \cdots \mathfrak{q}_s$ hence $\mathfrak{p}_1$ contains one of the $\mathfrak{q}_i$'s. We may assume after reordering that $\mathfrak{p}_1 \supset \mathfrak{q}_1$; hence $\mathfrak{p}_1 = \mathfrak{q}_1$ as both ideals are maximal. By (3) we can delete $\mathfrak{p}_1$ on both sides of (25.2.1.1); we then iterate the argument.

□

Corollary 25.2.2. *Let $I \in I(\mathcal{O}_K)$ be a fractional ideal. Then there exist unique* distinct *maximal ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r \subset \mathcal{O}_K$ and integers $k_1, \ldots, k_r \in \mathbb{Z}$ such that*[12]

$$I = \mathfrak{p}_1^{k_1} \cdots \mathfrak{p}_r^{k_r}.$$

Proof. Exercise. □

Corollary 25.2.3. *Let K be a number field and $n \in \mathbb{Z}_{>0}$. If $\gcd(n, h(\mathcal{O}_K)) = 1$ then $\mathcal{O}_K$ satisfies $SP(n)$.*

Proof. As in Theorem 16.3.1: if $\alpha, \beta \in \mathcal{O}_K \smallsetminus \{0\}$ satisfy $(\alpha, \beta) = (1)$ and $\alpha\beta = \gamma^n$ for some $\gamma \in \mathcal{O}_K$ then unique factorisation of ideals plus the fact that $(\alpha, \beta) = (1)$ imply that $(\alpha) = \mathfrak{a}^n$ and $(\beta) = \mathfrak{b}^n$ for some ideals $\mathfrak{a}, \mathfrak{b} \subset \mathcal{O}_K$. As $\gcd(n, h(\mathcal{O}_K)) = 1$ the group $Cl(\mathcal{O}_K)$ does not contain elements of order dividing n other than the identity element. It follows that $\mathfrak{a}, \mathfrak{b}$ are principal, proving the corollary. □

Exercise 25.2.4. Prove that the ring $A \subset \mathbb{C}$ consisting of *all* algebraic integers is a Bézout domain. Here are the main steps:

(1) Let $\alpha, \beta \in \mathbb{C}$ be algebraic integers. We want to prove that the ideal $(\alpha, \beta) \subset A$ is principal. Let $K \subset \mathbb{C}$ be the smallest field containing α, β. Prove that K is a number field.
(2) Let $I = (\alpha, \beta) \subset \mathcal{O}_K$. Deduce that there there exist $k \geq 1$ and $\gamma \in \mathcal{O}_K$ such that $I^k = (\gamma)$.
(3) Pick $\delta \in \mathbb{C}$ such that $\delta^k = \gamma$. Let $L = K(\delta)$; prove that $(\alpha, \beta)\mathcal{O}_L = \delta\mathcal{O}_L$. Conclude.

Exercise 25.2.5. Let K be a number field with ring of integers $\mathcal{O}_K$.

(1) Prove that $\mathcal{O}_K$ is a UFD if and only if it is a principal ideal domain (i. e., if and only if the class number of $\mathcal{O}_K$ equals one).
(2) Prove that every ideal of $\mathcal{O}_K$ is generated by at most two elements (Hint: pick a random non-zero element in the ideal, then use the Chinese remainder theorem to choose the second one).

25.3. **Regular primes and Fermat last theorem.** As a special case of Corollary 25.2.3 we obtain:

Corollary 25.3.1. *Let p be a prime such that $p \nmid h(\mathbb{Z}(\zeta_p))$. Then $\mathbb{Z}[\zeta_p]$ satisfies $SP(p)$.*

Odd primes p such that $p \nmid h(\mathbb{Z}(\zeta_p))$ are called *regular primes*. For these primes, Kummer was able to show that the Fermat equation $X^p + Y^p = Z^p$ has no integer solutions such that $XYZ \neq 0$. The previous corollary is of course a key step in the argument, and it is the reason why it only works for regular primes. We will now discuss the following weak form of Kummer's result, known as the *first case of Fermat's last theorem*:

$$x^p + y^p = z^p, \quad x, y, z \in \mathbb{Z}, \quad p \text{ regular } \Rightarrow p \mid xyz.$$

Let us perform the first step in the proof:

Lemma 25.3.2. *Let p be a prime and $x, y, z \in \mathbb{Z}$ such that $x^p + y^p = z^p$, $\gcd(x, y) = 1$ and $p \nmid z$. If $0 \leq i, j \leq p - 1$ and $i \neq j$ then $(x + \zeta_p^i y, x + \zeta_p^j y) = \mathbb{Z}[\zeta_p]$.*

[12]Note for the pedantic reader: $r = 0$ is allowed, in which case the empty product of ideals is the ideal (1).

Proof. Fix $0 \leq i < j \leq (p-1)$ and let $I = (x+\zeta_p^i y, x+\zeta_p^j y) \subset \mathbb{Z}[\zeta_p]$. Then I contains $\zeta_p^i(1-\zeta_p^{j-i})y$ hence, by Lemma 22.3.1, it contains $(1-\zeta_p)y$.

On the other hand I also contains $\zeta_p^{j-i}(x+\zeta_p^i y) - (x+\zeta_p^j y) = x(\zeta_p^{j-i}-1)$, hence it contains $x(\zeta_p - 1)$. Hence $(x(1-\zeta_p), y(1-\zeta_p)) \subset I$. As $\gcd(x,y) = 1$ we deduce that $(1-\zeta_p) \subset I$. The ideal $(1-\zeta_p)$ is maximal, as $\mathbb{Z}[\zeta_p]/(1-\zeta_p) \simeq \mathbb{Z}[X]/(\Phi_p(X), 1-X) \simeq \mathbb{F}_p$. It follows that either $I = (1-\zeta_p)$ or $I = (1)$.

If $I = (1-\zeta_p)$ then $x^p + y^p \equiv 0 \pmod{1-\zeta_p}$, hence $(1-\zeta_p) \mid z$, i.e. $z \in (1-\zeta_p)\mathbb{Z}[\zeta_p] \cap \mathbb{Z}$. It follows from Lemma 22.3.1 that $p \mid z$, contradiction. Hence $I = (1)$, as we wanted to prove. □

It follows from the previous lemma and from Corollary 25.3.1 that, if p is *regular* and $x^p+y^p = z^p$ with $\gcd(x,y) = 1$ and $p \nmid xyz$, then there exist $u \in \mathbb{Z}[\zeta_p]^\times$ and $\alpha \in \mathbb{Z}[\zeta_p]$ such that $x + \zeta_p y = u\alpha^p$. We will see in the last problem session how to deduce a contradiction from this equality; we refer the interested reader to [37, Chapter 9] for an explanation of how to remove the assumption that $p \nmid xyz$.

25.3.3. **Regular primes, Bernoulli numbers and the Riemann zeta function.* In view of our previous results, it is natural to wonder how many primes are regular. The first bad news is that not all primes are regular: the first counterexamples are $37, 59, 67, 101...$ In fact, it is currently not known whether there are infinitely many regular primes; on the other hand, it has been proved that there are infinitely many primes which are *not* regular. Therefore in a sense Kummer's approach to the Fermat equation is the wrong one. However, it is hard to overestimate how far-reaching the consequences of the ideas developed by Kummer are! We have already explained how Kummer's ideas gave birth to Algebraic Number Theory. Furthermore, Kummer also looked for a criterion to determine whether a given prime p is regular, and he came up with the following discovery:

(25.3.3.1) p is regular $\Leftrightarrow p$ does not divide the numerator of B_k for any $k \in \{2, 4, 6, \ldots, p-3\}$;

the numbers B_k are *Bernoulli numbers*, defined by the formula

$$\frac{t}{e^t - 1} = \sum_{n \geq 0} B_n \frac{t^n}{n!}.$$

Let $\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}$ be the Riemann zeta function. Euler computed

$$\forall\ k \geq 1,\ \zeta(2k) = (-1)^{k-1} B_{2k} \frac{2^{2k-1}}{(2k)!}\pi^{2k} \quad \text{e. g. for } k = 1,\ \zeta(2) = B_2 \cdot \pi^2 = \frac{\pi^2}{6}.$$

Something is going on here! In view of (25.3.3.1), Euler's formula seems to suggest the existence of a relation between the arithmetic of cyclotomic fields $\mathbb{Q}(\zeta_p)$ and values of the Riemann zeta function. In fact Herbrand and Ribet proved that each $B_k, 0 < k < p-1$, controls the order of a precise part of (the p-part of) the class group of $\mathbb{Q}(\zeta_p)$. This story turns out to be related with Iwasawa's result mentioned in 23.2.10, and such a relation led to deep discoveries by Mazur, Wiles, Kolyvagin, Thaine and Rubin... [37, Chapter 15].

25.4. ***Algebraic properties of Dedekind domains.** The proof we gave of the main properties of rings of integers of number fields relied crucially on *finiteness results*:

$$\text{Finiteness of } Cl(\mathcal{O}_K) \rightsquigarrow \text{ Group structure on } Cl(\mathcal{O}_K) \rightsquigarrow \text{ Unique factorisation of ideals.}$$

This approach was proposed by Hurwitz, and has the benefit of making proofs quite easy and highlighting the importance of finiteness of the class group, which is a key arithmetic property of number fields. However the drawback of this method is that it does *not* work for several interesting rings: for example, in view of Example 22.4.3, one may wonder whether the ring $B = \mathbb{C}[X,Y]/(Y^2 - X^3 + 2)$ enjoys similar properties as rings of integers of number fields. It turns out that $Cl(B)$ is *not* finite; however, it is a group, and unique factorisation of ideals in B holds. This is a consequence of the following result, which also applies to rings of integers of number fields because of Proposition 22.4.1.

Theorem 25.4.1. *Let A be a Dedekind domain with fraction field K. Then*

1. *Every fractional ideal $I \subset K$ is invertible, i.e. there exists a fractional ideal J such that $IJ = (\alpha)$ for some $\alpha \in K^\times$.*
2. *The quotient $Cl(A) = \{\text{fractional ideals of } A\}/\{\text{principal fractional ideals}\}$ is a group.*

(3) *Every non-zero ideal $I \subset A$ can be written as a product of prime ideals, unique up to reordering.*

Proof. The key point is to prove (1): once we have it, (2) follows immediately, and then the arguments we used for rings of integers of number fields work in general to prove unique factorisation of ideals. For completeness let us list the main steps in the proof of (1); the reader can fill in the details as an exercise, or see [21, Chapter 3].

(1) We may assume that $I \subset A$ is a non-zero ideal. Pick $0 \neq \alpha \in I$ and set
$$J = \{\beta \in A \mid \beta I \subset (\alpha)\};$$
then $J \subset A$ is a non-zero ideal such that $JI \subset (\alpha)$. We wish to show that equality holds.

(2) Let $L = \alpha^{-1}IJ \subset A$. Assume that $L \subsetneq A$. We claim that if this is the case then there exists $\gamma \in K \smallsetminus A$ such that $\gamma L \subset A$. Assume the claim for the moment. Then, as $J \subset L$, we deduce that $\gamma J \subset \gamma L \subset A$; as $\gamma IJ \subset (\alpha)$ we find $\gamma J \subset J$.

(3) Choose a finite set $(j_1, \dots, j_n)$ of generators of the ideal J. The inclusion $\gamma J \subset J$ implies that there is a square matrix $\mathbf{M}$ with A-coefficients such that
$$(j_1, \dots, j_n)\gamma = (j_1, \dots, j_n)\mathbf{M}$$
which implies that γ is integral over A by Lemma 20.3.3. As A is integrally closed we deduce that $\gamma \in A$, contradicting the fact that $\gamma \in K \smallsetminus A$.

(4) It remains to prove the following statement: if A is a Dedekind domain with fraction field K and $I \subsetneq A$ is a proper ideal then there exists $\gamma \in K \smallsetminus A$ such that $\gamma I \subset A$.

To prove this, let $\alpha \in I$ be a non-zero element (if $I = 0$ the statement is clearly true). First of all, by the exercise below the ideal (α) contains a product of maximal ideals. Take $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ such that $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (\alpha)$ and r is minimal. Since I is a proper ideal of A it is contained in a maximal ideal, which must be equal to one of the $\mathfrak{p}_i$'s. Say $\mathfrak{p}_1 \supset I \supset (\alpha) \supset \mathfrak{p}_1 \cdots \mathfrak{p}_r$. By minimality of r we have $(\alpha) \not\supset \mathfrak{p}_2 \cdots \mathfrak{p}_r$ so we can take $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_r \smallsetminus (\alpha)$. Then we find $\gamma = \frac{\beta}{\alpha} \in K \smallsetminus A$ and $\beta I \subset \mathfrak{p}_2 \cdots \mathfrak{p}_r \mathfrak{p}_1 \subset (\alpha)$ hence $\gamma I \subset A$. □

Exercise 25.4.2. Let A be a noetherian ring. Prove that every non-zero ideal of A contains a product of non-zero prime ideals (Hint: assume that the claim is false, take a counterexample which is maximal with respect to inclusion and obtain a contradiction).

26. Lecture 26: Factorisation of primes in quadratic and cyclotomic fields

In this lecture we study how prime numbers decompose into a product of prime ideals in the ring of integers of a number field; we will examine in detail the case of the rings of integers of quadratic and cyclotomic fields.

26.1. **Split, inert and ramified primes.** Let K be a number field and $n = [K : \mathbb{Q}]$. Let p be a prime number. Then the ideal $(p) = p\mathcal{O}_K$ can be written uniquely as

$$(p) = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2}\dots\mathfrak{p}_d^{e_d} \tag{26.1.0.1}$$

for some $\mathfrak{p}_1, \dots, \mathfrak{p}_d$ distinct non-zero prime ideals of $\mathcal{O}_K$. In particular $\mathfrak{p}_i^{e_i} + \mathfrak{p}_j^{e_j} = (1)$ if $i \neq j$. We say that $\mathfrak{p}_1, \dots, \mathfrak{p}_d$ are the prime ideals *lying above* p. If $\mathfrak{p}_i$ lies above p then $\mathfrak{p}_i \cap \mathbb{Z}$ is an ideal of $\mathbb{Z}$ containing $p\mathbb{Z}$. As $1 \notin \mathfrak{p}_i$ we deduce that $\mathfrak{p}_i \cap \mathbb{Z} = p\mathbb{Z}$. Therefore the inclusion $\mathbb{Z} \subset \mathcal{O}_K$ induces an inclusion of finite fields

$$\mathbb{F}_p \subset \mathcal{O}_K/\mathfrak{p}_i.$$

Definition 26.1.1. *Let K be a number field, p a prime number and $\mathfrak{p} \subset \mathcal{O}_K$ a prime ideal lying above p.*

1. *The* inertia degree *of $\mathfrak{p}$, denoted by $f_\mathfrak{p}$, is the dimension of $\mathcal{O}_K/\mathfrak{p}$ as an $\mathbb{F}_p$-vector space.*
2. *The* ramification index *of $\mathfrak{p}$, denoted by $e_\mathfrak{p}$, is the greatest among the positive integers e such that $\mathfrak{p}^e \mid (p)$. In other words, it is the exponent of $\mathfrak{p}$ in the factorisation* (26.1.0.1).
3. *We say that p is* unramified *in K if the ramification index of every prime ideal above p is 1. Equivalently, p is unramified in K if $p\mathcal{O}_K$ factors as a product of distinct prime ideals. A prime which is not unramified is called* ramified.
4. *We say that p* splits *(or* is totally split*) in K if $p\mathcal{O}_K$ factors as a product of $[K : \mathbb{Q}]$ distinct prime ideals.*
5. *We say that p is* inert *in K if $p\mathcal{O}_K$ is a prime ideal.*

Let p be a prime number; factor (p) as in (26.1.0.1). It follows from the Chinese remainder theorem 8.1.2 that

$$\mathcal{O}_K/(p) \simeq \mathcal{O}_K/\mathfrak{p}_1^{e_1} \times \mathcal{O}_K/\mathfrak{p}_2^{e_2} \times \dots \times \mathcal{O}_K/\mathfrak{p}_d^{e_d}.$$

As the left hand side is an $\mathbb{F}_p$-vector space of dimension n, we see that:

1. For $1 \leq i \leq d$ the inertia degree of $\mathfrak{p}_i$ is at most n. Equality occurs if and only if p is inert.
2. We have $d \leq n$, and equality holds if and only if p is totally split. In this case $e_\mathfrak{p} = f_\mathfrak{p} = 1$ for every $\mathfrak{p} \mid (p)$.
3. A prime p is unramified in K if and only if the ring $\mathcal{O}_K/(p)$ has no non-zero *nilpotent elements* (recall: if A is a ring and $a \in A$ then a is nilpotent if there exists $k > 0$ such that $a^k = 0$). A ring without non-zero nilpotent elements is called *reduced*. Hence we find that

$$p \text{ is unramified in } K \Leftrightarrow \text{ the ring } \mathcal{O}_K/(p) \text{ is reduced.} \tag{26.1.1.1}$$

In fact the inertia degree, ramification index and the number of prime ideals above a prime p satisfy the following relation, which is the subject of the next exercise; the reader acquainted with Riemann surfaces may find this equation familiar:

$$\sum_{\mathfrak{p} \mid (p)} e_\mathfrak{p} f_\mathfrak{p} = [K : \mathbb{Q}]. \tag{26.1.1.2}$$

Exercise 26.1.2.
1. Let $k \geq 0$ and $\mathfrak{p}$ a prime ideal of $\mathcal{O}_K$ lying above a prime number p. Show that $\mathfrak{p}^k/\mathfrak{p}^{k+1}$ is a one dimensional vector space over $\mathcal{O}_K/\mathfrak{p}$.
2. Deduce that $[\mathcal{O}_K : \mathfrak{p}^k] = [\mathcal{O}_K : \mathfrak{p}]^k$.
3. Prove equation (26.1.1.2).
4. Let I, J be non-zero ideals of $\mathcal{O}_K$. Then $N(IJ) = N(I)N(J)$. Show that the equality $N(IJ) = N(I)N(J)$ fails in general for ideals I, J in quadratic rings whose discriminant is not fundamental.

Let us also mention the following important result connecting discriminants and ramification.

Theorem 26.1.3. *Let K be a number field and let $\Delta(\mathcal{O}_K) \in \mathbb{Z}$ be the discriminant of its ring of integers. For a prime number p,*

$$p \text{ is ramified in } \mathcal{O}_K \Leftrightarrow p \mid \Delta(\mathcal{O}_K);$$

in particular, there are finitely many prime numbers which ramify in K.

Proof. For the proof of $\Rightarrow$ see Problem Session 5; for the other implication see for example [21]. □

26.1.4. **A geometric analogue.* (cf. Example 22.4.3)

Let $P(X,Y) \in \mathbb{C}[X,Y]$ be an irreducible polynomial, and let C be the plane curve with equation $P(X,Y) = 0$. Assume that $P(X,Y)$ has degree n and, for simplicity, say that the coefficient of Y^n in $P(X,Y)$ is non zero (e. g. take $P(X,Y) = Y^2 - X, n = 2$). The map $(x,y) \mapsto x$ induces a map

$$p : C \to \mathbb{C};$$

for every $x \in \mathbb{C}$, write

$$p^{-1}(x) = \{(x,y_1), \dots, (x,y_d))\};$$

then $\{y_1, \dots, y_d\}$ is the set of roots of the degree n polynomial $P(x,Y) \in \mathbb{C}[Y]$. Letting e_{y_i} be the multiplicity of the root y_i we obtain

$$\sum_{(x,y)\in p^{-1}(x)} e_y = n.$$

This formula resembles equation (26.1.1.2)! The only difference is that there is no contribution coming from inertia here: this has to do with the fact that we are working over $\mathbb{C}$, which does not have non-trivial algebraic extensions. To see inertia appearing one can for example replace $\mathbb{C}$ by $\mathbb{R}$ (or by a finite field).

Notice that $p^{-1}(x)$ consists of strictly less than n elements - i. e. x is ramified in C - if and only if the polynomial $P(x,Y)$ has multiple roots. Observe that

$$\mathbb{C}[X,Y]/(P(X,Y),(X-x)) \simeq \mathbb{C}[Y]/(P(x,Y)) \simeq \prod_{(x,y)\in p^{-1}(x)} \mathbb{C}[Y]/(Y-y)^{e_y}.$$

Hence x is ramified if and only if $\mathbb{C}[X,Y]/(P(X,Y),(X-x))$ is not reduced. This is the geometric counterpart of what we observed in (26.1.1.1).

Recall that $P(x,Y)$ has a multiple root if and only if $P(x,Y)$ and $\frac{\mathrm{d}}{\mathrm{d}Y}P(x,Y)$ have a common root. Hence the ramification locus is the image via p of the set of points satisfying the system of equations

$$P(X,Y) = 0$$
$$\frac{\partial}{\partial Y}P(X,Y) = 0.$$

Therefore x is ramified in C if and only if it is a root of the *resultant* $\mathrm{res}_Y(P(X,Y), \frac{\partial}{\partial Y}P(X,Y)) \in \mathbb{C}[X]$. In particular the ramification locus consists of *finitely many* points.

Example 26.1.5. Let $C : X = P(Y)$ and

$$p : C \to \mathbb{C}$$
$$(x,y) \mapsto x$$

and let $n = \deg(P(Y)) \geq 1$. Then $x \in \mathbb{C}$ is ramified if and only if the equation $P(Y) = x$ has a double root. Hence the ramification locus consists of points of the form $P(y)$ where y satisfies $\frac{\mathrm{d}}{\mathrm{d}Y}P(y) = 0$.

26.2. **Factorisation of primes in the ring of integers of a quadratic field.**

Proposition 26.2.1. *Let $d \equiv 0, 1 \pmod 4$ be a fundamental discriminant (which is not a square), and let $\mathcal{O}_d$ be the quadratic ring of discriminant d. Let p be a prime number; then:*

1. *p is ramified in $\mathcal{O}_d$ if and only if $p \mid d$.*
2. *If p is odd, then p splits in $\mathbb{Q}(\sqrt{d})$ if and only if the equation $x^2 \equiv d \pmod p$ has a solution.*
3. *If p is odd, then p is inert in $\mathbb{Q}(\sqrt{d})$ if and only if the equation $x^2 \equiv d \pmod p$ has no solution.*

(4) *If* $d \equiv 1 \pmod 4$ *then* 2 *splits (resp. is inert) in* $\mathbb{Q}(\sqrt{d})$ *if* $d \equiv 1 \pmod 8$ *(resp.* $d \equiv 5 \pmod 8$*).*

Proof. We have $\mathcal{O}_d = \mathbb{Z}[X]/(P_\alpha(X))$, where $P_\alpha(X)$ is the characteristic polynomial of $\alpha = \frac{\sqrt{d}}{2}$ if $d \equiv 0 \pmod 4$ and $\alpha = \frac{1+\sqrt{d}}{2}$ if $d \equiv 1 \pmod 4$.

Assume that $d \equiv 0 \pmod 4$ and let $d' = \frac{d}{4}$. Then $\mathcal{O}_d = \mathbb{Z}[X]/(X^2 - d')$ and $\mathcal{O}_d/(p) = \mathbb{F}_p[X]/(X^2 - d')$. The latter ring is reduced if and only if p is odd and $p \nmid d'$; this proves (1) in this case. Now assume that p is unramified. Then p splits if and only if $\mathcal{O}_d/(p)$ is not an integral domain, i.e. if and only if $X^2 - d' \in \mathbb{F}_p[X]$ is reducible. As d' is a square modulo an odd prime p if and only if $4d'$ is a square, we obtain (2) and (3).

Let us now take $d \equiv 1 \pmod 4$. Then $\mathcal{O}_d = \mathbb{Z}[X]/(X^2 - X + \frac{1-d}{4})$. We have $X^2 - X + \frac{1-d}{4} = (X - \frac{1}{2})^2 - \frac{d}{4}$. Hence for $p \neq 2$ we see that $\mathbb{F}_p[X]/(X^2 - X + \frac{1-d}{4})$ is reduced if and only if $p \nmid d$, and in this case p splits if and only if $\frac{d}{4}$ is a square modulo p, which is equivalent to d being a square modulo p.

Finally, for $p = 2$ we have $X^2 - X + \frac{1-d}{4} = X^2 + X \in \mathbb{F}_2[X]$ if $d \equiv 1 \pmod 8$ and $X^2 - X + \frac{1-d}{4} = X^2 + X + 1 \in \mathbb{F}_2[X]$ if $d \equiv 5 \pmod 8$, from which the last point in the proposition follows. □

26.2.2. *The reciprocity map.* Let $d \equiv 0, 1 \pmod 4$ be a fundamental discriminant. Let us denote by $\mathcal{P}_d$ the set of prime numbers not dividing $2d$. Let us introduce the map

$$\begin{aligned} \mathrm{rec}_d : \mathcal{P}_d &\to \{\pm 1\} \\ p &\mapsto \begin{cases} 1 \text{ if } p \text{ splits in } \mathbb{Q}(\sqrt{d}) \\ -1 \text{ if } p \text{ is inert in } \mathbb{Q}(\sqrt{d}). \end{cases} \end{aligned}$$

The previous proposition tells us that the map rec_d coincides with the map given by the Legendre symbol

$$\begin{aligned} \mathcal{P}_d &\to \{\pm 1\} \\ p &\mapsto \left(\frac{d}{p}\right). \end{aligned}$$

We have seen in the examples in 8.3 that, for $d = -3, -4, -20$, the map rec_d is *periodic* with period d: in other words, in those examples the behaviour of the ideal generated by a prime belonging to $\mathcal{P}_d$ in the quadratic ring $\mathcal{O}_d$ *only depends on the residue class of* p modulo d. We established this making crucial use of quadratic reciprocity. In the next lecture we will prove the following theorem, stating that the same phenomenon occurs for arbitrary quadratic rings, and we will deduce quadratic reciprocity from it.

Theorem 26.2.3. *The map* $\mathrm{rec}_d : \mathcal{P}_d \to \{\pm 1\}$ *is periodic with period* d*.*

Remark 26.2.4. We encourage the reader to take time to think about this theorem, which is one of the most important results in this course. While the Legendre symbol $\left(\frac{a}{b}\right)$ is clearly periodic in the entry a, there is no reason a priori why it should be periodic in b. Let us offer a more concrete - but equally non-trivial - reformulation of the theorem in the special case $d \equiv 0 \pmod 4$. In this situation the theorem states that primes $p \in \mathcal{P}_d$ such that the *quadratic* polynomial $X^2 - d$ has a root modulo p belong to a finite number of arithmetic progressions (which are *linear* objects).

26.3. **Factorisation of primes in the ring of integers of a cyclotomic field.** Let p be an odd prime. Recall that the field $\mathbb{Q}(\zeta_p)$ is a number field of degree $p - 1$, with ring of integers $\mathbb{Z}[\zeta_p] = \mathbb{Z}[X]/(\Phi_p(X))$, where $\Phi_p(X) = X^{p-1} + X^{p-2} + \ldots + 1$. Hence, as in the previous section, the factorisation of a prime number l in $\mathbb{Z}[\zeta_p]$ depends on how the polynomial $\Phi_p(X)$ factors in $\mathbb{F}_l[X]$.

Lemma 26.3.1. *The only prime which ramifies in* $\mathbb{Q}(\zeta_p)$ *is* p*.*

Proof. We have $\Phi_p(X) = \frac{X^p - 1}{X - 1} \equiv (X - 1)^{p-1} \pmod p$. Hence $\mathbb{Z}[\zeta_p]/(p) \simeq \mathbb{F}_p[X]/(X - 1)^{p-1}$ is not reduced, and p is ramified - in fact, we had already seen that $(1 - \zeta_p)^{p-1} = p\mathbb{Z}[\zeta_p]$.

Let us now show that every prime $l \neq p$ is unramified in $\mathbb{Z}[\zeta_p]$. Equivalently, we have to prove that $\Phi_p(X) \in \mathbb{F}_l[X]$ factors as a product of distinct irreducible polynomials. Suppose by contradiction that $\Phi_p(X) = Q^2 \cdot R$ for some $R \in \mathbb{F}_l[X]$ and $Q \in \mathbb{F}_l[X] \smallsetminus \mathbb{F}_l$. Then $X^p - 1 =$

$(X-1)Q^2 \cdot R$ hence Q divides both X^p-1 and its derivative. The derivative of X^p-1 is pX^{p-1} which is coprime with X^p-1 in $\mathbb{F}_l[X]$, contradiction. □

We now want to determine how (the ideals generated by) primes $l \neq p$ factor in (the ring of integers of) $\mathbb{Q}(\zeta_p)$; this boils down to understanding how $\Phi_p(X)$ factors modulo l.

Proposition 26.3.2. *Let $l \neq p$ be a prime, and f the order of l in $\mathbb{F}_p^\times$ (i. e. the smallest positive integer such that $l^f \equiv 1 \pmod p$). Then $\Phi_p(X)$ factors as a product of $\frac{p-1}{f}$ distinct factors of degree f in $\mathbb{F}_l[X]$.*

Proof. We know that $\Phi_p(X) \in \mathbb{F}_l[X]$ factors as a product of distinct irreducible polynomials. Let $P \in \mathbb{F}_l[X]$ be an irreducible factor of $\Phi_p(X)$; let $d = \deg P$, $K = \mathbb{F}_l[X]/(P)$ and $x \in K$ the image of $X \in \mathbb{F}_l[X]$. Then K has cardinality l^d, and x is an element of order p in $K^\times$, hence $l^d \equiv 1 \pmod p$. Furthermore the elements $1, x, \dots, x^{p-1} \in K^\times$ are distinct, hence they are all the roots of the polynomial $X^p - 1 \in \mathbb{F}_l[X]$. In particular $K^\times$ contains all the roots of $\Phi_p(X) \in \mathbb{F}_l[X]$.

If Q is another irreducible factor of $\Phi_p(X) \in \mathbb{F}_l[X]$, it follows from the previous discussion that there is $\zeta \in K^\times$ such that $Q(\zeta) = 0$. Hence the map $\mathbb{F}_l[X] \to K$ sending X to ζ factors through an injection $\mathbb{F}_l[X]/(Q) \hookrightarrow K$. We have proved that, if P and Q are any two irreducible factors of $\Phi_p(X)$, of degrees d and e, the inequality $e \leq d$ holds. By symmetry we must have $d = e$. Hence we learn that all irreducible factors of $\Phi_p(X)$ have the same degree d, which satisfies $l^d \equiv 1 \pmod p$. Therefore the order f of l in $\mathbb{F}_p^\times$ divides d. It remains to prove that $f = d$.

Notice that every element $y \in K$ satisfies the equation $y^{l^d} = y$, hence K contains all the roots of the polynomial $X^{l^d} - X$. Let $L = \{y \in K \mid x^{l^f} = x\} \subset K$. Then L is a subfield of K; as $f \mid d$ the field L has l^f elements and, as $l^f \equiv 1 \pmod p$, we have $X^p - 1 \mid X^{l^f-1} - 1$, so L contains all the roots of $\Phi_p(X)$. The same argument as before shows that there is an injection $K \hookrightarrow L$, hence $K = L$ and the proof is complete. □

Corollary 26.3.3. *Let p be an odd prime and let l be a prime.*

1. *l is ramified in $\mathbb{Q}(\zeta_p)$ if and only if $l = p$.*
2. *Let $l \neq p$ be a prime and f the order of l in $\mathbb{F}_p^\times$. Then $l\mathbb{Z}[\zeta_p]$ factors as a product of $\frac{p-1}{f}$ distinct prime ideals of inertia degree f.*
3. *l splits completely in $\mathbb{Q}(\zeta_p)$ if and only if $l \equiv 1 \pmod p$.*

Proof. (1) was proved in Lemma 26.3.1; (2) follows from the previous proposition and from the isomorphism $\mathbb{Z}[\zeta_p] \simeq \mathbb{Z}[X]/(\Phi_p(X))$. (3) is a special case of (2). □

Remark 26.3.4. Observe that, in particular, the previous corollary tells us that the factorisation of a prime l in $\mathbb{Z}[\zeta_p]$ only depends on the congruence class of l modulo p. We stated in Theorem 26.2.3 (to be proved in the next lecture) that the same phenomenon for quadratic fields, so it is natural to wonder for which number fields this property holds. The answer is: precisely for those which are contained in a cyclotomic field $\mathbb{Q}(\zeta_n)$ for some $n \geq 1$. This remarkable result is one of the main achievements of *class field theory*, an extremely deep and elegant branch of Algebraic Number Theory developed in the first half of the 20th century by Artin, Hasse, Furtwängler, Hilbert, Takagi, Tate (among others). See [9, Chapter 2].

26.3.5. *Properties of finite fields.* Our computations in the proof of Proposition 26.3.2 are related to the basic properties of finite fields, which we list here for the reader's convenience.

1. Let F be a finite field. Then the multiplicative group $F^\times$ is *cyclic*, i. e. there exists $a \in F^\times$ such that every element of $F^\times$ is a power of a. More generally, every finite subgroup of the multiplicative group of a field is cyclic.
2. Every finite field F has cardinality l^k for some prime l and positive integer $k \geq 1$. The polynomial $X^{l^k} - X$ factors as a product of linear factors in F.
3. For every prime l and every $k \geq 1$ there exists a field with l^k elements, unique up to isomorphism. It is denoted by $\mathbb{F}_{l^k}$. The map sending $x \in \mathbb{F}_{l^k}$ to x^l is a field automorphism of $\mathbb{F}_{l^k}$, called the *Frobenius morphism*.
4. If k, k' are positive integers then $\mathbb{F}_{l^k} \subset \mathbb{F}_{l^{k'}}$ if and only if $k \mid k'$.

Proof. (1) The key point is that, if F is a field, a polynomial $P(X) \in F[X]$ of degree n has at most n roots in F. There are several proofs of the fact that a finite subgroup $G \subset F^\times$ is

cyclic using this result. Here's a quick one using what we already know: as G is a finite abelian group it is isomorphic to a product $\prod_{i=1}^{r} \mathbb{Z}/q_i^{j_i}\mathbb{Z}$ where the q_i's are primes and $j_i \geq 1$ for every i. We claim that the primes q_i are distinct, from which the fact that G is cyclic follows. Assume by contradiction that, say, $q_1 = q_2 = q$. Then $\mathbb{Z}/q^{j_1}\mathbb{Z} \times \mathbb{Z}/q^{j_2}\mathbb{Z} \subset F^\times$ contains q^2 solutions of the equation $X^q = 1$, absurd.

(2) A finite field F must contain $\mathbb{F}_l$ for some l, hence it is a finite dimensional vector space over $\mathbb{F}_l$ of degree $k \geq 1$. It follows that $F^\times$ has cardinality $l^k - 1$, hence every $a \in F^\times$ satisfies $a^{l^k-1} = 1$. Therefore the polynomial $X^{l^k} - X$ has all its roots in F.

(3) To construct a field with l^k elements one (formally) adjoins to $\mathbb{F}_l$ the roots of the polynomial $P = X^{l^k} - X$. Uniqueness follows from the fact that any field with l^k elements is the "smallest" field containing $\mathbb{F}_l$ and the roots of P. Such a field is called the *splitting field* of P, and one proves that it is unique up to isomorphism; we omit the details. The fact that $x \mapsto x^l$ is a ring morphism of $\mathbb{F}_{l^k}$ follows from the fact that l divides the binomial coefficients $\binom{l}{i}$ for $1 \leq i \leq l-1$.

(4) If $\mathbb{F}_{l^k} \subset \mathbb{F}_{l^{k'}}$ then $k = \dim_{\mathbb{F}_l} \mathbb{F}_{l^k} \mid k' = \dim_{\mathbb{F}_l} \mathbb{F}_{l^{k'}}$. Conversely, if $k \mid k'$ then $X^{l^k} - X \mid X^{l^{k'}} - X$, hence $\{x \in \mathbb{F}_{l^{k'}} \mid x^{l^k} = x\}$ is a subfield of $\mathbb{F}_{l^{k'}}$ with l^k elements.

□

27. Lecture 27: The quadratic reciprocity law

27.1. **Gauss lemma on the Legendre symbol.** Let $d \equiv 0, 1 \pmod 4$ be an integer which is not a square, and assume that it is a fundamental discriminant. In the last lecture we have introduced the reciprocity map

$$\begin{aligned} \mathrm{rec}_d : \mathcal{P}_d &\to \{\pm 1\} \\ p &\to \left(\frac{d}{p}\right) \end{aligned}$$

where $\mathcal{P}_d$ is the set of prime numbers not dividing $2d$. The map rec_d encodes the behaviour of the ideal generated by a prime $p \in \mathcal{P}_d$ in the quadratic ring $\mathcal{O}_d$. In this lecture we want to show the following theorem, which implies the periodicity of the reciprocity map stated in Theorem 26.2.3.

Theorem 27.1.1. *Let $d \equiv 0, 1 \pmod 4$ be a non-zero integer. The map $\mathcal{P}_d \to \{\pm 1\}$ sending p to $\left(\frac{d}{p}\right)$ is periodic with period d, i. e., if $p, q \in \mathcal{P}_d$ and $p \equiv q \pmod d$ then $\left(\frac{d}{p}\right) = \left(\frac{d}{q}\right)$.*

As we mentioned, the above theorem implies Theorem 26.2.3; in fact it is slightly more general, as we do not require d to be a fundamental discriminant. Working in this generality will be useful later when we will explore the relation between the theorem and quadratic reciprocity. The key input for our proof of the theorem is the following description of the Legendre symbol found by Gauss. Fix a prime p and choose a set of representatives $\{x_1, \dots, x_{\frac{p-1}{2}}\}$ of the quotient $\mathbb{F}_p^\times / \{\pm 1\}$. Take an integer a coprime to p. For $1 \leq i \leq \frac{p-1}{2}$ we have

$$a x_i = \varepsilon_i x_{\sigma(i)}$$

for a unique $\sigma(i) \in \{1, \dots, (p-1)/2\}$ and $\varepsilon_i \in \{\pm 1\}$. Furthermore the map $i \mapsto \sigma(i)$ is a permutation of the set $\{1, \dots, (p-1)/2\}$.

Lemma 27.1.2. *(Gauss) Let a be an integer coprime to p. Then*

$$\left(\frac{a}{p}\right) = \prod_{i=1}^{\frac{p-1}{2}} \varepsilon_i.$$

Proof.

$$\begin{aligned} a^{\frac{p-1}{2}} \prod_{i=1}^{\frac{p-1}{2}} x_i \equiv \prod_{i=1}^{\frac{p-1}{2}} (a x_i) \equiv \prod_{i=1}^{\frac{p-1}{2}} \varepsilon_i x_{\sigma(i)} \equiv \prod_{i=1}^{\frac{p-1}{2}} \varepsilon_i \prod_{i=1}^{\frac{p-1}{2}} x_i \pmod p \\ \Rightarrow a^{\frac{p-1}{2}} \equiv \prod_{i=1}^{\frac{p-1}{2}} \varepsilon_i \pmod p. \end{aligned}$$

Hence the lemma follows from the fact that $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod p$ (Euler's criterion). □

Exercise 27.1.3. Use Gauss lemma to prove that, for an odd prime p:

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}, \quad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

Question 27.1.4. Take a prime $p \equiv 1 \pmod 3$, and consider the problem of describing *cubes* in $\mathbb{F}_p^\times$. What is the analogue of the Legendre symbol in this situation, and how does the above lemma generalise?

27.2. **Periodicity of the reciprocity map: proof.** We will describe the proof of Theorem 27.1.1 under the assumption that $d > 0$, which will be in force from now on; we leave it to the reader to modify the argument to cover the case $d < 0$. Let us choose as set of representatives of $\mathbb{F}_p^\times / \{\pm 1\}$ the set $A = \{1, \dots, \frac{p-1}{2}\}$. We set $k = \frac{d}{4}$ if $d \equiv 0 \pmod 4$ and $k = d$ if $d \equiv 1 \pmod 4$. For each

$i \in A$ we have $ki \equiv \varepsilon_i \sigma(i) \pmod{p}$, where $\varepsilon_i \in \{\pm 1\}$ and $\sigma(i) \in A$. For $i \in A$ we have

$$ki \in [0, \frac{p}{2}] \Leftrightarrow i \in [0, \frac{p}{2k}] \Rightarrow \varepsilon_i = 1;$$
$$ki \in [\frac{p}{2}, p] \Leftrightarrow i \in [\frac{p}{2k}, \frac{p}{k}] \Rightarrow \varepsilon_i = -1;$$
$$ki \in [p, \frac{3p}{2}] \Leftrightarrow i \in [\frac{p}{k}, \frac{3p}{2k}] \Rightarrow \varepsilon_i = 1;$$
$$ki \in [\frac{3p}{2}, 2p] \Leftrightarrow i \in [\frac{3p}{2k}, \frac{2p}{k}] \Rightarrow \varepsilon_i = -1;$$
$$\dots\dots\dots$$

Notice that as p is coprime to d none of the points $ki, i \in A$ is an endpoint of one of the above intervals. In view of Gauss lemma the value $\left(\frac{d}{p}\right) = \left(\frac{k}{p}\right)$ equals the parity of the cardinality of the set

$$A \cap \left([\frac{p}{2k}, \frac{p}{k}] \cup [\frac{3p}{2k}, \frac{2p}{k}] \cup \ldots \cup [\frac{(2j+1)p}{2k}, \frac{(2j+2)p}{2k}] \cup \cdots \right). \tag{27.2.0.1}$$

Observe that p is now at the *numerator*, while k is at the *denominator*, so we may hope to use the above description to prove periodicity in p of the symbol $\left(\frac{k}{p}\right)$.

Let us first determine the largest positive integer e such that the interval $[\frac{(2e-1)p}{2k}, \frac{ep}{k}]$ can have non-empty intersection with A. For $i \in A$ we have $ki \leq \frac{kp}{2}$, hence:

(1) if k is even then the last interval in (27.2.0.1) which can contain an integer $1 \leq i \leq \frac{p}{2}$ is $[\frac{(k-1)p}{2k}, \frac{kp}{2k}]$, so $e = \frac{k}{2}$;
(2) if k is odd then the last interval in (27.2.0.1) which can contain an integer $1 \leq i \leq \frac{p}{2}$ is $[\frac{(k-2)p}{2k}, \frac{(k-1)p}{2k}]$, so $e = \frac{k-1}{2}$.

Hence $2e$ is the largest even integer not exceeding k; in particular e *only depends on* k, and we need to study the parity of the number of integers in the union of intervals

$$[\frac{p}{2k}, \frac{p}{k}] \cup [\frac{3p}{2k}, \frac{2p}{k}] \cup \ldots \cup [\frac{(2j+1)p}{2k}, \frac{(2j+2)p}{2k}] \cup \cdots \cup [\frac{(2e-1)p}{2k}, \frac{ep}{k}]. \tag{27.2.0.2}$$

Let us distinguish two cases:

Case $d \equiv 0 \pmod 4$**:** if we replace p by a prime $q = p + 4ak$ for some $a \in \mathbb{Z}_{>0}$ then e is unchanged. On the other hand making such a replacement each interval

$$[\frac{(2j+1)p}{2k}, \frac{(2j+2)p}{2k}] \text{ in (27.2.0.2) gets replaced by}$$
$$[\frac{(2j+1)p}{2k} + (2j+1)2a, \frac{(2j+2)p}{2k} + (2j+2)2a].$$

So we see that the endpoints of each interval are shifted by even integers, hence the *parity* of the number of integers contained in each interval is unchanged. Therefore $\left(\frac{k}{p}\right) = \left(\frac{k}{q}\right)$ if p and q are primes in $\mathcal{P}_d$ which are congruent modulo $4k = d$.

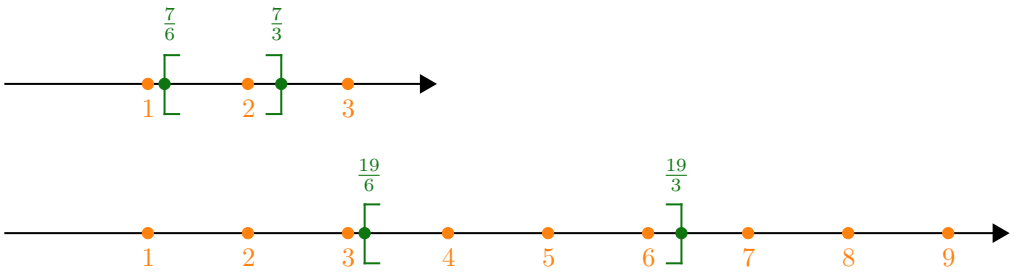


FIGURE 5. The case $d = 12, p = 7, q = 19$.

Case $d \equiv 1 \pmod 4$: if $q > p \in \mathcal{P}_d$ are congruent modulo $k = d$ then, as $q - p$ is even, we can write $q - p = 2ak$ for some $a \in \mathbb{Z}_{>0}$. Replacing p by $q = p + 2ak$ each interval

$$[\frac{(2j+1)p}{2k}, \frac{(2j+2)p}{2k}] \text{ in (27.2.0.2) gets replaced by}$$
$$[\frac{(2j+1)p}{2k} + (2j+1)a, \frac{(2j+2)p}{2k} + (2j+2)a].$$

In particular the parity of the number of integers in each interval in (27.2.0.2) is unchanged (resp. changes) if a is even (resp. a is odd). As $d \equiv 1 \pmod 4$ the number e of intervals in (27.2.0.2) is even, hence the parity of the number of integers in (27.2.0.2) is unchanged, so $\left(\frac{k}{p}\right) = \left(\frac{k}{q}\right)$.

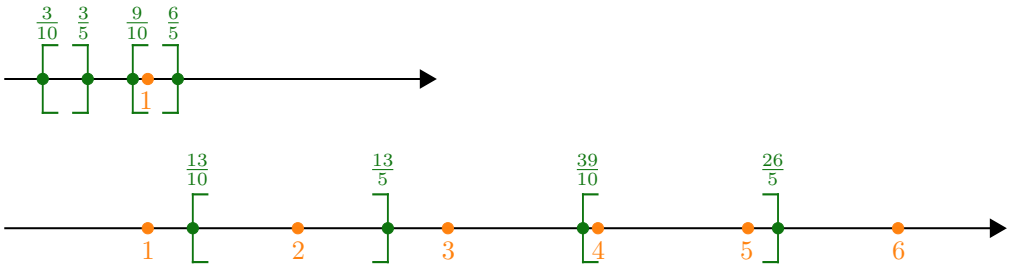


FIGURE 6. The case $d = 5, p = 3, q = 13$.

27.3. **The quadratic reciprocity law.** The periodicity of the reciprocity map is a manifestation of the *quadratic reciprocity law*. In fact we can deduce quadratic reciprocity from the theorem we proved: take $q > p$ two odd primes.

(1) If $p \equiv q \equiv 1 \pmod 4$ write $q - p = 4k$. Applying Theorem 27.1.1 with $d = 4k$ we find
$$\left(\frac{p}{q}\right) = \left(\frac{-4k}{q}\right) = \left(\frac{4k}{q}\right) = \left(\frac{4k}{q-4k}\right) = \left(\frac{4k}{p}\right) = \left(\frac{q}{p}\right).$$
(2) If $p \equiv q \equiv 3 \pmod 4$ write $q - p = 4k$. Then
$$\left(\frac{p}{q}\right) = \left(\frac{-4k}{q}\right) = -\left(\frac{4k}{q}\right) = -\left(\frac{4k}{q-4k}\right) = -\left(\frac{4k}{p}\right) = -\left(\frac{q}{p}\right).$$

The case $p \not\equiv q \pmod 4$ requires a slight extension of the arguments explained above which is left to the reader.

Remark 27.3.1. The quadratic reciprocity law really is a very important and remarkable statement, although it may take long time to fully appreciate it. [13] It was first proved by Gauss in 1796; the first proofs were published in [11], and over the years Gauss found (at least) 8 proofs of quadratic reciprocity. The approach we presented is due to Scholz (1939). In order to convince the reader of the everlasting interest in quadratic reciprocity, let us also point out the following recent, high-tech proof of quadratic reciprocity making use of sophisticated algebraic topology machinery: [4].

Exercise 27.3.2. (1) Adapt the above argument to prove Theorem 27.1.1 in the case $d < 0$. Furthermore, prove the missing case of quadratic reciprocity.

(2) Let $a, b \in \mathbb{Z} \smallsetminus \{0\}$ be two integers such that $\gcd(2a, b) = 1$ and $b > 0$. Factor $b = p_1 \cdots p_r$. Define the *Jacobi symbol*
$$\left(\frac{a}{b}\right) = \prod_{i=1}^{r} \left(\frac{a}{p_i}\right).$$
Prove that, if $a, b > 0$ are coprime and odd, then

$$\left(\frac{a}{b}\right) = (-1)^{\frac{a-1}{2}\frac{b-1}{2}} \left(\frac{b}{a}\right). \tag{27.3.2.1}$$

[13] At least, this has been the case for the author of this text.

One way to do this is to deduce the formula from quadratic reciprocity for the Legendre symbol; another strategy is the following:
   (a) extend Gauss lemma to the Jacobi symbol;
   (b) realise that the arguments used in the proof of Theorem 27.1.1 did not quite use the fact that p and q are prime; generalise Theorem 27.1.1 and deduce (27.3.2.1).

(3) We will now prove that, if $a > 0$ is not a square, then there exist infinitely many primes p such that a is not a square modulo p. In other words, if the equation $X^2 \equiv a \pmod{p}$ has a solution for all but finitely many primes p, then a must be a square. We will assume $a \neq 2$ for simplicity (for the case $a = 2$ see [16, Chapter 5]).
   (a) Show that we may assume that a is squarefree. Hence we can write $a = 2^e p_1 \cdots p_r$ with p_i odd for $1 \leq i \leq r$, $p_i \neq p_j$ if $i \neq j$ and $r \geq 1$ (as we are assuming $a \neq 2$).
   (b) Let $q_1, \ldots, q_k$ be a finite list of odd prime numbers not including any factor of a. To prove our result it suffices to show that there exists a prime q different from $q_1, \ldots, q_k$ such that $\left(\frac{a}{q}\right) = -1$. Pick a positive integer b such that

$$\begin{aligned} b &\equiv 1 \pmod{8} \\ b &\equiv 1 \pmod{p_i}, i = 1 \ldots, r-1 \\ b &\equiv c \pmod{p_r} \text{ where } c \text{ is any quadratic nonresidue } \pmod{p_r} \\ b &\equiv 1 \pmod{q_i}, i = 1 \ldots, k. \end{aligned}$$

   (c) Show that $b \equiv 1 \pmod{8} \Rightarrow \left(\frac{2}{b}\right) = 1$; deduce that $\left(\frac{a}{b}\right) = -1$.
   (d) Deduce that there is a prime factor q of b such that $\left(\frac{a}{q}\right) = -1$; conclude.
   (e) Try to prove the statement you just established without using quadratic reciprocity, until you are fully convinced that its contribution is non-trivial and crucial.

(4) Show that, given $b > 0$, the Jacobi symbol $\left(\frac{a}{b}\right)$ can be computed performing at most $C \log_2(b)$ operations of division with remainder (allowing the remainder to be a negative number), for a suitable constant C not depending on a, b.

Remark 27.3.3. We learn from the last exercise that quadratic reciprocity allows to compute the Jacobi symbol in a *polynomial* time. This has applications to the *Solovay-Strassen primality test*, which is based on the following idea: there are (a lot of) counterexamples to Euler's criterion for the symbol $\left(\frac{a}{b}\right)$ if b is not prime. So if we are given an odd integer b and we want to check if it is prime, we can pick an integer $2 \leq a \leq b-2$, compute $a^{\frac{b-1}{2}} \pmod{b}$ (which can be done in polynomial time), and compute the symbol $\left(\frac{a}{b}\right)$. If we find different numbers we can conclude that b is not prime; if not, we can try another a. In fact, assuming the Generalised Riemann Hypothesis, this idea can be used to construct an algorithm which determines if an integer is prime in a polynomial time, see [7]. In 2004 Agrawal, Kayal and Saxena constructed a different primality test and were able to prove (unconditionally) that it runs in polynomial time.

28. Lecture 28: Thue's theorem - part I

This and the next lecture are devoted to the proof of Thue's result on Diophantine approximation of algebraic numbers.

28.1. **Thue's theorem: statement and general strategy.** In lecture 19 we defined the irrationality exponent $e(\alpha)$ of a real number α to be the supremum of the real numbers δ such that the inequality

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{|q|^\delta}$$

has infinitely many solutions $\frac{p}{q} \in \mathbb{Q}$. Liouville's theorem states that if α is an algebraic number of degree d then $e(\alpha) \leq d$, and Dirichlet's theorem implies that the previous inequality is an equality for $d = 2$. On the other hand we asked ourselves in question 19.3.10 whether one can improve Liouville's estimate in the case $d \geq 3$. This question was answered affirmatively by Thue, who proved the following remarkable result:

Theorem 28.1.1 (Thue, 1909)**.** *Let $\alpha \in \mathbb{R}$ be an algebraic number of degree $d > 1$. For every $\delta > \frac{d}{2} + 1$ the inequality*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{|q|^\delta} \tag{28.1.1.1}$$

has finitely many solutions $\frac{p}{q} \in \mathbb{Q}$. In other words, $e(\alpha) \leq \frac{d}{2} + 1$.

If $d \geq 3$ then $\frac{d}{2} + 1 < d$, hence Thue's theorem answers question 19.3.10 affirmatively. As observed in Lecture 19, this yields finiteness results for integral solutions of Thue equations, hence of Mordell equations.

Corollary 28.1.2.
(1) *Let $P(X, Y) \in \mathbb{Z}[X, Y]$ be a homogeneous polynomial of degree at least 3, irreducible in $\mathbb{Q}[X, Y]$. For every $m \in \mathbb{Z} \smallsetminus \{0\}$ the equation $P(X, Y) = m$ has finitely many integral solutions.*
(2) *For every $k \in \mathbb{Z} \smallsetminus \{0\}$ the equation $Y^2 = X^3 + k$ has finitely many integral solutions.*

Example 28.1.3.
(1) The equation $X^3+3X^2Y+6XY^2+2Y^3 = 1$, which we found in (17.1.2.1), has finitely many integral solutions.
(2) We know that the equation $X^2 - 2Y^2 = 1$ has infinitely many solutions. However, for any integer $n \geq 3$, the equation $X^n - 2Y^n = 1$ has finitely many integral solutions.

The general strategy to prove Thue's theorem is similar to the one we used to establish Liouville's theorem, although the execution is substantially more complicated. Before giving the details let us list the main steps. Fix $\delta > \frac{d}{2} + 1$; take two solutions $a_1 = \frac{p_1}{q_1}, a_2 = \frac{p_2}{q_2}$ of the inequality (28.1.1.1). Here is our plan:

- **Step 1: construction of an auxiliary polynomial:** we will construct a polynomial in two variables $A(X, Y) = P(X) - YQ(X) \in \mathbb{Z}[X, Y]$ vanishing with *high order* at (α, α) and having *bounded* coefficients. We will not do this explicitly, but instead use Siegel's lemma.
- **Step 2: an upper bound:** as a_1, a_2 are close to α and A vanishes at (α, α), a straightforward application of Taylor's formula will allow us to bound above $|A(a_1, a_2)|$. In fact we will also need to bound certain derivatives $\frac{\partial^j A}{\partial X^j}(a_1, a_2)$, in view of the next step.
- **Step 3: the non-vanishing step:** this is the hardest step: we would like to argue that $A(a_1, a_2) \neq 0$, so that we get a lower bound on $|A(a_1, a_2)|$, but this is not clear unfortunately; we will show the weaker fact that $\frac{\partial^j A}{\partial X^j}(a_1, a_2) \neq 0$ for some *small enough* j.
- **Step 4: a lower bound:** Given the non-vanishing step, this will follow from the fact that a non-zero integer has absolute value at least one.
- **Step 5: comparing upper bound and lower bound:** assuming that (28.1.1.1) has infinitely many solutions, we will pick a_1, a_2 with *very large* denominators and check that comparing the upper and lower bounds in steps 2 and 4 leads to a contradiction.

We will now fill in the details of each of the above steps; we will need to make precise the estimates we mentioned, and verify that everything fits together leading to the desired contradiction in the last step. There are various slightly different ways of doing this; we will mostly follow Hindry's account of the argument [15, Chapter 6].

28.2. **Thue's theorem: first steps in the proof.**

28.2.1. *Preliminaries.* We fix a real algebraic number α of degree $d \geq 2$, and a real number $\delta > \frac{d}{2} + 1$. We choose:

(1) a real number $0 < \eta < 1$;
(2) two rational numbers $a_1 = \frac{p_1}{q_1}, a_2 = \frac{p_2}{q_2}$ with $q_1, q_2 > 1$, verifying the inequality (28.1.1.1).

We will later specify more precisely how to make the above choices.

Throughout the proof, positive constants *only depending on* α will be denoted by the letter C variously decorated. For a polynomial $P(X) \in \mathbb{Z}[X]$, we will denote by $\|P\|$ the maximum of the absolute values of the coefficients of P. If $j \geq 0$ is an integer we will let $D^j P = \frac{1}{j!} P^{(j)}$, where $P^{(j)} = \frac{\mathrm{d}^j P}{\mathrm{d}X^j}$.

Let $P, Q \in \mathbb{Z}[X]$. The following inequalities hold true; their verification is left as an exercise.

(1) $\|P + Q\| \leq \|P\| + \|Q\|$.
(2) $\|PQ\| \leq (\deg(P) + 1)\|P\|\|Q\|$.
(3) For $j \geq 0$, $\|D^j P\| \leq \binom{\deg(P)}{j}\|P\| \leq 2^{\deg(P)}\|P\|$.

We will need the following lemma.

Lemma 28.2.2. *There exist positive integers c, C_0 (depending only on α) with the following property: for every integer $r \geq 1$ write $\alpha^r = c_{r,0} + c_{r,1}\alpha + \ldots + c_{r,d-1}\alpha^{d-1}$, where $c_{r,s} \in \mathbb{Q}$ for $0 \leq s \leq d-1$. Then $c^r c_{r,s} \in \mathbb{Z}$ and $|c_{r,s}| \leq C_0^r$ for $0 \leq s \leq d-1$.*

Proof. Take $c \in \mathbb{Z}_{>0}$ such that $c \cdot c_{d,s} \in \mathbb{Z}$ for $0 \leq s \leq d-1$ and let $C_0 = 1 + \max\{|c_{d,s}|, s < d\}$. We will show by induction on r that c, C_0 satisfy the conclusion of the lemma. This is clear for $r < d$, so assume $r \geq d$. We have

$$\begin{aligned}\alpha^r &= \alpha \cdot (c_{r-1,0} + c_{r-1,1}\alpha + \ldots + c_{r-1,d-1}\alpha^{d-1}) \\ &= c_{r-1,0}\alpha + c_{r-1,1}\alpha^2 + \ldots + c_{r-1,d-2}\alpha^{d-1} + c_{r-1,d-1}(c_{d,0} + c_{d,1}\alpha + \ldots + c_{d,d-1}\alpha^{d-1}) \\ &= c_{r-1,d-1}c_{d,0} + (c_{r-1,0} + c_{r-1,d-1}c_{d,1})\alpha + \ldots + (c_{r-1,d-2} + c_{r-1,d-1}c_{d,d-1})\alpha^{d-1}.\end{aligned}$$

As $c \cdot c_{d,s} \in \mathbb{Z}$ and $c^{r-1} \cdot c_{r-1,s} \in \mathbb{Z}$ by induction we deduce that $c^r c_{r,s} \in \mathbb{Z}$ for $0 \leq s \leq d-1$. Furthermore $|c_{r,s}| \leq C_0^{r-1}|1 + c_{d,s}| \leq C_0^r$. □

28.2.3. *Step 1: construction of an auxiliary polynomial.* Let $T \geq 1$ be an integer and $D = \left\lfloor \frac{dT(1+\eta)}{2} \right\rfloor$. There exists a non-zero polynomial $A(X,Y) = P(X) - YQ(X) \in \mathbb{Z}[X,Y]$ and a constant C_1 such that $\max\{\deg(P), \deg(Q)\} \leq D$, $\max\{\|P\|, \|Q\|\} \leq C_1^{\frac{T}{\eta}}$ and

$$\frac{\partial^j A}{\partial X^j}(\alpha, \alpha) = 0, \;\; 0 \leq j \leq T-1.$$

Proof. This is a direct application of Siegel's lemma: we are looking for polynomials

$$P(X) = x_0 + x_1 X + \ldots + x_D X^D, \; Q(X) = y_0 + y_1 X + \ldots + y_D X^D, x_i, y_i \in \mathbb{Z}$$

such that, for $0 \leq j \leq T-1$,

$$0 = D^j P(\alpha) - \alpha D^j Q(\alpha) = \sum_{s=j}^{D} x_s \binom{s}{j} \alpha^{s-j} - \sum_{s=j}^{D} y_s \binom{s}{j} \alpha^{s-j+1}.$$

Using the notation of Lemma 28.2.2 we can write the right hand side above as

$$\sum_{l=0}^{d-1} \left(\sum_{s=j}^{D} \binom{s}{j} (x_s c_{s-j,l} - y_s c_{s-j+1,l}) \right) \alpha^l,$$

hence we need to find solutions x_i, y_i the system of linear equations with $\mathbb{Z}$-coefficients

$$\sum_{s=j}^{D} c^{D+1} \binom{s}{j} (x_s c_{s-j,l} - y_s c_{s-j+1,l}) = 0, \;\; 0 \leq l \leq d-1, 0 \leq j \leq T-1. \tag{28.2.3.1}$$

We have dT equations in $2(D+1)$ unknowns, and our choice of D ensures that $2(D+1) > dT(1+\eta) > dT$, hence we may apply Siegel's lemma. Letting M be the maximum of the absolute

values of the coefficients of the linear equations in (28.2.3.1), we obtain a non-zero solution of our linear system with absolute values of all the x_i, y_i bounded above by

$$(2(D+1)M)^{\frac{dT}{2(D+1)-dT}} \leq (2(D+1)M)^{\frac{1}{\eta}}$$

Finally, by Lemma 28.2.2 we have

$$M \leq 2^D (cC_0)^{D+1} \leq (2c^2C_0^2)^D.$$

Pick $\tilde{C} \geq 2c^2C_0^2$ such that $\tilde{C}^D \geq 2(D+1)$ - notice that there exists $\tilde{C}$ such that the latter inequality holds for every $D \geq 1$. Then using the definition of D and the fact that $0 < \eta < 1$ we obtain

$$(2(D+1)M)^{\frac{1}{\eta}} \leq \tilde{C}^{\frac{2D}{\eta}} \leq \tilde{C}^{\frac{2dT}{\eta}}$$

hence, setting $C_1 = \tilde{C}^{2d}$, we have a non-zero integral solution of the linear system (28.2.3.1) with absolute value of the x_i, y_i not exceeding $C_1^{\frac{T}{\eta}}$. □

28.2.4. *Step 2: an upper bound.* Let $T \geq 1$ be an integer, D the number defined in Step 1 and $A(X,Y) = P(X) - YQ(X) \in \mathbb{Z}[X,Y]$ the polynomial constructed in Step 1. There is a constant $C_2 > 0$ such that for $0 \leq j \leq \frac{T}{2}$ we have

$$|D^j P(a_1) - a_2 D^j Q(a_1)| \leq \max\{q_1^{-\delta(T-j)}, q_2^{-\delta}\} C_2^{D+\frac{T}{\eta}}.$$

Proof. Using Taylor's formula in one variable we obtain:

$$\begin{aligned}
D^j P(a_1) - a_2 D^j Q(a_1) &= \sum_{h\geq 0} \binom{j+h}{h} D^{j+h}P(\alpha)(a_1-\alpha)^h - a_2 \sum_{h\geq 0} \binom{j+h}{h} D^{j+h}Q(\alpha)(a_1-\alpha)^h \\
&= \sum_{h\geq 0} \binom{j+h}{h} (D^{j+h}P(\alpha) - \alpha D^{j+h}Q(\alpha))(a_1-\alpha)^h \\
&\quad - (a_2-\alpha) \sum_{h\geq 0} \binom{j+h}{h} D^{j+h}Q(\alpha)(a_1-\alpha)^h \\
&\overset{Step\ 1}{=} \sum_{h\geq T-j} \binom{j+h}{h} (D^{j+h}P(\alpha) - \alpha D^{j+h}Q(\alpha))(a_1-\alpha)^h \\
&\quad - (a_2-\alpha) \sum_{h\geq 0} \binom{j+h}{h} D^{j+h}Q(\alpha)(a_1-\alpha)^h. \qquad (28.2.4.1)
\end{aligned}$$

Now we have $\|D^{j+h}P\| \leq 2^D\|P\|$ and $\|D^{j+h}Q\| \leq 2^D\|Q\|$, hence

$$|(D^{j+h}P(\alpha) - \alpha D^{j+h}Q(\alpha))| \leq 2(D+1)2^D \max\{1, |\alpha|^{D+1}\} \max\{\|P\|, \|Q\|\}.$$

Furthermore in the first sum after the last equality in (28.2.4.1), non-zero terms can occur only for $j+h \leq D$: therefore there are at most $D+1$ non-zero terms in the sum. By assumption $|a_1-\alpha|^h < q_1^{-\delta(T-j)}$ for $h \geq T-j$, hence, choosing a constant $\overline{C}$ such that $\overline{C}^D \geq 2(D+1)^2 4^D \max\{1, |\alpha|^{D+1}\}$, we obtain

$$\left| \sum_{h\geq T-j} \binom{j+h}{h} (D^{j+h}P(\alpha) - \alpha D^{j+h}Q(\alpha))(a_1-\alpha)^h \right| \leq \overline{C}^D \max\{\|P\|, \|Q\|\} q_1^{-\delta(T-j)}.$$

In the same way one shows that the absolute value of the espression in the last line of (28.2.4.1) is bounded above by $\overline{\overline{C}}^D \|Q\| q_2^{-\delta}$ for a suitable constant $\overline{\overline{C}}$. Finally, we know by Step 1 that $\max\{\|P\|, \|Q\|\} \leq C_1^{\frac{T}{\eta}}$, hence we find

$$|D^j P(a_1) - a_2 D^j Q(a_1)| \leq (\overline{C} + \overline{\overline{C}})^D C_1^{\frac{T}{\eta}} \max\{q_1^{-\delta(T-j)}, q_2^{-\delta}\} \leq C_2^{D+\frac{T}{\eta}} \max\{q_1^{-\delta(T-j)}, q_2^{-\delta}\}$$

where $C_2 = \max\{\overline{C} + \overline{\overline{C}}, C_1\}$. □

29. Lecture 29: Thue's theorem - part II

In this lecture we will conclude the proof of Thue's theorem.

29.1. **The Wronskian & conclusion of the proof of Thue's theorem.** The notation introduced in 28.2.1 will be in force throughout this lecture. We also fix an integer $T \geq 1$ we set $D = \left\lfloor \frac{dT(1+\eta)}{2} \right\rfloor$ as in the last lecture. We have proved the existence of a non-zero polynomial $A(X,Y) = P(X) - YQ(X) \in \mathbb{Z}[X,Y]$ with the following properties:

(1) $\deg P \leq D, \deg Q \leq D$.
(2) $\max\{\|P\|, \|Q\|\} \leq C_1^{\frac{T}{\eta}}$.
(3) $\frac{\partial^j A}{\partial X^j}(\alpha,\alpha) = 0$ for $0 \leq j \leq T-1$.
(4) For $0 \leq j \leq \frac{T}{2}$, $|D^j P(a_1) - a_2 D^j Q(a_1)| \leq \max\{q_1^{-\delta(T-j)}, q_2^{-\delta}\} C_2^{D+\frac{T}{\eta}}$.

To proceed further, we need an estimate in the other direction, i. e. a lower bound on $|D^j P(a_1) - a_2 D^j Q(a_1)|$; in particular, we need to find j such that this quantity is non-zero. This *non-vanishing* step is the key point, and often the hardest one to establish, in the proof of several results in Diophantine approximation and transcendence theory. To perform this step in our situation we will make use of some basic properties of the *Wronskian*.

Definition 29.1.1. *The Wronskian of P, Q is the polynomial*

$$W(X) = P^{(1)}(X)Q(X) - P(X)Q^{(1)}(X).$$

The key property of the Wronskian is that $W = 0$ if and only if P, Q differ by multiplication by a constant. Indeed, the derivative of $\frac{P}{Q}$ is $\frac{W}{Q^2}$, hence it vanishes if $W = 0$. It follows that $P = \lambda Q$ for some $\lambda \in \mathbb{Q}$.

For $j \geq 0$, the j-th derivative of the Wronskian is given by the following formula, which can be proved by induction:

$$W^{(j)} = \sum_{i=0}^{j} \binom{j}{i} (P^{(j-i)}Q^{(i+1)} - Q^{(j-i)}P^{(i+1)}). \tag{29.1.1.1}$$

29.1.2. *Step 3: the non-vanishing step.* There is a constant C_3 such that

$$D^j P(a_1) - a_2 D^j Q(a_1) \neq 0 \text{ for some } j \leq 1 + C_3 \frac{T}{\eta \log q_1}.$$

Proof. We learn from (29.1.1.1) that, given an integer $h \geq 0$, *if* $W^{(h)}(a_1) \neq 0$ then $P^{(j)}Q^{(i)}(a_1) - Q^{(j)}P^{(i)}(a_1) \neq 0$ for some $1 \leq i, j \leq h+1$, hence $P^{(j)}(a_1) - a_2 Q^{(j)}(a_1) \neq 0$ for some $1 \leq j \leq h+1$. Therefore we need to bound above the order of vanishing $\mathrm{ord}_{a_1}(W)$ of W at a_1. First of all, let us observe that W is non-zero. Indeed, if $W = 0$ then $P = \lambda Q$ for some $\lambda \in \mathbb{Q}$, as observed above. This yields $A(X,Y) = P(X) - YQ(X) = Q(X)(\lambda - Y)$; by construction we have $\frac{\partial^j A}{\partial X^j}(\alpha,\alpha) = 0$ for $j \leq T-1$, hence we deduce that $(X-\alpha)^T$ divides $Q(X)$. It follows that $(P_\alpha^{\min})^T \mid Q$, hence $\deg Q \geq dT$. But $dT > \lfloor \frac{dT(1+\eta)}{2} \rfloor = D \geq \deg Q$, and we find a contradiction.

We have proved that $W \neq 0$. We now claim that there is a constant C_3 such that

$$\mathrm{ord}_{a_1}(W) \leq C_3 \frac{T}{\eta \log q_1}.$$

We will now prove this claim, which completes the execution of Step 3. Take $t \geq 1$ such that $(X-a_1)^t \mid W(X)$. We have $a_1 = \frac{p_1}{q_1}$ and we may assume that p_1, q_1 are coprime, hence the polynomial $(q_1 X - p_1)^t \in \mathbb{Z}[X]$ is primitive (i. e. the greatest common divisor of its coefficients is one). Hence (a version of) Gauss lemma implies that $(q_1 X - p_1)^t$ divides $W(X)$ in $\mathbb{Z}[X]$. In particular q_1^t divides the leading coefficient of W, which is *non-zero* because $W \neq 0$. Therefore $\|W\| \geq q_1^t$; on the other hand:

$$\|W\| \leq \|P^{(1)}Q\| + \|PQ^{(1)}\| \leq 2D(D+1)\max\{\|P\|, \|Q\|\}^2 \leq 2D(D+1)C_1^{\frac{2T}{\eta}} \leq \breve{C}^{\frac{T}{\eta}}$$

for a suitable constant $\breve{C}$. So

$$t \log q_1 \leq \frac{T}{\eta} \log \breve{C} \Rightarrow \mathrm{ord}_{a_1}(W) \leq \log \breve{C} \frac{T}{\eta \log q_1}$$

and we can take $C_3 = \log \breve{C}$. □

29.1.3. *Step 4: a lower bound.* If $D^j P(a_1) - a_2 D^j Q(a_1) \neq 0$ then

$$|D^j P(a_1) - a_2 D^j Q(a_1)| \geq \frac{1}{q_1^{D-j} q_2}.$$

Indeed $D^j P$ and $D^j Q$ have degree at most $D - j$ and integer coefficients; once again, the estimate follows from the fact that a non-zero integer has absolute value at least one.

29.1.4. *Step 5: comparing upper and lower bound.* Now assume by contradiction that (28.1.1.1) has infinitely many solutions. Then, for any chosen η, we can pick $a_1 = \frac{p_1}{q_1}$ with q_1 large enough to guarantee that $1 + C_3 \frac{T}{\eta \log q_1} \leq \frac{T}{2}$ for any $T \geq 3$. By Step 3 we know that there is $0 \leq j \leq 1 + C_3 \frac{T}{\eta \log q_1}$ such that $D^j P(a_1) - a_2 D^j Q(a_1) \neq 0$. Combining Step 2 and Step 4 we obtain

$$\frac{1}{q_1^{D-j} q_2} \leq \max\{q_1^{-\delta(T-j)}, q_2^{-\delta}\} C_2^{D+\frac{T}{\eta}}. \tag{29.1.4.1}$$

We now pick $a_2 = \frac{p_2}{q_2}$ with $q_2 \geq q_1^3$, and we set $T = \lfloor \frac{\log q_2}{\log q_1} \rfloor$, so that $q_1^T \leq q_2 \leq q_1^{T+1}$. It follows that

$$q_2^{-\delta} \leq q_1^{-\delta T} \leq q_1^{-\delta(T-j)}$$

hence (29.1.4.1) yields

$$\frac{1}{q_1^{D-j+T+1}} \leq q_1^{-\delta(T-j)} C_2^{D+\frac{T}{\eta}} \Rightarrow -D + j - T - 1 \leq -\delta(T-j) + \left(D + \frac{T}{\eta}\right) \frac{\log C_2}{\log q_1}.$$

We have $D \leq \frac{dT(1+\eta)}{2}$ by definition, hence we obtain:

$$-dT\frac{1+\eta}{2} + j - T - 1 \leq -\delta(T-j) + \left(dT\frac{1+\eta}{2} + \frac{T}{\eta}\right) \frac{\log C_2}{\log q_1}.$$

The last term on the right is bounded by $\frac{T}{\eta \log q_1} C_4$ for some constant C_4, hence we obtain

$$T\left(-\frac{d}{2} - \frac{d\eta}{2} - 1 + \delta - \frac{C_4}{\eta \log q_1}\right) \leq j\delta - j + 1 \leq j\delta + 1.$$

Now we have $j \leq 1 + C_3 \frac{T}{\eta \log q_1}$, hence

$$\begin{aligned} &T\left(-\frac{d}{2} - \frac{d\eta}{2} - 1 + \delta - \frac{C_4}{\eta \log q_1}\right) \leq \left(1 + C_3 \frac{T}{\eta \log q_1}\right)\delta + 1 \\ \Rightarrow &T\left(-\frac{d}{2} - \frac{d\eta}{2} - 1 + \delta - \frac{C_4}{\eta \log q_1} - \delta\frac{C_3}{\eta \log q_1}\right) \leq \delta + 1. \end{aligned}$$

Finally, as $\delta > 1 + \frac{d}{2}$ we can first pick η small enough to ensure that $\delta - \frac{d}{2} - \frac{d\eta}{2} - 1 > 0$. Then we can choose q_1 large enough to guarentee that the requirement at the beginning of Step 5 is satisfied, and in addition the quantity in parentheses in the last inequality is strictly positive. Then as q_2 goes to infinity so does T, hence we obtain a contradiction.

29.2. **Final remarks on Thue's theorem.**

29.2.1. *Non-effectiveness.* As you may have noticed, while the basic idea to prove Thue's theorem resembles the approach we have used for Liouville's theorem, Thue's result is much harder to prove. More seriously, there is a *crucial* change in what is happening in the last step of the proof of the two theorems: examining the argument used for Liouville's theorem you can check that it gives an explicit *upper bound* on the denominator of a solution of the inequality

$$\left|\frac{p}{q} - \alpha\right| < \frac{1}{|q|^\delta} \tag{29.2.1.1}$$

for $\delta > d$. In particular, this allows in principle to find all the solutions. On the contrary, *no such a bound* can be extracted from the argument we gave for Thue's theorem; in other words, the proof described above is *not effective*. The reason is that what we really showed is the following: *if we know that there is a solution* $a_1 = \frac{p_1}{q_1}$ of the inequality (29.2.1.1) with q_1 large enough - so that all the requirements in Step 5 are satisfied - *then we can bound* the size of the denominator

of *another* solution $a_2 = \frac{p_2}{q_2}$ with q_2 sufficiently larger than q_1. While this is certainly enough to prove finiteness of the set of rational numbers verifying (29.2.1.1), it does *not* allow us to find the solutions in practice. The problem is that to get an explicit bound we first need to know the existence of one solution $a_1 = \frac{p_1}{q_1}$ with q_1 large enough; we can start looking for such a solution, but if we have not found it yet we cannot be sure whether it exists or not, and the above argument is of no use to determine the solutions of (29.2.1.1). For the same reason, while in Corollary 28.1.2 we deduced from Thue's theorem that Mordell equations have finitely many integral solutions, our arguments do not allow us to solve these equations in practice.

29.2.2. *Further developements.* Thue's groundbreaking work inspired several mathematicians who refined his technique and improved his result over the years: Siegel showed that $e(\alpha) \leq 2\sqrt{d}$ for every real algebraic number α of degree $d > 1$. To do this he used auxiliary polynomials $A(X, Y)$ with arbitrary degree in Y. Gelfand and Dyson were able to replace $2\sqrt{d}$ by $\sqrt{2d}$. Finally, Roth - first British Fields medal and professor at Imperial College from 1966 [14] - was able to prove in 1955 that $e(\alpha) = 2$ for *every* algebraic irrational number! He was awarded the Fields medal in 1958 for this astonishing result; in the words of Davenport, who presented Roth's work at the International Congress of Mathematicians in 1958, "The achievement is one that speaks for itself; it closes a chapter, and a new chapter will now be opened. Roth's theorem settles a question which is both of a fundamental nature and of extreme difficulty. It will stand as a landmark in mathematics for as long as mathematics is cultivated."

Finally, let us mention that the theorems quoted above are all ineffective; an effective proof of finiteness of the set of integral solutions of Thue equations was given, by completely different methods, by Baker in the 70s.

[14]Biographical note for London people: Roth was actually born in Prussia, but moved to London as a child to escape the Nazis. He became a pupil of the St. Paul's school in Barnes, just south of Hammersmith Bridge. He later did his Master and PhD at UCL.

30. Lecture 30: Problem session V

(1) Let K be a number field with ring of integers $\mathcal{O}_K$. The aim of this exercise is to prove the following result:

Theorem 30.0.1. *If a prime number p is ramified in K then p divides $\Delta(\mathcal{O}_K)$. In particular, only finitely many primes ramify in K.*

Let us remark that the implication in the above theorem is actually an equivalence, although we shall not prove this. I am grateful to Tony Scholl for pointing out to me the following argument (which seems to be due to Hecke).

(a) Let $\alpha \in \mathcal{O}_K$. Prove that $Tr(\alpha^p) \equiv Tr(\alpha)^p \equiv Tr(\alpha) \pmod{p}$.
(b) Let p be a prime number which ramifies in K. Write $p\mathcal{O}_K = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_d^{e_d}$; we may (and will) assume that $e_1 > 1$. Show that there exists $\alpha \in \mathfrak{p}_1^{e_1-1}\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_d^{e_d} \smallsetminus p\mathcal{O}_K$.
(c) Let α be an element in $\mathfrak{p}_1^{e_1-1}\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_d^{e_d} \smallsetminus p\mathcal{O}_K$. Prove that, for every $\beta \in \mathcal{O}_K$, the congruence $Tr(\alpha\beta) \equiv 0 \pmod{p}$ holds.
(d) Let $e_1, \ldots, e_n$ be a $\mathbb{Z}$-basis of $\mathcal{O}_K$. Write $\alpha = a_1e_1 + \ldots + a_ne_n$ with $a_1, \ldots, a_n \in \mathbb{Z}$. Show that there exists $1 \leq i \leq n$ such that $p \nmid a_i$. Deduce that the matrix whose $(i,j)-$th entry is $Tr(e_ie_j)$, when reduced modulo p, has rank strictly smaller than n. Conclude.

(2) Dirichlet proved (by analytic means) that if a and b are coprime integers then the arithmetic progression $an + b, n \in \mathbb{Z}$ contains infinitely many prime numbers. In this exercise we will prove the following special case of Dirichlet's result:

Theorem 30.0.2. *Let l be a prime. There exist infinitely many prime numbers p such that $p \equiv 1 \pmod{l}$.*

(a) Let $P(X) \in \mathbb{Z}[X]$ be a non constant polynomial. Let $S_P = \{p \text{ prime} : \exists\, n \in \mathbb{Z}, p \mid P(n)\}$. Prove that S_P is infinite (try to adapt Euclid's argument).
(b) Let $n \in \mathbb{Z}$ and let p be a prime such that $p \mid \Phi_l(n)$. Show that the order of n in $\mathbb{F}_p^\times$ is either 1 or l.
(c) Conclude.
(d) Deduce that there are infinitely many primes which split completely in $\mathbb{Q}(\zeta_l)$.

(3) In this exercise we study bounds for the class number of imaginary quadratic fields. First of all we will show that there are imaginary quadratic fields whose ring of integers has arbitrarily large class number. I thank Daniel Kohen for suggesting the following argument.

(a) Let $k > 0$ be a squarefree integer which is congruent to 7 modulo 8. Let A be the ring of integers of $\mathbb{Q}(i\sqrt{k})$. Prove that there exists an ideal $\mathfrak{p} \subset A$ such that $N(\mathfrak{p}) = 2$.
(b) Fix an ideal $\mathfrak{p}$ as in the previous point. If $\frac{k}{4} > 2$ prove that $\mathfrak{p}$ is not principal. Deduce that $h(A) \geq 2$.
(c) More generally, if $\frac{k}{4} > 2^g$ then show that none of the ideals $\mathfrak{p}, \mathfrak{p}^2, \ldots, \mathfrak{p}^g$ can be principal. Deduce that $h(A) \geq g + 1$.
(d) What happens if you try to adapt the previous strategy to the case of real quadratic fields?

We now switch to upper bounds for the class number: prove that there exists a real number $C > 0$ such that, for every imaginary quadratic field K, the inequality $h(\mathcal{O}_K) \leq C\Delta(\mathcal{O}_K)$ holds.[15]

(4) Let A be a Dedekind domain with finitely many prime ideals. Prove that A is a principal ideal domain.

Now recall your favourite example of ring of integers of a number field which is not a PID, and deduce that there are infinitely many prime numbers.

(5) Let p be a prime number and $K = \cup_{n\geq 1}\mathbb{Q}(\zeta_{p^k})$. Let $A = \cup_{n\geq 1}\mathbb{Z}[\zeta_{p^k}] \subset K$.
(a) Is A noetherian? Is it a UFD?
(b) Let $\phi : A \to A$ be the map sending x to x^p. Is ϕ a ring homomorphism?
(c) Show that ϕ induces a surjective ring homomorphism $\bar{\phi} : A/pA \to A/pA$.

[15] Much better estimates can be proved, but they require analytic methods which we have not discussed.

(6) In this exercise we complete the proof of the first case of Fermat's last theorem for a regular prime $p > 3$. Assume by contradiction that there are $x, y, z \in \mathbb{Z}$ such that $p \nmid xyz$ and $x^p + y^p = z^p$. Let $K = \mathbb{Q}(\zeta_p)$.
   (a) Show that the only roots of unity in K are of the form $\pm\zeta_p^a, a \in \mathbb{Z}$.
   (b) Let $u \in \mathcal{O}_K^\times$ and $\tau = \frac{u}{\bar{u}}$; show that $\tau = \pm\zeta_p^a$ for some $a \in \mathbb{Z}$.
   (c) Let $\lambda = 1 - \zeta_p$. Show that for every field automorphism $\rho : K \to K$ the congruence $\rho(u) \equiv u \pmod{\lambda\mathcal{O}_K}$ holds.
   (d) Deduce that $\tau = \zeta_p^a$ for some $a \in \mathbb{Z}$, hence $u = \zeta^{-2b}\bar{u}$ for some $b \in \mathbb{Z}$.
   (e) Deduce that $\zeta^b u$ is a real number.
   (f) Show that there exist $u \in \mathcal{O}_K^\times \cap \mathbb{R}$, $0 \le b \le p-1$ and $\beta \in \mathcal{O}_K$ such that $x + \zeta_p y = \zeta_p^b u\beta$ and $\beta \equiv n \pmod{p\mathcal{O}_K}$ for some $n \in \mathbb{Z}$.
   (g) Prove that

$$\zeta_p^{-b}(x + \zeta_p y) \equiv \zeta_p^b(x + \zeta_p^{-1} y) \pmod{p\mathcal{O}_K}, \text{ hence } x + \zeta_p y - \zeta_p^{2b} x - \zeta_p^{2b-1} y \in p\mathcal{O}_K.$$

   (h) Show that either (a) $\zeta_p^{2b} = 1$, or (b) $\zeta_p^{2b-1} = 1$, or (c) $\zeta_p^{2b+1} = 1$.
   (i) Prove that cases $(a), (c)$ lead to a contradiction.
   (j) In case (b), show that $x \equiv y \pmod{p}$. Write the Fermat equation as $x^p + (-z)^p = (-y)^p$ and repeat the argument above; deduce that $x \equiv -z \pmod{p}$, and obtain a contradiction.

31. The Diophantine equation $aY^2 + bY + c = dX^n$

We end these notes discussing the work [19]. Using Thue's theorem, Landau and Ostrowski were able to prove in 1919 a general finiteness result for integral solutions of a class of Diophantine equations including Mordell equations. The proof relies on the arithmetic properties of quadratic rings studied in this course.

31.1. **The theorem of Landau and Ostrowski.**

Theorem 31.1.1. *([19]) Let $n \geq 3$ be an integer. Let $a, b, c, d \in \mathbb{Z}$ be integers such that $a, d \neq 0$ and $b^2 - 4ac \neq 0$. The equation*

$$aY^2 + bY + c = dX^n$$

has finitely many integral solutions.

In particular, for $a = d = 1$, $b = 0$ and $n = 3$ we obtain finiteness of integral solutions of Mordell equations. The theorem has other interesting consequences. For example you can deduce from it the following result, originally proved by Polya.

Corollary 31.1.2. *Let a, b, c be integers such that $a \neq 0$ and $b^2 - 4ac \neq 0$. The greatest prime factor of $ay^2 + by + c$ tends to infinity as the absolute value of $y \in \mathbb{Z}$ goes to infinity.*

31.1.3. Let us start the proof of Theorem 31.1.1 observing that we can rewrite the equation as

$$(2aY + b)^2 - (b^2 - 4ac) = 4adX^n.$$

Hence we may (and will) restrict to the study of the equation

$$Y^2 - k = lX^n \tag{31.1.3.1}$$

with $k, l \neq 0$ and $n \geq 3$. Let $(x, y) \in \mathbb{Z}^2$ be a solution of the above equation. We write, as usual

$$(y - \sqrt{k})(y + \sqrt{k}) = lx^n;$$

now the idea is to use unique factorisation in $\mathbb{Z}$ (if k is a square) or $\mathbb{Z}[\sqrt{k}]$ to reduce the above equation to a finite number of Thue equations, i. e. equations of the form $P_i(X, Y) = a_i$ where $P_i \in \mathbb{Z}[X, Y]$ is a homogeneous polynomial and $a_i \in \mathbb{Z} \smallsetminus \{0\}$. Thue's theorem then implies that such an equation has finitely many integral solutions as long as P_i is not a power of a homogeneous linear or quadratic polynomial - this is a slight improvement of Corollary 28.1.2.

31.2. **Back to Mordell equations.** Our strategy will be close in spirit to the one used to prove Corollary 16.3.3 concerning the Mordell equation $Y^2 + k = X^3$ for $k > 0$ squarefree congruent to 1, 2 modulo 4, assuming that $3 \nmid h(\mathbb{Z}[i\sqrt{k}])$. In that setting, given $(x, y) \in \mathbb{Z}^2$ such that $y^2 + k = x^3$, we argued as follows:

(1) Using the hypothesis that k is squarefree and congruent to 1, 2 modulo 4, we proved that $(y + i\sqrt{k}, y - i\sqrt{k}) = \mathbb{Z}[i\sqrt{k}]$, hence $y + i\sqrt{k} = \mathfrak{a}^3$ for some ideal $\mathfrak{a} \subset \mathbb{Z}[i\sqrt{k}]$.
(2) Using the assumption that $3 \nmid h(\mathbb{Z}[i\sqrt{k}])$ we deduced that $y + i\sqrt{k} = \varepsilon\alpha^3$ for some $\alpha \in \mathbb{Z}[i\sqrt{k}]$ and some $\varepsilon \in \mathbb{Z}[i\sqrt{k}]^\times$.
(3) As $k > 0$ we were able to absorb units in the cube, and deduce that $y + i\sqrt{k} = (a + bi\sqrt{k})^3$ for some $a, b \in \mathbb{Z}$. This yields a Thue equation involving a reducible polynomial of degree 3, which we explicitly solved.

The key observation is that one can in fact reduce a Mordell equation to a finite number of Thue equations *without* the assumptions used in the previous points. Let us explain the rough ideas, which we will then apply in the more general context of Theorem 31.1.1.

(1) The assumption that k is squarefree and congruent to 1, 2 modulo 4 is unnecessary: we can always work with the ring of integers $\mathcal{O}_K \subset K = \mathbb{Q}(\sqrt{-k})$. Even if the elements $y \pm \sqrt{-k}$ are not coprime, there are only finitely many ideals containing both, hence we will get finitely many equations of the form $(y + \sqrt{-k}) = \mathfrak{b}\mathfrak{a}^3$, where $\mathfrak{a}, \mathfrak{b}$ are ideals of $\mathcal{O}_K$.
(2) The assumption that $3 \nmid h(\mathcal{O}_K)$ can be removed: the ideal $\mathfrak{a}$ will perhaps not be principal, but as the class group is finite there are finitely many possibilities for $\mathfrak{a}$ up to multiplication by a principal fractional ideal. A computation in the class group will give rise to equations of the form $s(y + \sqrt{-k}) = \varepsilon\beta\alpha^3$ for suitable $\beta \in \mathcal{O}_K$ and $s \in \mathbb{Z}$ - varying among finitely many possilities - and for some $\varepsilon \in \mathcal{O}_K^\times$.

(3) If $k<0$ then $\mathcal{O}_K^\times$ is infinite, and we cannot absorb the unit ε into the cube. However we know that $\mathcal{O}_K^\times \simeq \{\pm 1\}\times\mathbb{Z}$, hence up to multiplication by a cube there are three choices for ε. So we obtain finitely many equations of the form $s(y+\sqrt{-k})=\gamma\delta^3$.
(4) Comparing each equation in the previous point with the one obtained applying the involution $a+b\sqrt{-k}\mapsto a-b\sqrt{-k}$ we obtain a Thue equation.

31.3. **Proof of theorem 31.1.1.**

31.3.1. *The equation $Y^2-m^2=lX^n$.* I will explain in detail how to show that equation (31.1.3.1) has finitely many integral solutions when k is a square. The argument relies on unique factorisation in $\mathbb{Z}$ and Thue's theorem. After that I will outline the proof for k not a square: it is an exercise using all the arithmetic properties of quadratic rings you learned so far, so you may find it instructive to fill in the details.

Let $m,l\neq 0$ and $n\geq 3$ be integers. We want to show that the equation $Y^2-m^2=lX^n$ has finitely many integral solutions. Let $(x,y)\in\mathbb{Z}^2$ be such that $(y-m)(y+m)=lx^n$ and $x\neq 0$. The greatest common divisor of $y+m$ and $y-m$ divides $2m$. Hence a prime dividing $y+m$ but not dividing $2lm$ must appear in the factorisation of $y+m$ with an exponent which is a multiple of n. Calling $p_1,\dots,p_r$ the distinct prime factors of $2lm$ we deduce that we can write $y+m=\pm p_1^{e_1}\cdots p_r^{e_r}a^n$ for some $e_1,\dots,e_r\in\mathbb{Z}_{\geq 0}$ and some $a\in\mathbb{Z}$. Absorbing n-powers we obtain equations of the form

$$y+m=bt^n,\ b\in I$$

where $I\subset\mathbb{Z}$ is a finite set. The same argument shows that $y-m$ satisfies an equation of the form $y-m=cu^n$, with the integer c belonging to a finite set J. Now for each $b\in I$ and $c\in J$ we obtain

$$2m=bt^n-cu^n.$$

The polynomial bT^n-cU^n is not a power of a linear or quadratic polynomial (otherwise setting $U=1$ we would obtain a polynomial with repeated roots) hence by Thue's theorem each of the above equations has finitely many integral solutions, and the same is true for our original equation.

31.3.2. *The equation $Y^2-k=lX^n$ for k not a square.* Now assume that k is not a square; let $K=\mathbb{Q}(\sqrt{k})$, endowed with the involution $\alpha\mapsto\alpha'$ sending $\sqrt{k}$ to $-\sqrt{k}$. Let $\mathcal{O}_K$ be the ring of integers of K. If $\mathfrak{a},\mathfrak{b}\subset\mathcal{O}_K$ are two non-zero ideals giving rise to the same element in the class group of $\mathcal{O}_K$ we write $\mathfrak{a}\sim\mathfrak{b}$. Hence $\mathfrak{a}\sim\mathfrak{b}$ if and only if there exist $\alpha,\beta\in\mathcal{O}_K\smallsetminus\{0\}$ such that $\alpha\mathfrak{a}=\beta\mathfrak{b}$. Furthermore in this case α can be taken to be a non-zero integer.

Let $(1,\tau)$ be a basis of $(\mathcal{O}_K,+)$. Take $x,y\in\mathbb{Z}$ such that $y^2-k=lx^n$ and notice that $x\neq 0$.

Step 1: Let $\mathfrak{p}_1,\dots,\mathfrak{p}_r$ be the prime ideal factors of $(2\sqrt{kl})$. There exist nonnegative integers $e_1,\dots,e_r\in\mathbb{Z}_{\geq 0}$ and an ideal $\mathfrak{a}\subset\mathcal{O}_K$ such that

$$(y+\sqrt{k})=\mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}\mathfrak{a}^n.$$

Hence there exists a finite set I of ideals of $\mathcal{O}_K$ such that

$$(y+\sqrt{k})=\mathfrak{b}\mathfrak{t}^n \text{ for some } \mathfrak{b}\in I,\mathfrak{t}\subset\mathcal{O}_K.$$

Step 2: Let $\mathfrak{c}_1,\dots,\mathfrak{c}_h$ be ideals giving a set of representatives of the class group of $\mathcal{O}_K$. For each $\mathfrak{c}_i$ let $\mathfrak{d}_i$ be an ideal representing the inverse of $\mathfrak{c}_i$ in the class group. According to the image of $\mathfrak{t}$ in the class group we obtain finitely many equations

$$(y+\sqrt{k})=\mathfrak{b}\mathfrak{t}^n \text{ for some } \mathfrak{b}\in I,\mathfrak{t}\sim\mathfrak{c}_i.$$

This implies that $\mathfrak{b}\sim\mathfrak{d}_i^n$ hence $d\mathfrak{b}=\delta\mathfrak{d}_i^n$ for some $d\in\mathbb{Z}\smallsetminus\{0\}$ and $\delta\in\mathcal{O}_K\smallsetminus\{0\}$. We choose such d and δ and write each of the above equations as

$$(d(y+\sqrt{k}))=(\delta\gamma^n) \text{ for some } \gamma\in\mathcal{O}_K.$$

We deduce that $d(y+\sqrt{k})=\varepsilon\delta\gamma^n$ for some $\varepsilon\in\mathcal{O}_K^\times$.

Step 3: As $\mathcal{O}_K^\times$ is finitely generated each of the above equation gives rise to finitely many equations of the form $d(y+\sqrt{k})=\zeta\gamma^n$.

Step 4: Now write $\gamma = a + b\tau$ with $a, b \in \mathbb{Z}$. We find

$$d(y + \sqrt{k}) = \zeta\gamma^n, \;\; d(y - \sqrt{k}) = \zeta'(\gamma')^n \Rightarrow \frac{2d\sqrt{k}}{\tau - \tau'} = \frac{\zeta(a + b\tau)^n - \zeta'(a + b\tau')^n}{\tau - \tau'}.$$

The left hand side of the above equation is a non-zero integer, and the right hand side is a homogeneous polynomial in a, b of degree n with integer coefficients. Furthermore it is not a power of a linear or quadratic homogeneous polynomial, hence each equation has finitely many integral solutions, and the theorem is proved.

Epilogue

We have learned how the arithmetic properties of number fields can be used to solve certain Diophantine equations. In particular we have seen that unique factorisation can be restored in rings of integers of number fields introducing *ideals*, and that *finiteness* of the ideal class group allows us to control failure (in the naive sense) of unique factorisation. In some cases, this enabled use to prove general theorems on the solutions of Diophantine equations (e. g. Mordell equations). For example, you are now able to solve the Diophantine equation

$$Y^2 = X^3 - 2021.$$

We discovered that *quadratic reciprocity* is a powerful tool to solve Diophantine equations. It also gives us non-trivial information about representation of integers by binary quadratic forms, and it is intimately related to the behaviour of primes in quadratic fields.

Finally, we studied *Diophantine approximation* and some of its applications to Diophantine equations, e. g. Pell and Thue equations.

We have also sometimes asked ourselves questions to which we couldn't answer - and bumped into open questions as well. Number Theory is alive and well!

Let me end this text with some questions which naturally emerge from what we have learned, and will hopefully lead you to learn or discover something interesting.

(1) We have solved several Mordell equations $Y^2 = X^3 + k$, but by no means all of them! How do we find the integer solutions of an arbitrary Mordell equation? For example, what about the equation $Y^2 = X^3 - 26$? And what about rational solutions? What can we say of the solutions of equations of the form $Y^2 = P(X)$ where $P(X) \in \mathbb{Z}[X]$ is an arbitrary polynomial?

(2) Similarly, we have solved equations of the form $p = X^2 + nY^2$ for certain values of n. What happens for general n? And what if we look at quadratic forms in more variables?

(3) We have studied *algebraic numbers* and their approximation by rational numbers throughout the course. What about *transcendental numbers*?

(4) We have seen that the quadratic reciprocity law governs the behaviour of primes in quadratic fields; concretely, it exhibits an unexpected regularity in the set of primes p such that a given quadratic polynomial with integral coefficients has a root modulo p. What about higher degree polynomials? Shouldn't there be a more general reciprocity law? What should its connection with the arithmetic of number fields be? In fact, what is a reciprocity law?

References

[1] Vladimir I. Arnold, *Dynamics, Statistics and Projective Geometry of Galois Fields* , Cambridge University Press, 2010
[2] Michael A. Bennett, Amir Ghadermarzi, *Mordell's equation: a classical approach*, available here
[3] B. J. Birch, J. R. Merriman, *Finiteness theorems for binary forms with given discriminant*, Proc. London Math. Soc., 1972, available here
[4] Dustin Clausen, *p-adic J-homomorphisms and a product formula*, available here
[5] Chris J. Conidis, Pace P. Nielsen, Vandy Tombs, *Transfinitely valued Euclidean domains have arbitrary indecomposable order type*, available here
[6] Keith Conrad, *Remarks about Euclidean domains*, available here
[7] Keith Conrad, *The Solovay-Strassen test*, available here
[8] William A. Coppel, *Number Theory: an introduction to Mathematics*, Springer, 2006
[9] David A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication*, Wiley, 2013
[10] Richard Dedekind, *Theory of algebraic integers*, Cambridge University Press, 2008
[11] Carl Friedrich Gauss, *Disquisitiones Arithmeticae*, Yale University Press, 1965
[12] Allen Hatcher, *Topology of numbers*, available here
[13] Roger Heath-Brown, *Primes represented by $x^3 + 2y^3$*, Acta Math., Vol. 186, n. 1, 2001
[14] David Hilbert, *The theory of algebraic number fields*, Springer, 1998
[15] Marc Hindry, *Arithmetics*, Springer, 2011
[16] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, Springer (2nd edition), 1998
[17] Ernst Kummer, *Extrait d'une Lettre de M. Kummer à M. Liouville*, Journal de Mathématiques Pures et Appliquées, vol. 12, 1847, available here
[18] Ernst Kummer, *Sur les nombres complexes qui sont formés avec les nombres entiers réels et les racines de l'unité*, Journal de Mathématiques Pures et Appliquées, vol. 12, 1847, available here
[19] Edmund Landau, Alexander Ostrowski, *On the diophantine equation $aY^2 + by + c = dx^n$*, Proceedings of the London Mathematical Society, 1921
[20] Joseph Liouville, *Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques*, available here
[21] Daniel A. Marcus, *Number fields*, Springer, 1995
[22] Yuri Matiyasevich, *Hilbert's tenth problem: what can we do with Diophantine equations?*, available here
[23] Louis J. Mordell, *Diophantine equations*, Diophantine equations, Academic Press, 1969
[24] Louis J. Mordell, *The infinity of rational solutions of $y^2 = x^3 + k$* , Journal of the London Mathematical Society, Volume s1-41, Issue 1, 1966
[25] David Mumford, *The red book of varieties and schemes*, Springer, 1988
[26] Jan Nekovář, *Number theory*, notes for a course at UPMC, available here
[27] I. Niven, *A simple proof that π is irrational*, Bull. Amer. Math. Soc., vol. 53, n. 6, 1947
[28] Bjorn Poonen, *Undecidability in number theory*, Notices Amer. Math. Soc. 55, 2008
[29] Pierre Samuel, *Algebraic theory of numbers*, Dover Publications, 1970
[30] Pierre Samuel, *Unique factorization*, The American Mathematical Monthly, 1968, available here
[31] Carl L. Siegel, *Lectures on the Geometry of Numbers*, Springer, 1989
[32] Joseph H. Silverman, John Tate, *Rational points on elliptic curves*, Springer, 1994
[33] Christophe Soulé, *An introduction to arithmetic groups*, available here
[34] Elias S. Stein, Rami Shakarchi, *Complex analysis*, Princeton University Press, 2003
[35] Alexander Stasinski, *Finiteness of the class group of basic arithmetic rings*, preprint, 2019, available here
[36] Mak Trifkovic, *Algebraic theory of quadratic numbers*, Universitext, Springer, 2013.
[37] Lawrence C. Washington, *Introduction to cyclotomic fields*, Springer, GTM 83, 1996
[38] André Weil, *On analogy in Mathematics*, Notices of the AMS, vol. 52, 1940, available here
[39] Don Zagier, *A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares*, The American Mathematical Monthly, Vol. 97, n. 2, 1990