Franz Lemmermeyer

Algebraic Number Theory

December 19, 2005

Contents

1. Fermat, Euler, and Nonunique Factorization

The purpose of this chapter is mainly motivational. More precise definitions of the objects we will study will be given later.

1.1 Euler and Quadratic Irrationals

Algebraic number theory was born when Euler used algebraic numbers to solve diophantine equations such as $y^2 = x^3 - 2$: Fermat had claimed that $(x, y) = (3, 5)$ is the only solution in natural numbers, and Euler gave a "proof" by writing

$$x^3 = y^2 + 2 = (y - \sqrt{-2}\,)(y + \sqrt{-2}\,) \tag{1.1}$$

and working with the ring $\mathbb{Z}[\sqrt{-2}\,] = \{a + b\sqrt{-2} : a, b \in \mathbb{Z}\}$.

The problem with Euler's idea was that he did not justify all of his claims. Arguing that the two factors on the right hand side of (1.1) were coprime[1], he concluded that each factor had to be a perfect cube,[2] i.e. that $y + \sqrt{-2} = (a + b\sqrt{-2}\,)^3$ for certain $a, b \in \mathbb{Z}$. Comparing real and imaginary parts yields $y = a^3 - 6ab^2 = a(a^2 - 6b^2)$ and $1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$. The last equation tells us that $b \mid 1$, hence $b = \pm 1$. Moreover, $3a^2 - 2b^2 = 1$, hence $a = \pm 1$. Plugging these solutions into $y = a(a^2 - 6b^2)$ shows that $y = \pm 5$ and thus $x = 3$, proving Fermat's claim.

In order to understand why Euler's argument is not sufficient, let us consider the diophantine equation $y^2 = x^2 - 5$. Imitating Euler's proof, we find

$$x^2 = y^2 + 5 = (y - \sqrt{-5}\,)(y + \sqrt{-5}\,). \tag{1.2}$$

Since the two factors are "coprime", both of them must be squares; but from $y + \sqrt{-5} = (a + b\sqrt{-5}\,)^2$ we get the equation $1 = 2ab$, which does not have any solutions in integers. This seems to suggest that $y^2 = x^2 - 5$ does not have any integral solutions; but actually $(x, y) = (3, 2)$ is one.[3]

[1] Here is the first problem: he does not really define what this means.

[2] This is the second problem: Euler knows that this argument works inside the natural numbers; in fact a proof can be found in Euclid's elements. But Euler does not explain why this should work in $\mathbb{Z}[\sqrt{-2}\,]$.

[3] Of course there is no need to invoke algebraic numbers for solving $y^2 = x^2 - 5$, because we can write the equation in the form $5 = x^2 - y^2 = (x - y)(x + y)$.

1.2 Fermat and the Harbingers of Nonunique Factorization

The examples above suggest the following question: why does an argument that works well for numbers of the form $a + b\sqrt{-2}$ go wrong for numbers of the form $a + b\sqrt{-5}$? The reason for this strange behavior would not be uncovered until the mid-19th century, although traces of it can be tracked back to the work of Fermat. One of his more famous theorems claims that every prime of the form $4n + 1$ can be written as the sum of two squares. The heart of Fermat's proof was the fact that a divisor of a number of the form $x^2 + y^2$ with $\gcd(x, y) = 1$ also can be represented in the form $x^2 + y^2$; thus from $5 \cdot 13 = 8^2 + 1^2$ we may conclude that 5 and 13 are sums of two squares.[4]

Fermat also found by induction that the same claim holds for the quadratic form $x^2 + 2y^2$; on the other hand he knew that it failed for $x^2 + 5y^2$ because

$$21 = 1^2 + 5 \cdot 2^2 = 4^2 + 5 \cdot 1^2, \tag{1.3}$$

yet 3 and 7 cannot be represented in this form.

The connection with Euler's use of quadratic irrationals becomes apparent when we write (1.3) in the form

$$3 \cdot 7 = (1 + 2\sqrt{-5}\,)(1 - 2\sqrt{-5}\,) = (4 + \sqrt{-5}\,)(4 + \sqrt{-5}\,). \tag{1.4}$$

We claim that the factors in these factorizations are all irreducible in the ring $R = \mathbb{Z}[\sqrt{-5}\,]$, and do not differ just by units.

Before we prove this, let us recall the relevant notations. In a domain R (a commutative ring with 1 and without zero divisors), we say that $b \mid a$ (b divides a) if there exists a $c \in R$ with $a = bc$. The divisors of 1 are called units and form a group $R^\times$.

A nonunit $p \in R \setminus R^\times$ is called

- irreducible if it only has trivial factorizations: $p = ab$ for $a, b \in R$;
- prime if $p \mid ab$ for any $a, b \in R$ implies that $p \mid a$ or $p \mid b$.

We know that primes are always irreducible, and that, in unique factorization domains, irreducibles are prime.

Before we can see why 3 is irreducible in $\mathbb{Z}[\sqrt{-5}\,]$ we have to determine the units of this ring. This is quite easy:

Proposition 1.1. *Let $m < -1$ be an integer; then the units of the ring $R = \mathbb{Z}[\sqrt{m}\,] = \{a + b\sqrt{-m} : a, b \in \mathbb{Z}\}$ are $R^\times = \{-1, +1\}$.*

In $\mathbb{Z}$, the prime 5 only has four possible divisors; going through all possibilities easily shows that $(\pm 3, \pm 2)$ are the only integral solutions.

[4] From the modern point of view his proof essentially is a "translation" of the fact that $\mathbb{Z}[i]$ is Euclidean into a language avoiding algebraic numbers.

Before we prove this result, let us put $N(a+b\sqrt{m}) = (a+b\sqrt{m})(a-b\sqrt{m}) = a^2 - mb^2$; this is called the norm of $a+b\sqrt{m}$. It is clear from the definition that the norm is multiplicative, i.e., that $N(\alpha\beta) = N(\alpha)N(\beta)$. In fact, if we call $\alpha' = a - b\sqrt{m}$ the conjugate of $\alpha = a + b\sqrt{m}$, then a simple calculation shows that $N(\alpha\beta) = (\alpha\beta)(\alpha\beta)' = \alpha\alpha'\beta\beta' = N(\alpha)N(\beta)$.

Lemma 1.2. *An element $\varepsilon = a+b\sqrt{m} \in R = \mathbb{Z}[\sqrt{m}\,]$ (here m is a nonsquare integer) is a unit if and only if $N\varepsilon = \pm 1$.*

Proof. If ε is a unit, then there is some $\eta \in R$ with $\varepsilon\eta = 1$. Applying the norm gives $N(\varepsilon)N(\eta) = N(1) = 1$. This is an equation in $\mathbb{Z}$, hence $N(\varepsilon) = N(\eta) = \pm 1$.

Conversely, assume that $\varepsilon = a + b\sqrt{m} \in R$ satisfies $N(\varepsilon) = \pm 1$. Then $\frac{1}{\varepsilon} = \frac{a-b\sqrt{m}}{a^2-mb^2} = \pm(a - b\sqrt{m}\,) =: \eta$ satisfies $\varepsilon\eta = 1$, hence $\varepsilon \in R^\times$. □

Now we are ready to give the

Proof of Prop. 1.1. From Lemma 1.2 we know that $\varepsilon = a + b\sqrt{m} \in R^\times$ is a unit if and only if $N\varepsilon = a^2 - mb^2 = \pm 1$. Since $m < 0$, this is equivalent to $a^2 - mb^2 = 1$, and for $m < -1$ this holds if and only if $b = 0$ and $a = \pm 1$. □

In particular, all the factors in (1.4) are nonunits. Now assume that 3 is reducible in R, i.e., there are nonunits $\alpha, \beta \in R$ with $3 = \alpha\beta$. Taking norms shows that $9 = N(3) = N(\alpha)N(\beta)$. Since α and β are nonunits, and since $N\alpha > 0$, we conclude that $N\alpha = N\beta = 3$. Writing $\alpha = a + b\sqrt{-5}$ shows that $3 = a^2 + 5b^2$; but this is impossible in integers.

The same line of reasoning shows that all the factors in (1.4) are irreducible. This is not enough to conclude that $\mathbb{Z}[\sqrt{-5}\,]$ does not have unique factorization. In fact, consider the factorizations

$$\sqrt{2} \cdot \sqrt{2} = (2 + \sqrt{2}\,)(2 - \sqrt{2}\,)$$

in $R = \mathbb{Z}[\sqrt{2}\,]$. All the factors in there are irreducible since their norm is ± 2; yet the two factorizations do not differ substantially because the factors differ by units: in fact, $2 + \sqrt{2} = \sqrt{2} \cdot (1 + \sqrt{2}\,)$, and $\varepsilon = 1 + \sqrt{2}$ is a unit in R.

On the other hand, 3 and, say, $1 + 2\sqrt{-5}$ do not differ by a unit since their quotient $\frac{1}{3} + \frac{2}{3}\sqrt{-5}$ is not an element of R.

This shows that $\mathbb{Z}[\sqrt{-5}\,]$ does not have unique factorization, and that this is a consequence of Fermat's observation on divisors of numbers of the form $x^2 + 5y^2$. This fact is also responsible for the erroneous result that our second "proof" above has produced. If $\mathbb{Z}[\sqrt{-5}\,]$ had unique factorization, the given proof would actually be correct, as the following lemma shows:

Lemma 1.3. *Assume that R is a unique factorization domain. If $a, b \in R$ are coprime and if $ab = c^n$ for some $c \in R$, then there exists a unit $u \in R^\times$ and elements $r, s \in R$ such that $a = ur^n$ and $b = u^{-1}s^n$.*

Proof. Since R has unique factorization, a is the product of a unit and certain prime powers. Since a and b are coprime, these primes do not divide b; since ab is an n-th power, the exponent of each prime in the factorization of a must be a multiple of n. This proves the claim. $\square$

This result does not hold in $\mathbb{Z}[\sqrt{-5}\,]$: here $(2+\sqrt{-5}\,)(2-\sqrt{-5}\,) = 3^2$ is a square; since the factors $(2+\sqrt{-5}\,)$ and $(2-\sqrt{-5}\,)$ are irreducible (!) and do not differ by a unit, they must be coprime. Yet $\pm(2+\sqrt{-5}\,)$ is not a square (again because $2+\sqrt{-5}$ is irreducible).

The insight that nonunique factorization is responsible for the failure of Euler's method in certain cases became common knowledge in the middle of the 19th century and is connected with the work of Dirichlet, Jacobi, Eisenstein, Liouville, Kummer, and Dedekind.

1.3 Dedekind's Ideals

How can we save unique factorization in rings like $\mathbb{Z}[\sqrt{-5}\,]$? In order to motivate the answer, consider Hilbert's example of the set of integers $M = \{1, 5, 9, \ldots, 4n+1, \ldots\}$. In this monoid, the factorization $9 \cdot 49 = 21 \cdot 21$ shows that unique factrization does not hold. The different factorizations can, however, be explained by introducing the "ideal numbers" 3 and 7 and observing that $9 \cdot 49 = 21 \cdot 21$ comes from pairing up the factors in the ideal factorization $441 = 3^2 7^2$ in two different ways.

Now let us do the same in $\mathbb{Z}[\sqrt{-5}\,]$ by introducing the ideals. Recall that an ideal $\mathfrak{a}$ in a ring R is a set closed with respect to addition and multiplication by ring elements:

$$a, b \in \mathfrak{a} \Longrightarrow a + b \in \mathfrak{a};$$
$$a \in \mathfrak{a}, r \in R \Longrightarrow ra \in \mathfrak{a}.$$

Given elements $a_1, \ldots, a_n \in R$ we can define an ideal $\mathfrak{a} = (a_1, \ldots, a_n) = \{\sum r_i a_i : r_i \in R\}$; ideals of the form $\mathfrak{a} = (a) = aR$ are called principal ideals.

Ideals can be multiplied: we simply let $\mathfrak{a}\mathfrak{b} = \{\sum a_i b_i : a_i \in \mathfrak{a}, b_i \in \mathfrak{b}\}$ be the set of all finite sums of products of elements of $\mathfrak{a}$ and $\mathfrak{b}$. In particular, this implies that e.g. $(a)(b) = (ab)$, $(a)(b_1, b_2) = (ab_1, ab_2)$, $(a_1, a_2)(b_1, b_2) = (a_1b_1, a_1b_2, a_2b_1, a_2b_2)$, etc. Moreover, the ideal $(1) = R$ consisting of all ring elements is a neutral element with respect to this multiplication.

The factorization (1.4) of elements in $R = \mathbb{Z}[\sqrt{-5}\,]$ immediately implies a corresponding factorization of principal ideals

$$(3) \cdot (7) = (1 + 2\sqrt{-5}\,)(1 - 2\sqrt{-5}\,) = (4 + \sqrt{-5}\,)(4 + \sqrt{-5}\,).$$

But whereas the elements in (1.4) were irreducible, the ideals are not. In fact, write $\mathfrak{p} = (3, 1+\sqrt{-5}\,)$, $\mathfrak{p}' = (3, 1-\sqrt{-5}\,)$, $\mathfrak{q} = (7, 4+\sqrt{-5}\,)$, and $\mathfrak{q}' = (7, 1-\sqrt{-5}\,)$. Then we find

$$\begin{aligned}\mathfrak{p}\mathfrak{p}' &= (9, 3(1-\sqrt{-5}\,), 3(1+\sqrt{-5}\,), 6)\\ &= (3)(3, 1-\sqrt{-5}, 1+\sqrt{-5}, 2) = (3)(1) = (3),\end{aligned}$$

because any ideal containing 3 and 2 also contains $3-2=1$ and hence is the unit ideal. Similarly we get $\mathfrak{q}\mathfrak{q}' = (7)$; this calculation is left to the reader. Thus we have $(3)(7) = (\mathfrak{p}\mathfrak{p}')(\mathfrak{q}\mathfrak{q}')$, and we may hope that the other factorizations can be explained similarly. This does work indeed:

$$\begin{aligned}\mathfrak{p}\mathfrak{q} &= (21, 3(4+\sqrt{-5}\,), 7(1+\sqrt{-5}\,), (1+\sqrt{-5}\,)(4+\sqrt{-5}\,))\\ &= (4+\sqrt{-5}\,)(4-\sqrt{-5}, 3+\sqrt{-5}, 1+\sqrt{-5}\,)\\ &= (4+\sqrt{-5}\,)\end{aligned}$$

because the ideal $(4-\sqrt{-5}, 3+\sqrt{-5}, 1+\sqrt{-5}\,)$ contains $7 = 4-\sqrt{-5}+3+\sqrt{-5}$ and $2 = (3+\sqrt{-5}\,)-(1+\sqrt{-5}\,)$, hence $1 = 7-3\cdot 2$. Note that $(3+\sqrt{-5}\,)$ does not denote an ideal here: it must denote a number inside brackets because the left hand side 2 is a number, and because we have not defined the difference of ideals.

Similarly we find $\mathfrak{p}'\mathfrak{q}' = (4+\sqrt{-5}\,)$, $\mathfrak{p}\mathfrak{q}' = (1-2\sqrt{-5}\,)$, and $\mathfrak{p}'\mathfrak{q} = (1+2\sqrt{-5}\,)$. Thus the nonunique factorization of elements in (1.4) turns into the equality

$$(21) = \mathfrak{p}\mathfrak{p}'\mathfrak{q}\mathfrak{q}'$$

of ideals, from which the factorizations of principal ideals

$$(21) = (3)(7) = (1+2\sqrt{-5}\,)(1-2\sqrt{-5}\,) = (4+\sqrt{-5}\,)(4+\sqrt{-5}\,)$$

results from pairing the ideals $\mathfrak{p}$, $\mathfrak{p}'$, $\mathfrak{q}$ and $\mathfrak{q}'$ in different ways.

The first goal now is to prove that this is not accidental, and that factorization into prime ideals holds in any ring of integers of an algebraic number field. This can be shown in various degrees of abstraction. In the next chapter, we give a down and dirty way of doing this in quadratic number fields.

Exercises

1.1 Find units $\neq \pm 1$ in the rings $\mathbb{Z}[\sqrt{m}\,]$ for $m = -1, 2, 3, 5, 6, 7$.

1.2 Prove that $(a_1, a_2)(b_1, b_2) = (a_1b_1, a_1b_2, a_2b_1, a_2b_2)$ for ideals in some commutative ring. Generalize.

1.3 Show that $6 = (1+\sqrt{-5}\,)(1-\sqrt{-5}\,) = 2\cdot 3$ is another example of nonunique factorization in $\mathbb{Z}[\sqrt{-5}\,]$.

1.4 Show that $6 = 2\cdot 3 = (2+\sqrt{-2}\,)(2-\sqrt{-2}\,)$ is not an example of nonunique factorization in $\mathbb{Z}[\sqrt{-2}\,]$.

1.5 Explain the different factorizations in Exercise 1 using the ideals $\mathfrak{p} = (2, 1+\sqrt{-5}\,)$, $\mathfrak{q} = (3, 1+\sqrt{-5}\,)$, and $\mathfrak{q}' = (3, 1-\sqrt{-5}\,)$. Show that

1. $(2, 1-\sqrt{-5}\,) = \mathfrak{p}$;
2. $\mathfrak{p}^2 = (2)$;
3. $\mathfrak{q}\mathfrak{q}' = (3)$;
4. $\mathfrak{q}^2 = (2+\sqrt{-5}\,)$.

1.6 Discuss the factorizations $6 = 2 \cdot 3 = -\sqrt{-6}^2$ in $\mathbb{Z}[\sqrt{-6}\,]$ and $6 = 2 \cdot 3 = (2+\sqrt{10}\,)(-2+\sqrt{10}\,)$ in $\mathbb{Z}[\sqrt{10}\,]$.

1.7 Prove that the only unique factorization domains of the form $\mathbb{Z}[\sqrt{m}\,]$ with $m \leq 1$ are those for $m = 1$ and $m = 2$.
Hints. First consider the case $m \equiv 2 \bmod 4$. If $m > 2$, it is composite, say $m = ab$. Now consider the factorizations $m = ab = -\sqrt{-m}^2$. If m is odd and $m \neq 1$, then have a look at the factorization of $m + 1$.

1.8 The last exercise showed that unique factorization domains are rare among the rings $\mathbb{Z}[\sqrt{m}\,]$ with $m \leq 1$. The situation is better for $m > 1$; nevertheless show that $\mathbb{Z}[\sqrt{m}\,]$ does not have unique factorization if $m = 2n$ with $n \equiv 1 \bmod 4$. Does the proof also work if $n \equiv 3 \bmod 4$?

2. Elementary Arithmetic of Quadratic Number Fields

In this chapter we introduce the rings of integers in quadratic number fields and discuss some of them that happen to be unique factorization domains.

2.1 Algebraic Integers

In the last chapter we have studied some rings of the form $\mathbb{Z}[\sqrt{m}\,]$. It turned out, however, that these are not always the right domains to work with. The reason becomes apparent in the following example.

Consider the ring $R = \mathbb{Z}[\sqrt{-3}\,]$. There we have the factorization $2 \cdot 2 = (1+\sqrt{-3}\,)(1-\sqrt{-3}\,)$ into irreducibles, showing that R does not have unique factorization. The problem is that these factorizations cannot be explained by ideal factorization. In fact, consider the ideal $\mathfrak{a} = (1+\sqrt{-3}\,)$; then $\mathfrak{a}^2 = (-2+2\sqrt{-3}\,) = (2)\mathfrak{a}'$ with $\mathfrak{a}' = (1-\sqrt{-3}\,)$. Multiplying through by $\mathfrak{a}$ shows that $\mathfrak{a}^3 = (8)$. If we had unique factorization into prime ideals, this would imply $\mathfrak{a} = (2)$. But two principal ideals are equal if and only if their generators differ by a unit, hence we would have to conclude that $\frac{1+\sqrt{-3}}{2}$ is a unit; in fact, it is not even an element in R.

Help comes from studying Fermat's Last Theorem for exponent 3: for solving $x^3 + y^3 = z^3$ we could factor the left hand side as

$$x^3 + y^3 = (x+y)(x^2 - xy + y^2).$$

The quadratic factor is irreducible in $\mathbb{Z}$, but can be factored in $\mathbb{C}$ as

$$x^2 - xy + y^2 = (x + y\rho)(x + y\rho^2),$$

where $\rho = \frac{-1+\sqrt{-3}}{2}$ is a primitive cube root of unity, i.e. a complex number $\rho \neq 1$ with the property $\rho^3 = 1$. Thus for studying this diophantine equation it seems we should work with the ring (!) $\mathbb{Z}[\rho] = \{a + b\rho : a, b \in \mathbb{Z}\}$; this ring contains $\mathbb{Z}[\sqrt{-3}\,]$ properly because $2\rho + 1 = \sqrt{-3}$.

Norm and Trace

Before we give the final definition of the "correct" rings of integers, let us introduce some notation. Consider the quadratic number field

$$K = \mathbb{Q}(\sqrt{m}\,) = \{a + b\sqrt{m} : a, b \in \mathbb{Q}\}.$$

This is a Galois extension of $\mathbb{Q}$, i.e., there are two automorphisms, the identity and the conjugation map σ sending $\alpha = a + b\sqrt{m} \in K$ to $\sigma(\alpha) = \alpha' = a - b\sqrt{m}$. Clearly $\sigma^2 = 1$, and $\mathrm{Gal}(K/\mathbb{Q}) = \{1, \sigma\}$. It is obvious that $\alpha \in K$ is fixed by σ if and only if $b = 0$, that is, if and only if $\alpha \in \mathbb{Q}$. We say that K is real or complex quadratic according as $m > 0$ or $m < 0$.

The element $\alpha = a + b\sqrt{m} \in K$ is a root of the quadratic polynomial $P_\alpha(X) = X^2 - 2aX + a^2 - mb^2 \in \mathbb{Q}[X]$; its second root $\alpha' = a - b\sqrt{m}$ is called the *conjugate* of α. We also define

$$\begin{array}{llll} \mathrm{N}\alpha & = \alpha\alpha' & = a^2 - mb^2 & \text{the } \textit{norm} \text{ of } \alpha, \\ \mathrm{Tr}\,\alpha & = \alpha + \alpha' & = 2a & \text{the } \textit{trace} \text{ of } \alpha, \text{ and} \\ \mathrm{disc}(\alpha) & = (\alpha - \alpha')^2 & = 4mb^2 & \text{the } \textit{discriminant} \text{ of } \alpha. \end{array}$$

The basic properties of norm and trace are

Proposition 2.1. *For all $\alpha, \beta \in K$ we have $N(\alpha\beta) = N\alpha\, N\beta$ and $\mathrm{Tr}(\alpha + \beta) = \mathrm{Tr}\,\alpha + \mathrm{Tr}\,\beta$. Moreover $N\alpha = 0$ if and only if $\alpha = 0$, $\mathrm{Tr}\,\alpha = 0$ if and only if $\alpha = b\sqrt{m}$, and $\mathrm{disc}(\alpha) = 0$ if and only if $\alpha \in \mathbb{Q}$.*

Proof. Left as an exercise. □

In particular, the norm is a group homomorphism $K^\times \longrightarrow \mathbb{Q}^\times$, and the trace is a group homomorphism from the additive group $(K, +)$ to the additive group $(\mathbb{Q}, +)$.

The Power of Linear Algebra

Let $K \subseteq L$ be fields; then L may be viewed as a K-vector space: the vectors are the elements from L (they form an additive group), the scalars are the elements of K, and the scalar multiplication is the restriction of the usual multiplication in L. The dimension $\dim_K L$ of L as a K-vector space is called the *degree* of L/K and is denoted by $(L : K)$.

Clearly $K = \mathbb{Q}(\sqrt{m}\,)$ has degree 2 over $\mathbb{Q}$: a basis is given by$\{1, \sqrt{m}\}$ since every element of K can be written uniquely as a $\mathbb{Q}$-linear combination of 1 and $\sqrt{m}$.

In algebraic number theory, fields of higher degree are also studied; for example,

$$\mathbb{Q}(\sqrt[3]{2}\,) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} : a, b, c \in \mathbb{Q}\}$$

is a number field of degree 3 with basis $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$.

Norm and trace can be defined in arbitrary number fields by generalizing the following approach: Let $\{1, \omega\}$ denote a basis of $K = \mathbb{Q}(\sqrt{m}\,)$ as a $\mathbb{Q}$-vector space (for example, take $\omega = \sqrt{m}$). Multiplication by α is a linear map because $\alpha(\lambda\beta + \mu\gamma) = \lambda(\alpha\beta) + \mu(\alpha\gamma)$ for $\lambda, \mu \in \mathbb{Q}$ and $\beta, \gamma \in K$. Now once

a basis is chosen, linear maps can be represented by a matrix; in fact, all we have to do is compute the action of $\alpha = a + b\omega$ on the basis $\{1, \omega\}$.

To this end let us identify $a + b\sqrt{m}$ with the vector $\binom{a}{b}$; then 1 and $\sqrt{m}$ correspond to $\binom{1}{0}$ and $\binom{0}{1}$. The images of these vectors under multiplication by α are, in light of $\alpha \cdot 1 = a + b\omega$ and $\alpha \cdot \omega = bm + a\omega$ for $\omega = \sqrt{m}$, the vectors $\binom{a}{b}$ and $\binom{mb}{a}$. Thus multiplication by α is represented by the matrix $M_\alpha = \left(\begin{smallmatrix} a & mb \\ b & a \end{smallmatrix}\right)$. Now we see that $\mathrm{N}(\alpha) = \det M_\alpha$ and $\mathrm{Tr}(\alpha) = \mathrm{Tr}\, M_\alpha$. It is an easy exercise to show that the norm and the trace in this definition do not depend on the choice of the basis.

From linear algebra we know that the characteristic polynomial of the matrix M_α is given by

$$\det(M_\alpha - XI) = \left| \begin{pmatrix} a - X & mb \\ b & a - X \end{pmatrix} \right| = X^2 - \mathrm{Tr}(\alpha)X + \mathrm{N}(\alpha) = P_\alpha(X).$$

We now say that α is **integral** if the characteristic polynomial $P_\alpha(X)$ has integral coefficients. Clearly α is integral if its norm and trace are ordinary rational integers. Thus all elements in $\mathbb{Z}[\sqrt{m}\,]$ are algebraic integers, but so are e.g. $\rho = \frac{-1+\sqrt{-3}}{2}$ and $\frac{1+\sqrt{5}}{2}$, as is easily checked. Moreover, a rational number $a \in \mathbb{Q}$ is integral if and only if $P_a(X) = X^2 - 2aX + a^2 = (X - a)^2$ has integral coefficients, which happens if and only if $a \in \mathbb{Z}$. This is a good sign: the integral numbers among the rationals according to our definition coincide with the integers!

Rings of Integers

Now let $\mathcal{O}_K$ denote the set of all algebraic integers in $K = \mathbb{Q}(\sqrt{m}\,)$, where m is a squarefree integer. In the following, we will determine $\mathcal{O}_K$ and show that it forms a ring.

Lemma 2.2. *We have $a + b\sqrt{m} \in \mathcal{O}_K$ if and only if $u = 2a$ and $v = 2b$ are integers with $u^2 - mv^2 \equiv 0 \bmod 4$.*

Proof. Assume that $\alpha = a + b\sqrt{m} \in \mathcal{O}_K$; then $u := 2a = \mathrm{Tr}(\alpha) \in \mathbb{Z}$ and $a^2 - mb^2 = \mathrm{N}(\alpha) \in \mathbb{Z}$. Multiplying the last equation through by 4 we find that $4mb^2$ must be an integer. Since m is squarefree, it cannot cancel any denominators in $4b^2$, hence $4b^2$ and therefore also $v := 2b$ are integers. Moreover, $u^2 - mv^2 = 4a^2 - 4mb^2 = 4\mathrm{N}(\alpha)$ is a multiple of 4, hence $u^2 - mv^2 \equiv 0 \bmod 4$.

Now assume that $u = 2a$ and $v = 2b$ are integers with $u^2 - mv^2 \equiv 0 \bmod 4$. Then for $\alpha = a + b\sqrt{m}$ we find that $P_\alpha(X) = X^2 - uX + \frac{1}{4}(u^2 - mv^2)$ has integral coefficients, hence $\alpha \in \mathcal{O}_K$. □

This lemma is now used to classify the algebraic integers in K:

Proposition 2.3. *We have*

$$\mathcal{O}_K = \begin{cases} \{a + b\sqrt{m} : a, b \in \mathbb{Z}\} & \textit{if } m \equiv 2, 3 \bmod 4, \\ \{\frac{a+b\sqrt{m}}{2} : a \equiv b \bmod 2\} & \textit{if } m \equiv 1 \bmod 4. \end{cases}$$

In particular, $\mathcal{O}_K = \mathbb{Z}[\sqrt{m}\,]$ is the ring of integers in K whenever $m \equiv 2, 3 \bmod 4$.

Proof. Assume that $a + b\sqrt{m}$ with $a, b \in \mathbb{Q}$ is an algebraic integer. Then $2a$, $2b$ and $a^2 - mb^2$ are integers by Lemma 2.2.

1. If $m \equiv 2 \bmod 4$, then $u^2 - mv^2 \equiv 0 \bmod 4$ for integers $u = 2a$ and $v = 2b$ implies that u and v are even, hence a and b are integers.

2. If $m \equiv 2 \bmod 4$, then $u^2 - mv^2 \equiv 0 \bmod 4$ for integers $u = 2a$ and $v = 2b$ can only happen if u and v have the same parity; if they are both odd, then $u^2 \equiv v^2 \equiv 1 \bmod 4$ and $u^2 - mv^2 \equiv 2 \bmod 4$: contradiction. Thus u and v are even, and a and b are integers.

3. Finally assume that $m \equiv 1 \bmod 4$. Again, $u^2 - mv^2 \equiv 0 \bmod 4$ if and only if u and v have the same parity. If u and v are both even, then a and b are integers; if not, then $u \equiv v \equiv 1 \bmod 2$ are both odd, and $a + b\sqrt{m} = \frac{u+v\sqrt{m}}{2}$ is an algebraic integer with trace u and norm $\frac{1}{2}(u^2 - mv^2)$. □

In the cases $m \equiv 2, 3 \bmod 4$, every integer in $\mathcal{O}_K$ can be written uniquely as a $\mathbb{Z}$-linear combination of 1 and $\sqrt{m}$: we say that $\{1, \sqrt{m}\}$ is an integral basis in this case. These are not unique: other examples are $\{1, a + \sqrt{m}\}$ for any $a \in \mathbb{Z}$ or $\{1 + \sqrt{m}, \sqrt{m}\}$.

In the case $m \equiv 1 \bmod 4$ we claim that $\mathcal{O}_K$ also has an integral basis, namely $\{1, \omega\}$ with $\omega = \frac{1}{2}(1 + \sqrt{m}\,)$. In fact, for any pair of integers $a, b \in \mathbb{Z}$, the number $a + b\omega = \frac{2a+b+b\sqrt{m}}{2}$ is integral since $2a + b \equiv b \bmod 2$; conversely, any integer $\frac{a+b\sqrt{m}}{2}$ with $a \equiv b \bmod 2$ can be written in the form $\frac{a-b}{2} + b\omega$ with $\frac{a-b}{2}, b \in \mathbb{Z}$. We have proved:

Corollary 2.4. *The ring $\mathcal{O}_K$ of integers in a quadratic number field K is a free abelian group, i.e., for*

$$\omega = \begin{cases} \sqrt{m} & \textit{if } m \equiv 2, 3 \bmod 4, \\ \frac{1+\sqrt{m}}{2} & \textit{if } m \equiv 1 \bmod 4 \end{cases}$$

we have $\mathcal{O}_K = \mathbb{Z} \oplus \omega\mathbb{Z}$.

Now that we have constructed the rings of integers in a quadratic number field, we want to prove that they are Dedekind rings, i.e., domains in which every ideal is the product of prime ideals in a unique way. As a first step we review the basics of ideals and modules in commutative rings – the actual proof of unique factorization into prime ideals will then actually be quite fast and easy.

2.2 Euclidean Rings

Note: I did not cover the following sections in class; the notes are taken from last year's elementary number theory.

A domain R is called a Euclidean ring if there exists a function $\nu : R \longrightarrow \mathbb{N}$ such that

E1 $\nu(r) = 0$ if and only if $r = 0$;
E2 for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = bq + r$ and $\nu(r) < \nu(b)$.

For example, $R = \mathbb{Z}$ is a Euclidean ring with respect to the absoulute value $\nu = |\cdot|$.

Lemma 2.5. *If R is Euclidean with respect to ν, then $\nu(b) = 1$ implies that $b \in R^\times$.*

Proof. Assume that $\nu(b) = 1$; then $1 = bq + r$ with $\nu(r) < \nu(b) = 1$. Thus $\nu(r) = 0$, hence $r = 0$ by [E1], and we have proved that $b \mid 1$, i.e, $b \in R^\times$. □

Let us now give a few important examples of Euclidean rings.

Proposition 2.6. *The ring $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ of Gaussian integers is Euclidean with respect to the norm function $N(x + iy) = x^2 + y^2$.*

Proof. Let us first show that the norm function is multiplicative. This means that $N[(a + bi)(c + di)] = N(a + bi) \cdot N(c + di)$, and is easily checked by computation.

Now assume that we are given elements $a = r + si$ and $b + t + ui$ in $\mathbb{Z}[i]$; then we need to find $q, r \in \mathbb{Z}[i]$ with $a = bq + r$ and $N(r) < N(b)$. Since N is multiplicative, this is equivalent to the statement that for every $p = \frac{a}{b} \in \mathbb{Q}(i) = \{x + yi : x, y \in \mathbb{Q}\}$ there is an element $q \in \mathbb{Z}[i]$ with $N(p - q) < 1$.

Now write $p - q = x + yi$ for $x, y \in \mathbb{Q}$, and let $q = c + di$ with $c, d \in \mathbb{Z}$ and $|x - c| \leq \frac{1}{2}$, $|y - d| \leq \frac{1}{2}$. Then $N(p - q) \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1$. □

This result can be generalized somewhat: the rings $\mathbb{Z}[\sqrt{m}\,]$ are Euclidean with respect to $\nu(a + b\sqrt{m}\,) = |a^2 - mb^2|$ for $m = -2, 2, 3$. In fact, there are more values of m for which these rings are Euclidean, but the proofs soon become very technical.

Proposition 2.7. *Let K be a field. Then the ring $K[X]$ of polynomials in one variable X with coefficients from K is Euclidean with respect to $\nu(f) = 2^{\deg f}$.*

Proof. Assume that $a, b \in R$ are nonzero polynomials. Then we have to find $q, r \in R$ with $a = bq + r$ and $\deg r < \deg b$. We do this by induction and long division.

First observe that the claim is trivial if $\deg a < \deg b$; thus we may assume that $\deg a \geq \deg b$. Then the claim is trivial if $\deg a = 0$, since this implies

$\deg b = 0$, hence b is a nonzero constant, hence a unit, and we can write $a = bq + 0$ with $q = ab^{-1}$.

Now assume that the claim is true for all polynomials a with $\deg a < m$, and write $a = a_m x^m + a_{m-1}x^{m-1} + \ldots + a_0$ and $b = a_n x^n + b_{n-1}x^{n-1} + \ldots + b_0$ with $m \geq n$. Then a and $b \cdot q_1$ with $q_1 = \frac{a_m}{b_n} X^{m-n}$ are polynomials of degree m with the same leading coefficient a_m, hence $r_1 = a - bq_1$ is a polynomial with degree $\deg r_1 < m$. By induction assumption, there exist polynomials q, r with $r_1 = bq + r$ and $\deg r < \deg b$. But now $a = bq_1 + r_1 = b(q_1 + q) + r$, and this proves the theorem. $\square$

Ideals

Our goal is to show that Euclidean rings are UFDs. This will have concrete applications; apart from showing again that e.g. $\mathbb{Z}$ has unique factorization, the fact that $\mathbb{Z}[i]$ is a UFD implies that every prime $p \equiv 1 \bmod 4$ is the sum of two integral squares. Even Lagrange's 4-squares theorem (every positive integer is the sum of 4 integral squares) can be proved by showing that the division algebra of quaternions $A = Q(i, j, k)$ with $i^2 = j^2 = k^2 = -1$ and $ij = -ji$, $ij = k$ contains a (left) Euclidean subring.

The proof that Euclidean rings are UFDs becomes simpler upon introducing another type of rings: principal ideal rings (PIDs). Thus what we actually will prove are the inclusions

$$\text{Euclidean Rings} \subset \text{Principal Ideal Rings} \subset \text{Unique Factorization Rings}.$$

Let R be a ring. A subring I of R is called an ideal if $IR \subseteq I$, i.e., if $ir \in I$ for all $i \in I$ and all $r \in R$.

Let me give you a few examples:

- In any ring, the set $(a) = \{ar : r \in R\}$ is an ideal for any $a \in R$. Such ideals are called principal ideals. In particular, $R = (1)$ and (0) are ideals.
- For $a, b \in R$, the set $(a, b) = \{ar + bs : r, s \in R\}$ is an ideal.
  In $\mathbb{Z}$, the ideal $I = (3, 5)$ contains $2 = 5 - 3$, 3, hence $1 = 3 - 2 = 3 - (5 - 3) = 2 \cdot 3 - 5$. But if $1 \in I$, then $m = 1 \cdot m \in I$ for any $m \in \mathbb{Z}$, and we conclude that $I = (1) = \mathbb{Z}$.
  Similarly, the ideal $I = (6, 9)$ contains $3 = 9 - 6$, hence $3m \in I$ for any $m \in \mathbb{Z}$. On the other hand, if $r \in I$, then $r = 6a + 9b = 3(2a + 3b)$ is a multiple of 3. This shows that $I = (3) = 3\mathbb{Z}$.
- More generally, the set $(a_1, \ldots, a_n) = \{r_1 a_1 + \ldots + r_n a_n : r_i \in R\}$ forms an ideal. Ideals of this form are called finitely generated. For any set of $a_i \in R$, $I = (a_1, a_2, \ldots)$ is defined to be the set of all **finite** R-linear combinations of the a_i; this is again an ideal.
- Every subring of $\mathbb{Z}$ is an ideal; im fact, the subrings of $\mathbb{Z}$ have the form $(m) = m\mathbb{Z}$ for some integer m, and these are all ideals: $a \in (m)$ implies $a = mb$ for some integer b; but then $ar = m(br) \in (m)$ for any integer r.

- The set I of ideals $\left(\begin{smallmatrix} a & 0 \\ c & d \end{smallmatrix}\right)$ with $a, c, d \in \mathbb{Z}$ forms a subring of the ring $M_2(\mathbb{Z})$ of 2×2-matrices with entries in $\mathbb{Z}$, but they do no not form an ideal because e.g. $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is not in I.
- In the polynomial ring $\mathbb{C}[X, Y]$ in two variables, the ideal (X, Y) is not principal (!).

A domain in which every ideal is principal is called a principal ideal domain. Checking whether a given ideal is principal or not is often a nontrivial task.

In order to become familiar with ideals, let us prove

Lemma 2.8. *Let R be a ring. Then $(b) \supseteq (a)$ if and only if $b \mid a$ (to contain is to divide).*

Proof. If $(b) \supseteq (a)$, then $a \in (b)$ and hence $a = bc$ for some $c \in R$. Thus $b \mid a$. The converse is also clear. □

Lemma 2.9. *Let R be a ring. Then $(a) = R$ if and only of $a \in R^\times$.*

Proof. From $(a) = (1)$ we deduce that $1 \in (a)$, hence there is some $r \in R$ with $ar = 1$. But then $a \in R^\times$. □

Lemma 2.10. *Let R be a ring. Then $(a) \subseteq (a, b)$ for any $b \in R$.*

This is trivial.

Lemma 2.11. *Let R be a ring. If $(a) \subseteq I$ and $(b) \subseteq I$, then $(a, b) \subseteq I$.*

Proof. This is clear by the definition of an ideal: from $a, b \in I$ we get $ar + bs \in I$ for all $r, s \in I$. □

The next result connects ideals to the notion of a greatest common divisor:

Proposition 2.12. *Let R be a PID. Then elements have a gcd. Moreover, $d = \gcd(a, b)$ for $a, b, d \in R$ if and only if $(a, b) = (d)$.*

Proof. Let $a, b \in R$. We have to show that there is some $d \in R$ satisfying the axioms of a gcd. Since R is a PID, we can write $(a, b) = (d)$ (such a d will not be unique). There are two things to show:

1. $d \mid a$, $d \mid b$: In fact, $a \in (a, b) = (d)$ implies $a = dr$ for some $r \in R$, hence $d \mid a$; similarly we find $d \mid b$.
2. $e \mid a$, $e \mid b \Longrightarrow e \mid d$: since $d \in (a, b)$ there exist $r, s \in R$ with $d = ar + bs$. Now the assumptions imply that e divides the right hand side, hence $e \mid d$

□

We have seen examples of this before when we showed that $(3, 5) = (1)$ and $(6, 9) = (3)$.

Principal Ideal Domains

Now we claim

Theorem 2.13. *Every Euclidean domain is a PID.*

Proof. Let I be an ideal in the Euclidean ring R. If $I = (0)$ we are done; thus assume that I is not the zero ideal. Let $a \in I$ be a nonzero element with minimal $f(a)$, where f is the Euclidean function. We claim that $I = (a)$.

In fact, let $b \in I$ and write $b = aq + r$ with $f(r) < f(a)$; since $a \in I$ and I is an ideal we know that $aq \in I$, hence $r = b - aq \in I$. By the definition of a we must have $r = 0$, and this shows that every element of I is a multiple of a, i.e., $I = (a)$. □

This provides us with many (but not all) PIDs. In our proof of unique factorization in $\mathbb{Z}$, the main problem was showing that irreducibles are prime. In PIDs, we get this for free:

Proposition 2.14. *In any PID irreducible elements are prime.*

Proof. Let $p \in R$ be irreducible, and assume that $p \mid ab$. If $p \mid a$ we are done, so assume that $p \nmid a$. We claim that $(a, p) = (1) = R$. In fact, write $(d) = (a, p)$. Then $d \mid p$, hence $p = dr$ for $d, r \in R$. Since p is irreducible, d or r must be a unit. If d is a unit, then $(a, p) = (1)$ as claimed, and if r is a unit, then $(d) = (p)$, hence $(a, p) = (p)$ and finally $p \mid a$: contradiction.

Thus $(a, p) = (1)$, hence there exist $r, s \in R$ with $ar + ps = 1$. But then $b = abr + aps$, and since $p \mid ab$, p divides the right hand side and therefore b. □

Next we have to show that every nonzero nonunit in a PID has a factorization into irreducibles. This is not at all obvious: consider e.g. the domain $D = \mathbb{Z}[\sqrt{2}, \sqrt[4]{2}, \sqrt[8]{2}, \ldots]$ containing $\mathbb{Z}$ and all roots $2^{1/2^n}$ for $n \geq 1$. Then 2 is not a unit, and it is not irreducible because $2 = \sqrt{2} \cdot \sqrt{2}$. But $\sqrt{2} = \sqrt[4]{2} \cdot \sqrt[4]{2}$ shows that $\sqrt{2}$ is also reducible, and this process can be continued indefinitely: although 2 is a nonunit, it is not a product of irreducibles because none of its factors is irreducible. In PIDs, this does not happen:

Proposition 2.15. *Let R be a PID. Then every $a \in R \setminus \{0\}$ has a factorization into a unit times irreducible elements.*

Proof. If a is a unit, we are done. If a is a nonunit then we claim that a has an irreducible factor. This is clear if a is irreducible; if not then it has a nontrivial factorization $a = a_1 b_1$. If a_1 is irreducible, we are done; if not, then there is a nontrivial factorization $a_1 = a_2 b_2$ etc. In this way we get a sequence of elements $a_1, a_2, \ldots$ with $\ldots, a_3 \mid a_2$, $a_2 \mid a_1$, $a_1 \mid a$. Consider the ideal $I = (a, a_1, a_2, \ldots)$. Since R is a PID, there is a $c \in R$ with $I = (c)$. Since I is the union of the ideals (a), (a_1), (a_2), $\ldots$, c must be an element of one of these, say $c \in (a_m)$. But then $(c) \subseteq (a_m)$ and $(a_m) \subseteq I = (c)$ imply that

$I = (a_m)$. Now $a_{m+1} \mid a_m$, as well as $a_m \mid a_{m+1}$ because $a_{m+1} \in I = (a_m)$: this implies that a_m and a_{m+1} differ by a unit, hence $a_m = a_{m+1}b_{m+1}$ is not a nontrivial factorization.

Thus we have shown that every nonzero nonunit a is divisible by an irreducible element. We now claim that a has a factorization into irreducibles. In fact, write $a = a_1b_1$ with a_1 irreducible. If b_1 is irreducible, we are done; if not, write $b_1 = a_2b_2$ with a_2 irreducible and continue. By the same argument as above this process must terminate, and after finitely many steps we have a factorization of a into irreducibles. □

Now we are ready to prove

Theorem 2.16. *Every PID is a UFD.*

Proof. We have already shown the following two facts:

1. Every element $\neq 0$ has a factorization into irreducible elements;
2. Irreducibles are primes.

Now assume that $a = p_1 \cdots p_r = q_1 \cdots q_s$ are factorizations into irreducibles. Since p_1 is prime and divides the right hand side, it must divide one of the factors, say $p_1 \mid q_1$. Since q_1 is irreducible, we must have $q_1 = p_1u_1$ for some unit u_1; replacing q_2 by q_2u_1 and cancelling p_1 shows that $p_2 \cdots p_r = q_2 \cdots q_s$. Now do induction on the number of irreducible factors just as in $\mathbb{Z}$. □

Residue Classes modulo Ideals

Ideals have played a role in our proof that Euclidean domains have unique factorization. The main purpose of ideals, however, is that they can be used to generalize the notion of residue classes modulo elements.

In fact, let R be a ring (as always commutative and with a multiplicative unit 1) and I an ideal in R. Then we say that $a \equiv b \bmod I$ if $a - b \in I$. If $I = (m)$ is principal, this is the usual definition: we have $a \equiv b \bmod (m)$ $\iff$ $a - b \in (m)$ $\iff$ $a - b = mr$ for some $r \in R$ $\iff$ $m \mid a - b$ $\iff$ $a \equiv b \bmod m$.

The residue class $a \bmod I$ is denoted by $[r]$ or $r + I$. Thus $r + I = \{a \in R : a \equiv r \bmod I\} = \{r + i : i \in I\}$. The set of residue classes modulo I is denoted by R/I. Note that $\mathbb{Z}/m\mathbb{Z}$ is equal to R/I for $R = \mathbb{Z}$ and $I = m\mathbb{Z} = (m)$.

Proposition 2.17. *The set R/I of residue classes modulo I forms a ring.*

We define the norm of an ideal I by $N(I) = \#R/I$. Note that the norm of an ideal might be infinite; for example, $\mathbb{Z}/(0)$ has infinitely many elements (distinct integers determine distinct residue classes modulo (0); similarly, $\mathbb{Z}[X]/(X)$ is infinite). Since R/I is a ring, we can form its unit group $(R/I)^\times$. We now define Euler's phi function for ideals in R by $\Phi(I) = \#(R/I)^\times$.

In the case $R = \mathbb{Z}$ we have proved that $\phi(p^n) = (p-1)p^{n-1}$ for positive primes (or $\phi(p) = (|p|-1)|p|^{n-1}$ for arbitrary primes). We did this by counting all the elements in $\mathbb{Z}/p^n\mathbb{Z}$ (there were p^n of them) and subtracting the number classes represented by multiples of p (there are p^{n-1} of them).

2.3 Arithmetic of $\mathbb{Z}[i]$

Units and Primes

Finding all units in $R = \mathbb{Z}[i]$ is easy. Assume that $a + bi$ is a unit; then $(a+bi)(c+di) = 1$, and taking the norm shows that $(a^2+b^2)(c^2+d^2) = 1$, which implies that $a^2+b^2 = 1$. This happens if and only if $a+bi \in \{\pm 1, \pm i\}$.

Proposition 2.18. *We have* $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$.

Now let us determine all the primes in $\mathbb{Z}[i]$. Assume that $a + bi$ is prime. Then $(a+bi) \mid (a+bi)(a-bi) = N(a+bi) = a^2+b^2$. Thus every prime divides a natural number a^2+b^2; writing this number as a product of primes in $\mathbb{N}$ and keeping in mind that $a+bi$ is a prime in $\mathbb{Z}[i]$ we find that $a+bi$ must divide one of the prime factors of a^2+b^2.

Lemma 2.19. *Every prime in* $\mathbb{Z}[i]$ *divides a prime in* $\mathbb{Z}$.

Thus in order to find all primes in $\mathbb{Z}[i]$ we only need to look at factors of primes in $\mathbb{Z}$. Of course primes in $\mathbb{Z}$ need not be prime in $\mathbb{Z}[i]$: for example, we have $5 = (1+2i)(1-2i)$.

Now assume that a prime $p \in \mathbb{N}$ factors nontrivially in $\mathbb{Z}[i]$; then $p = (a+bi)(c+di)$. Taking norms gives $p^2 = (a^2+b^2)(c^2+d^2)$. Since none of the factors is a unit, we must have $a^2+b^2 = c^2+d^2 = p$. Since $a^2+b^2 \equiv 0, 1, 2 \bmod 4$, primes of the form $p \equiv 3 \bmod 4$ are irreducible in $\mathbb{Z}[i]$, and since $\mathbb{Z}[i]$ is a UFD, they are prime (in algebraic number theory, primes in $\mathbb{Z}$ remaining prime in an extension are called inert).

Next $2 = i^3(1+i)^2$: thus 2 is a unit times a square (in algebraic number theory, such primes will be called ramified).

Finally, if $p \equiv 1 \bmod 4$, then $(\frac{-1}{p}) = +1$, hence $x^2 \equiv -1 \bmod p$ for some integer x. This implies $p \mid (x^2+1) = (x+i)(x-i)$. Now clearly p does not divide any of the factors since $\frac{x}{p} + \frac{1}{p}i$ is not a Gaussian integer. Thus p divides a product without dividing one of the factors, and this means p is not prime in $\mathbb{Z}$. Since irreducibles are prime, this implies that p must be reducible, i.e., it has a nontrivial factorization. We have seen above that this means that $p = a^2+b^2$ for $a, b \in \mathbb{Z}$: thus the first supplementary law plus unique factorization in $\mathbb{Z}[i]$ implies Fermat's two-squares theorem!

Note that this proof is a lot more involved than the simple proof we have given before; on the other hand, it is much clearer.

Let us get back to primes $p \equiv 1 \bmod 4$. We have seen that $p = a^2+b^2 = (a+bi)(a-bi)$. Can it happen that $a+bi$ and $a-bi$ differ only by a unit? If

$a+bi=(a-bi)\varepsilon$, then $\varepsilon=\frac{a+bi}{a-bi}=\frac{1}{p}(a+bi)^2=\frac{1}{p}(a^2-b^2+2abi)$. But this is not a Gaussian integers since $p\nmid 2ab$. Thus $a+bi$ and $a-bi$ are distinct primes (in algebraic number theory, we say that such primes split).

Theorem 2.20. *The ring $\mathbb{Z}[i]$ has the following primes:*

- $1+i$*, the prime dividing* 2*;*
- $a+bi$ *and* $a-bi$*, where* $p=a^2+b^2\equiv 1 \bmod 4$*;*
- *rational primes* $q\equiv 3 \bmod 4$*.*

In particular, $\mathbb{Z}[i]$ has infinitely many primes. We could have proved this also by Euclid's argument.

Residue Class Systems

Of course we can now define residue classes in $\mathbb{Z}[i]$: we say that $a+bi\equiv c+di \bmod r+si$ if $(r+si)\mid(a-c+(b-d)i)$. If $a+bi$ is a prime, how many residue classes are there?

This is easy for the prime $1+i$: we claim that every Gaussian integer is congruent to 0 or 1 mod $1+i$. In fact, $a+bi\equiv a-b \bmod (1+i)$ because $i\equiv -1 \bmod 1+i$. Now we can reduce $a-b \bmod 2$ since 2 is a multiple of $1+i$, and this proves the claim. Moreover, $1\not\equiv 0 \bmod 1+i$ since 1 is not divisible by $1+i$. Thus $\{0,1\}$ is a complete system of residues modulo $1+i$.

The same trick works for primes $c+di$ with norm $p\equiv 1 \bmod 4$: we have $i\equiv -\frac{c}{d} \bmod c+di$, hence $a+bi\equiv a-b\frac{c}{d} \bmod a+bi$. Thus every Gaussian integer is congruent to some integer modulo $a+bi$. Now we can reduce modulo p (this is a multiple of $a+bi$) and find that every Gaussian integer is congruent to some element $0,\ 1,\ \ldots,\ p-1$ modulo $a+bi$. Moreover, these elements are incongruent modulo $a+bi$: if $r\equiv s \bmod a+bi$ for $0\le r,s<p$, then $(a+bi)\mid(r-s)$; taking norms gives $p^2\mid(r-s)^2$, hence $p\mid(r-s)$ and finally $r=s$. Thus $\{0,1,\ldots,p-1\}$ is a complete system of residues modulo the prime $a+bi$ with norm p.

Finally, consider inert primes $q\equiv 3 \bmod 4$. Here we claim that $S=\{r+si : 0\le r,s<p\}$ is a complete system of residues modulo p (note that this set contains p^2 elements). It is clear that every $a+bi\equiv r+si \bmod p$ for some $r+si\in S$: just reduce a and b modulo p. We only have to show that no two elements in S are congruent modulo p. Assume therefore that $r+si\equiv t+ui \bmod p$ for $0\le r,s,t,u<p$. Then $p\mid(r-t+(s-u)i)$, i.e., $\frac{r-t}{p}+\frac{s-u}{p}i\in\mathbb{Z}[i]$. This happens if and only if $r\equiv t \bmod p$ and $s\equiv u \bmod p$, which implies $r=t$ and $s=u$.

We have proved:

Proposition 2.21. *The complete system of residues modulo a Gaussian prime $a+bi$ has exactly $N(a+bi)=a^2+b^2$ elements.*

Let us now add a level of abstraction and consider, for a prime $p=a^2+b^2\equiv 1 \bmod 4$, the map $\lambda:\mathbb{Z}/p\mathbb{Z}\longrightarrow\mathbb{Z}[i]/(a+bi):[r]_p\longmapsto[r]_{a+bi}$.

It obviously is a homomorphism because $\lambda([r]_p)\lambda([s]_p) = [r]_{a+bi}[s]_{a+bi} = [rs]_{a+bi} = \lambda([rs]_p)$. From what we have seen above, λ is surjective because every residue class modulo $a+bi$ is represented by one of the integers 0, 1, $\ldots p-1$. Is λ injective? Its kernel is $\ker\lambda = \{[r]_p : [r]_{a+bi} = [0]_{a+bi}\}$. Now $r \equiv 0 \bmod a+bi$ implies $p^2 \mid r^2$, hence $p \mid r$, hence $[r]_p = [0]_p$. Thus $\ker\lambda = \{[0]_p\}$, and λ is injective.

We have seen that $\lambda : \mathbb{Z}/p\mathbb{Z} \longrightarrow \mathbb{Z}[i]/(a+bi)$ is an isomorphism: the two residue class systems have the same number of elements, the same structure, and in particular, they are both fields with p elements.

What can we say about the residue class ring $R = \mathbb{Z}[i]/(p)$ for primes $p \equiv 3 \bmod 4$? Let us check that R is a domain, i.e., that it has no zero divisors. In fact, assume that $(a+bi)(c+di) \equiv 0 \bmod p$. Since p is a prime in $\mathbb{Z}[i]$, this implies $p \mid (a+bi)$ or $p \mid (c+di)$, hence $[a+bi]_p = [0]_p$ or $[c+di]_p = [0]_p$, and this shows that R is a domain.

Now we have

Proposition 2.22. *Any domain R with finitely many elements is a field.*

Proof. Let $a \in R$ be nonzero. We need to show that a is a unit, i.e. that there is a $b \in R$ with $ab = 1$. Let $n = \#R$; we claim that $a^{n-1} = 1$ (this is like Fermat's little theorem, and the proof is the same). Write $R \setminus \{0\} = \{a_1 = 1, a_2, \ldots, a_{n-1}\}$. Define elements b_j by $a_j a = b_j$. We claim that the b_j are just the a_j in some order. This will follow if we can show that no two b_j are equal. Assume therefore that $b_j = b_k$; then $a_j a = a_k a$. Since R is a domain, we may cancel (note that $ac = ad$ implies $a(c-d) = 0$, and since R has no zero divisors, this shows that $a = 0$ or $c = d$), and we get $a_j = a_k$.

Next we multiply all the equations $a_1 a = b_1$, $\ldots$, $a_{n-1}a = b_{n-1}$; since $\prod a_j = \prod b_j$ we conclude that $a^{n-1} = 1$.

Now we simply put $b = a^{n-2}$ and observe that $ab = 1$. □

We have proved:

Proposition 2.23. *Let $p \equiv 3 \bmod 4$ be prime. Then $\mathbb{Z}[i]/(p)$ is a field with p^2 elements.*

There also exist finite fields with p^2 elements for primes $p \equiv 1 \bmod 4$, but these cannot be constructed as residue class fields in $\mathbb{Z}[i]$.

2.4 Reciprocity Laws

Now pick a prime $\pi = a + bi \equiv 1 \bmod 2$; note that this means that a is odd and b is even. We have

Proposition 2.24 (Fermat's Little Theorem)**.** *For any element α coprime to π we have $\alpha^{N\pi - 1} \equiv 1 \bmod \pi$.*

The proof is analogous to the one in $\mathbb{Z}$: enumerate the $N\pi - 1$ elements α_i in $(Z[i]/\pi\mathbb{Z}[i])^\times$, then multiply them by α and show that the products $\beta_i \equiv \alpha\alpha_i \bmod \pi$ are just the α_i in some order. Then multiply etc.

As an example, let us compute $(1+i)^{N\pi-1} \bmod \pi$ for $\pi = 1+2i$. Then $N\pi = 5$ and $(1+i)^4 = (2i)^2 = -4 \equiv 1 \bmod 5$, hence modulo π.

Quadratic Reciprocity in $\mathbb{Z}[i]$

Since $\pi \equiv 1 \bmod 2$ we find that $N\pi = a^2 + b^2 \equiv 1 \bmod 4$. Thus

$$0 \equiv \alpha^{N\pi-1} - 1 = \big(\alpha^{\frac{N\pi-1}{2}} - 1\big)\big(\alpha^{\frac{N\pi-1}{2}} + 1\big) \bmod \pi,$$

hence $\alpha^{\frac{N\pi-1}{2}} \equiv \pm 1 \bmod \pi$ since π is prime. We now define the quadratic Legendre symbol $[\frac{\alpha}{\pi}] = \pm 1$ in $\mathbb{Z}[i]$ by

$$\left[\frac{\alpha}{\pi}\right] \equiv \alpha^{\frac{N\pi-1}{2}} \bmod \pi.$$

As an example, let us compute $[\frac{1+i}{1+2i}]$. We find $(1+i)^{(N\pi-1)/2} = (1+i)^2 = 2i \equiv -1 \bmod \pi$, hence $[\frac{1+i}{1+2i}] = -1$.

In order to prove a quadratic reciprocity law in $\mathbb{Z}[i]$ we collect a few simple properties of these symbols.

Proposition 2.25. *For elements $\alpha, \beta, \pi \in \mathbb{Z}[i]$ with $\pi \equiv 1 \bmod 2$ prime we have*

1. $[\frac{\alpha}{\pi}] = [\frac{\beta}{\pi}]$ *if* $\alpha \equiv \beta \bmod \pi$*;*
2. $[\frac{\alpha\beta}{\pi}] = [\frac{\alpha}{\pi}][\frac{\beta}{\pi}]$*;*
3. $[\frac{\alpha\beta}{\pi}] = +1$ *if* $\alpha \equiv \xi^2 \bmod \pi$.

Proof. The first and second follow directly from the definition. Assume now that $\alpha \equiv \xi^2 \bmod \pi$; then $\alpha^{\frac{N\pi-1}{2}} \equiv \xi^{N\pi-1} \equiv 1 \bmod \pi$ by Fermat's Little Theorem. □

We also will use a few results on the quadratic character of certain integers:

Proposition 2.26. *Let $p = a^2 + b^2$ be an odd prime, and suppose that a is odd. Then*

$$\left(\frac{a}{p}\right) = 1, \left(\frac{b}{p}\right) = \left(\frac{2}{p}\right), \text{ and } \left(\frac{a+b}{p}\right) = \left(\frac{2}{a+b}\right).$$

Proof. Using the quadratic reciprocity law, we get $(\frac{a}{p}) = (\frac{p}{a}) = +1$, because $p = a^2 + b^2 \equiv b^2 \bmod a$. Next the congruence $(a+b)^2 \equiv 2ab \bmod p$ shows that $(\frac{a}{p}) = (\frac{2b}{p})$, and this proves our second claim. Finally $2p = (a+b)^2 + (a-b)^2$ implies that $(\frac{a+b}{p}) = (\frac{p}{a+b}) = (\frac{2}{a+b})$. □

In order to prove the quadratic reciprocity law in $\mathbb{Z}[i]$, we write $\pi = a + bi, \lambda = c + di$; then $\pi \equiv \lambda \equiv 1 \bmod 2$ implies that $a \equiv c \equiv 1 \bmod 2$ and $b \equiv d \equiv 0 \bmod 2$. If $\pi = p \in \mathbb{Z}$ or $\lambda = \ell \in \mathbb{Z}$, the proof follows directly from the relations $\left[\frac{p}{\lambda}\right] = \left(\frac{p}{N\lambda}\right)$ and $\left[\frac{\pi}{\ell}\right] = \left(\frac{N\pi}{\ell}\right)$, which are easy to verify:

Proposition 2.27. *For primes $\pi \in \mathbb{Z}[i]$ and elements $a \in \mathbb{Z}$ coprime to π we have $\left[\frac{a}{\pi}\right] = \left(\frac{a}{N\pi}\right)$.*

Proof. We have $\left[\frac{a}{\pi}\right] \equiv a^{(N\pi-1)/2} \bmod \pi$ and $\left(\frac{a}{N\pi}\right) \equiv a^{(N\pi-1)/2} \bmod p$. This implies $\left[\frac{a}{\pi}\right] \equiv \left(\frac{a}{N\pi}\right) \bmod \pi$. If the symbols were different, then π must divide 2, hence $N\pi$ must divide 4, and this is nonsense since π has odd norm > 1. $\square$

Thus we may assume that $p = N\pi$ and $\ell = N\lambda$ are prime. We find immediately that $ai \equiv b \bmod \pi$ and $ci \equiv d \bmod \lambda$, hence we get

$$\left[\frac{\pi}{\lambda}\right] = \left[\frac{c}{\lambda}\right]\left[\frac{ac+bci}{\lambda}\right] = \left[\frac{c}{\lambda}\right]\left[\frac{ac+bd}{\lambda}\right].$$

Since $c \in \mathbb{Z}$ and $ac + bd \in \mathbb{Z}$, we find

$$\left[\frac{c}{\lambda}\right] = \left(\frac{c}{\ell}\right) \text{ and } \left[\frac{ac+bd}{\lambda}\right] = \left(\frac{ac+bd}{\ell}\right).$$

Using Proposition 2.26, we find

$$\left[\frac{\pi}{\lambda}\right] = \left(\frac{ac+bd}{\ell}\right). \tag{2.1}$$

But now $p\ell = (a^2+b^2)(c^2+d^2) = (ac+bd)^2+(ad-bc)^2 \equiv (ad-bc)^2 \bmod (ac+bd)$ implies $\left(\frac{\ell}{ac+bd}\right) = \left(\frac{p}{ac+bd}\right)$, and applying the quadratic reciprocity law in $\mathbb{Z}$ twice shows that

$$\left(\frac{ac+bd}{\ell}\right) = \left(\frac{\ell}{ac+bd}\right) = \left(\frac{p}{ac+bd}\right) = \left(\frac{ac+bd}{p}\right).$$

The quadratic reciprocity law in $\mathbb{Z}[i]$ follows by symmetry:

$$\left[\frac{\pi}{\lambda}\right] = \left(\frac{ac+bd}{\ell}\right) = \left(\frac{ac+bd}{p}\right) = \left[\frac{\lambda}{\pi}\right].$$

The supplementary laws follow immediately from (2.1) by putting $a = 0, b = 1$ or $a = b = 1$, and using quadratic reciprocity.

Quartic Reciprocity

Note that $N\pi \equiv 1 \bmod 4$ implies that we have

$$\begin{aligned} 0 &\equiv \alpha^{N\pi-1} - 1 = \left(\alpha^{\frac{N\pi-1}{2}} - 1\right)\left(\alpha^{\frac{N\pi-1}{2}} + 1\right) \\ &\equiv \left(\alpha^{\frac{N\pi-1}{4}} - 1\right)\left(\alpha^{\frac{N\pi-1}{4}} + 1\right)\left(\alpha^{\frac{N\pi-1}{4}} - i\right)\left(\alpha^{\frac{N\pi-1}{4}} + i\right) \bmod \pi, \end{aligned}$$

hence we can define a quartic power residue symbol $[\frac{\alpha}{\pi}]_4 \in \{1, i, -1, -i\}$ by demanding that

$$\left[\frac{\alpha}{\pi}\right]_4 \equiv \alpha^{\frac{N\pi-1}{4}} \bmod \pi.$$

This symbol satisfies the quartic reciprocity law, according to which

$$\left[\frac{\pi}{\lambda}\right]_4 = \left[\frac{\lambda}{\pi}\right]_4 (-1)^{\frac{N\pi-1}{4}\frac{N\lambda-1}{4}}$$

for any two primes $\pi \equiv \lambda \equiv 1 \bmod (2+2i)$. Its proof is much more difficult that that of the quadratic reciprocity law, but is quite easy using Gauss sums, which are objects defined using the theory of finite fields.

2.5 Gauss Sums

Eventually, I will put a section on quartic Gauss sums here and show how to use them for deriving the quartic reciprocity law.

2.6 Other quadratic number rings

The arithmetic of this ring is very similar to that of $\mathbb{Z}[i]$. In particular, it is a UFD. We will now use this fact to prove one of Fermat's claims; the proof itself is due to Euler, who worked with the algebraic integers $a + b\sqrt{-2}$ as if they were natural numbers, not worrying about defining what primes, gcd's etc. are or whether unique factorization holds.

Theorem 2.28. *The diophantine equation $y^2 = x^3 - 2$ has $(3, \pm 5)$ as its only solutions in integers.*

Proof. Write $x^3 = y^2 + 2 = (y + \sqrt{-2})(y + \sqrt{-2})$. Note that y must be odd (otherwise $y^2 + 2 \equiv 2 \bmod 4$, and no cube is $\equiv 2 \bmod 4$). Now let $\delta = \gcd(y + \sqrt{-2}, y + \sqrt{-2})$. Clearly $\delta \mid 2\sqrt{-2}$ (the difference of these values); thus δ is a power of $\sqrt{-2}$. On the other hand, if $\sqrt{-2} \mid (y \pm \sqrt{-2})$, then it divides the product of these factors, which is $x^3 = y^2 + 2$. But x is odd, hence $\sqrt{-2} \nmid \delta$.

We have seen that $y + \sqrt{-2}$ and $y + \sqrt{-2}$ are coprime and that their product is a cube. Since $\mathbb{Z}[\sqrt{-2}]$ is a UFD, this implies that the factors are cubes up to units. Since the only units are ± 1 and since these are cubes, it follows that $y + \sqrt{-2} = (a + b\sqrt{-2})^3$. Comparing real and imaginary parts we find $y = a^3 - 6ab^2$ and $1 = 3a^2b - 2b^3$. The last equation shows $1 = b(3a^2 - 2b^2)$, hence $b = \pm 1$ and therefore $a = \pm 1$. This shows $y = \pm 5$ and finally $x = 3$. $\square$

Exercises

2.1 Show that $\mathbb{Z}[\sqrt{m}\,]$ is norm-Euclidean for $m = -2, 2, 3$.

2.2 Let R be the ring of continuous functions $\mathbb{R} \longrightarrow \mathbb{R}$, where addition and multiplication are defined pointwise.
1. Determine the unit group $R^\times$;
2. does R contain irreducible elements?
3. for $a \in \mathbb{R}$ let $I_a = \{f \in R : f(a) = 0\}$; is I_a an ideal?
4. find an ideal in R that is not principal.

2.3 Prove that the ideal (X, Y) in $\mathbb{C}[X, Y]$ is not principal.

2.4 Use the Euclidean algorithm to compute $\gcd(7 - 6i, 3 - 14i)$.

2.5 Find the prime factorization of $-3 + 24i$. (Hint: first factor the norm).

2.6 Find $c \in \{0, 1, \ldots, 16\}$ such that $3 + 2i \equiv c \bmod 1 + 4i$.

2.7 Show that for any $\alpha \in \mathbb{Z}[i]$ with odd norm there is a unit $\varepsilon \in \mathbb{Z}[i]^\times$ such that $\alpha\varepsilon = a + bi$ with a odd, b even, and $a + b \equiv 1 \bmod 4$. Show also that this condition is equivalent to $a + bi \equiv 1 \bmod (2 + 2i)$.

2.8 Use Euclid's argument to show that there are infinitely many primes in $\mathbb{Z}[i]$.

2.9 Show that for primes $p = a^2 + b^2$ with b even we have $[\frac{a+bi}{a-bi}] = (\frac{2}{p})$.

2.10 Compute the quartic symbols $[\frac{1+2i}{1+4i}]_4$ and $[\frac{1+4i}{1+2i}]_4$, and check that the quartic reciprocity law holds for these elements.

2.11 Let $R = \mathbb{Z}[\sqrt{m}\,]$ with $m < -1$. Show that $R^\times = \{\pm 1\}$.

2.12 Show that $\mathbb{Z}[\sqrt{2}\,]$ contains infinitely many units.

2.13 Find all the prime elements in $\mathbb{Z}[\sqrt{-2}\,]$.

2.14 Use the ring $\mathbb{Z}[\sqrt{-2}\,]$ to construct finite fields with p^2 elements for primes $p \equiv 5, 7 \bmod 8$.

2.15 Solve the congruence $x^2 \equiv -1 \bmod 41$ and then compute $\gcd(x+i, 41)$ in $\mathbb{Z}[i]$.

2.16 Find infinitely many integers $x, y, z \in \mathbb{Z}$ with $x^2 + y^2 = z^3$.

2.17 Solving equations like $y^2 = x^3 + c$ is not always as easy as for $c = -2$. Show that solving $y^2 = x^3 + 1$ in the standard way leads to the new diophantine equation $a^3 - 2b^3 = 1$. How do you think mathematicians solve this last equation?

3. Modules and Ideals

In this chapter we will discuss modules and ideals in general, as well as in the rings of integers of quadratic number fields.

3.1 Modules

Let R be a commutative ring; an (additively written) abelian group M is said to be an R-module if there is a map $R \times M \longrightarrow M : (r, m) \longmapsto rm$ with the following properties:

- $1m = m$ for all $m \in M$;
- $r(sm) = (rs)m$ for all $r, s \in R$ and $m \in M$;
- $r(m + n) = rm + rn$ for all $r \in R$ and $m, n \in M$;
- $(r + s)m = rm + sm$ for all $r, s \in R$ and $m \in M$.

The most important examples are abelian groups G: they are all $\mathbb{Z}$-modules via $ng = g + \ldots + g$ (n terms) for $n > 0$ and $ng = -(-n)g$ for $n < 0$. In particular, a subring M of a commutative ring R is a $\mathbb{Z}$-module.

If M and N are R-modules, then so is $M \oplus N = \{(m, n) : m \in M, n \in N\}$ via the action $r(m, n) = (rm, rn)$.

In the following, $K = \mathbb{Q}(\sqrt{d}\,)$ is a quadratic number field, and $\{1, \omega\}$ is a basis of its ring of integers $\mathcal{O}_K$. Our first job is the classification of all $\mathbb{Z}$-modules in $\mathcal{O}_K$.

Proposition 3.1. *Let $M \subset \mathcal{O}_K$ be a $\mathbb{Z}$-module in $\mathcal{O}_K$. Then there exist natural numbers m, n and and integer $c \in \mathbb{Z}$ such that $M = [n, c + m\omega] := n\mathbb{Z} \oplus (c + m\omega)\mathbb{Z}$.*

Note that this says that every element in M is a unique $\mathbb{Z}$-linear combination of n and $c + m\omega$; the elements n and $c + m\omega$ are therefore called a basis of the $\mathbb{Z}$-module M in analogy to linear algebra. Actually, studying R-modules is a generalization of linear algebra in the sense that R-modules are essentially vector spaces with the field of scalars replaced by a ring.

Also observe that, in general, not every R-module has a basis; R-modules possessing a basis are called **free**, and the number of elements in a basis is called the **rank** of the R-module. Proposition 3.1 claims that all $\mathbb{Z}$-modules

in $\mathcal{O}_K$ are free of rank ≤ 2. In fact, the $\mathbb{Z}$-modules $M = \{0\} = [0,0]$, $M = \mathbb{Z} = [1,0]$ and $M = \mathcal{O}_K = [1,\omega]$ have ranks 0, 1 and 2, respectively.

Proof of Prop. 3.1. Step 1: defining m, n and c. Consider the subgroup $H = \{s : r + s\omega \in M\}$ of $\mathbb{Z}$. Every subgroup of $\mathbb{Z}$ has the form $m\mathbb{Z}$ for some integer m, hence in particular we have $H = m\mathbb{Z}$ for some $m \geq 0$. By construction, there is an integer $c \in \mathbb{Z}$ such that $c + m\omega \in M$. Finally, $M \cap \mathbb{Z}$ is a subgroup of $\mathbb{Z}$, hence $M \cap \mathbb{Z} = n\mathbb{Z}$ for some $n \geq 0$.

Step 2: showing that this definition works. We now claim that $M = n\mathbb{Z} \oplus (c + m\omega)\mathbb{Z}$. The inclusion $\supseteq$ is clear; assume therefore that $r + s\omega \in M$. Since $s \in H$ we have $s = um$ for some $u \in \mathbb{Z}$, and then $r - uc = r + s\omega - u(c + m\omega) \in M \cap \mathbb{Z}$, hence $r - uc = vn$. But then $r + s\omega = r - uc + u(c + m\omega) = vn + u(c + m\omega) \in n\mathbb{Z} \oplus (c + m\omega)\mathbb{Z}$. $\square$

Residue Classes

Given a $\mathbb{Z}$-submodule M of R we can form the quotient group R/M whose elements are expressions of the form $r + M$ for $r \in R$, with $r + M = s + M$ if and only if $r - s \in M$; addition is defined by $(r + M) + (s + M) = (r + s) + M$.

The number of elements in R/M is called the norm of the module M and will be denoted by $\mathrm{N}(M)$. In general, the norm $\mathrm{N}(M) = (R : M)$ will not be finite: just consider the module $M = \mathbb{Z} = [1,0]$ in some ring $R = \mathcal{O}_K$. Reducing $a + b\sqrt{m}$ modulo M gives $a + b\sqrt{m} \equiv b\sqrt{m} \bmod M$, and in fact we have $R/M = \{b\sqrt{m} + M : b \in \mathbb{Z}\}$ since $b\sqrt{m} \equiv b'\sqrt{m} \bmod M$ implies $b = b'$. In particular, $(R : M) = \infty$.

This cannot happen if the $\mathbb{Z}$-module M has rank 2. Note that a $\mathbb{Z}$-module $M = [n, c + m\omega)]$ in $\mathcal{O}_K$ has rank 2 if and only if $mn \neq 0$. Modules of maximal rank in $\mathcal{O}_K$ (in the case of quadratic extensions $K/\mathbb{Q}$ this means rank 2) are also called **full** modules. Now we claim

Proposition 3.2. *Let $M = [n, c + m\omega)]$ be a full $\mathbb{Z}$-module in $\mathcal{O}_K$. Then*

$$S = \{r + s\omega : 0 \leq r < n,\ 0 \leq s < m\}$$

is a complete residue system modulo M in $\mathcal{O}_K$, and in particular $\mathrm{N}(M) = mn$.

Proof. We first show that every $x + y\omega \in \mathcal{O}_K$ is congruent mod M to an element of S. Write $y = mq + s$ for some $q \in \mathbb{Z}$ and $0 \leq s < m$; then $x + y\omega - q(c + m\omega) = x' + s\omega$ for some integer x', hence $x + y\omega \equiv x' + s\omega \bmod M$. Now write $x' = nq' + r$ for $q' \in \mathbb{Z}$ and $0 \leq r < n$; then $x' + s\omega \equiv r + s\omega \bmod M$.

Now we claim that the elements of S are pairwise incongruent modulo M. Assume that $r + s\omega \equiv r' + s'\omega \bmod M$ for $0 \leq r, r' < n$ and $0 \leq s, s' < m$; then $r - r' + (s - s')\omega \in M$ implies that $s - s' \in m\mathbb{Z}$ and $r - r' \in n\mathbb{Z}$, hence $r = r'$ and $s = s'$. $\square$

We will also need a second way of characterizing the norm of modules in $\mathcal{O}_K$. In contrast to the results above, which are valid in more general orders (they hold, for example, in rings $\mathbb{Z}[\sqrt{-m}\,]$), this characterization of the norm only holds in the ring of integers $\mathcal{O}_K$ (also called the maximal order). In fact, the following lemma due to Hurwitz exploits that we are working in $\mathcal{O}_K$:

Lemma 3.3. *Let $\alpha, \beta \in \mathcal{O}_K$ and $m \in \mathbb{N}$. If $N\alpha$, $N\beta$ and $\operatorname{Tr} \alpha\beta'$ are divisible by m, then $m \mid \alpha\beta'$ and $m \mid \alpha'\beta$.*

Proof. Put $\gamma = \alpha\beta'/m$; then $\gamma' = \alpha'\beta/m$, and we know that $\gamma + \gamma' = (\operatorname{Tr} \alpha\beta')/m$ and $\gamma\gamma' = \frac{N\alpha}{m}\frac{N\beta}{m}$ are integers. But if the norm and the trace of some γ in a quadratic number field are integral, then we have $\gamma \in \mathcal{O}_K$. $\square$

Remark: the last sentence of the proof demands that any element in $\mathbb{Q}(\sqrt{m}\,)$ with integral norm and trace is in the ring. This means that the lemma holds in any subring of K containing $\mathcal{O}_K$, but not in smaller rings.

Proposition 3.4. *Let K be a quadratic number field with ring of integers $\mathcal{O}_K$ and integral basis $\{1, \omega\}$. If M is a full $\mathbb{Z}$-module in $\mathcal{O}_K$, then there is an $f \in \mathbb{N}$ and a module $\mathcal{O} = [1, g\omega]$ such that $MM' = f\mathcal{O}$ and $N(M) = fg$.*

$\mathbb{Z}$-modules $\mathcal{O} \subseteq \mathcal{O}_K$ containing $\mathbb{Z}$ are also called orders; the order $\mathcal{O}_K$ is, for obvious reasons, called the maximal order.

Proof. Using Proposition 3.1 we can write $M = [\alpha, \beta]$ for $\alpha, \beta \in \mathcal{O}_K$ (actually Prop. 3.1 is more precise, but this is all we need for now). Then $M' = [\alpha', \beta']$ and therefore $MM' = [N\alpha, \alpha\beta', \alpha'\beta, N\beta]$. Now there is some integer $f > 0$ with $f = \gcd(N\alpha, N\beta, \operatorname{Tr} \alpha\beta')$ (in $\mathbb{Z}$); Hurwitz's Lemma shows that $\frac{\alpha\beta'}{f}$ and $\frac{\alpha'\beta}{f}$ are integral; thus we get $MM' = [f][\frac{N\alpha}{f}, \frac{N\beta}{f}, \frac{\alpha\beta'}{f}, \frac{\alpha'\beta}{f}]$ (the generators of this $\mathbb{Z}$-module are all integral by Hurwitz's Lemma). In order to prove $MM' = f\mathcal{O}$ it is therefore sufficient to show that $1 \in [\frac{N\alpha}{f}, \frac{N\beta}{f}, \frac{\alpha\beta'}{f}, \frac{\alpha'\beta}{f}]$. But 1 is a $\mathbb{Z}$-linear combination of $\frac{N\alpha}{f}$, $\frac{N\beta}{f}$ and $\frac{\operatorname{Tr} \alpha\beta'}{f}$ (by the definition of f), hence in particular a $\mathbb{Z}$-linear combination of $\frac{N\alpha}{f}$, $\frac{N\beta}{f}$, $\frac{\alpha\beta'}{f}$ and $\frac{\alpha'\beta}{f}$. This proves the claim. $\square$

3.2 Ideals

An ideal I in some ring R is just a $\mathbb{Z}$-submodule of R that also is an R-module. In other words, I must satisfy $I + I = I$ (closed under addition) and $I \cdot R = I$ (closed under multiplication by ring elements).

The fact that $IR = I$ allows us to make the quotient group R/I into a ring via $(r + I) \cdot (s + I) = rs + I$. In fact, if $r + I = r' + I$ and $s + I = s' + I$, i.e., if $a = r - r' \in I$ and $b = s - s' \in I$, then $r's' + I = (r - a)(s - b) + I = rs + (ab - rb - sa) + I$, and this is equal to the coset $rs + I$ only if $ab - rb - sa \in I$; since $a, b \in I$ implies that $ab \in I$, this is equivalent to $rb + sa \in I$. Since I

is an ideal, we find $sa, rb \in I$, and this implies that multiplication is well defined.

Note that if I and J are ideals in R, then so are

$$\begin{aligned} I + J &= \{i + j : i \in I, j \in J\}, \\ IJ &= \{i_1 j_1 + \ldots + i_n j_n : i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J\}, \end{aligned}$$

as well as $I \cap J$. The index n in the product IJ is meant to indicate that we only form finite sums. If A and B are ideals in some ring R, we say that $B \mid A$ if $A = BC$ for some ideal C.

We say that an nonzero ideal $I \neq R$ is

- irreducible if $I = AB$ for ideals A, B implies $A = R$ or $B = R$;
- a prime ideal if $AB \subseteq I$ for ideals A, B always implies $A \subseteq I$ or $B \subseteq I$;
- a maximal ideal if $I \subseteq J \subseteq R$ for an ideal J implies $J = I$ or $J = R$.

In principal ideal rings, this coincides with the usual usage of prime and irreducible elements: an ideal (a) is irreducible (prime) if and only if a is irreducible (prime). In fact, $(r) \mid (s)$ is equivalent to $r \mid s$. In general domains, r may be irreducible whereas (r) factors into two ideals (necessarily not principal).

Prime ideals and maximal ideals can be characterized as follows:

Proposition 3.5. *An ideal I is*

- *prime in R if and only if R/I is an integral domain;*
- *maximal in R if and only if R/I is a field.*

Proof. R/I is an integral domain if and only if it has no zero divisors. But $0 = (r + I)(s + I) = rs + I$ is equivalent to $rs \in I$; if I is prime, then this implies $r \in I$ or $s \in I$, i.e., $r + I = 0$ or $s + I = 0$, and R/I is a domain. The converse is also clear.

Now let I be maximal and take some $a \in R \setminus I$; we have to show that $a + I$ has a multiplicative inverse. Since I is maximal, the ideal generated by I and a must be the unit ideal, hence there exist elements $m \in I$ and $r, s \in R$ such that $1 = rm + sa$. But then $(a + I)(s + I) = as + I = (1 - rm) + I = 1 + I$.

Conversely, assume that every coset $r + I \neq 0 + I$ has a multiplicative inverse. Then we claim that I is maximal. In fact, assume that M is an ideal strictly bigger than I. Then there is some $m \in M \setminus I$. Pick $r \in R$ sith $(m + I)(r + I) = 1 + I$; then $mr - 1 \in I \subset M$, and $m \in M$ now shows that $1 \in M$. $\square$

Note that an integral domain is a ring with 1 in which $0 \neq 1$; thus (1) is not prime since the null ring R/R only has one element.

It follows from this proposition that every maximal ideal is prime; the converse is not true in general. In fact, consider the ring $\mathbb{Z}[X]$ of polynomials with integral coefficients. Then $I = (X)$ is an ideal, and $R/I \simeq \mathbb{Z}$ is an integral domain but not a field, hence I is prime but not maximal.

Example. Now consider the domain $R = \mathbb{Z}[\sqrt{-5}\,]$ and the ideal $\mathfrak{p} = (2, 1 + \sqrt{-5}\,)$. We claim that $R/\mathfrak{p} \simeq \mathbb{Z}/2\mathbb{Z}$; this will imply that $\mathfrak{p}$ is prime, and even a maximal ideal.

We first prove that every element of R is congruent to 0 or 1 modulo $\mathfrak{p}$. This is easy: reducing $a + b\sqrt{-5}$ modulo 2 shows that every element is congruent to $a + b\sqrt{-5}$ mod (2) with $a, b \in \{0, 1\}$, i.e., to one of 0, 1, $\sqrt{-5}$, $1+\sqrt{-5}$.[1] Reducing these classes modulo $\mathfrak{p}$ we find that $\sqrt{-5} \equiv 1 \bmod \mathfrak{p}$ (the difference is in $\mathfrak{p}$ and $1+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}$). Thus every element is $\equiv 0, 1 \bmod \mathfrak{p}$. Moreover, these residue classes are different since $0 \equiv 1 \bmod \mathfrak{p}$ would imply $1 \in \mathfrak{p}$, which is not true: $1 = \alpha \cdot 2 + \beta \cdot (1 + \sqrt{-5}\,)$ is impossible for $\alpha, \beta \in R$, as a little calculation will show.

An important result is

Theorem 3.6 (Chinese Remainder Theorem)**.** *If A and B are ideals in R with $A + B = R$, then $R/AB \simeq R/A \oplus R/B$ as rings.*

Proof. Since $A + B = R$, there exist $a \in A$ and $b \in B$ such that $a + b = 1$. Consider the map $\phi : R/A \oplus R/B \longrightarrow R/AB$ defined by $\phi(r + A, s + B) = rb + sa + AB$. We claim that ϕ is a ring homomorphism. Checking that $\phi(r + A, s + B) + \phi(r' + A, s' + B) = \phi(r + r' + A, s + s' + B)$ is easy. Multiplication is more tricky: we have

$$\begin{aligned}
\phi(r + A, s + B)\phi(r' + A, s' + B) &= (rb + sa)(r'b + s'a) + AB \\
&= rr'b^2 + ss'a^2 + AB \\
&= rr'b(1 - a) + ss'a(1 - b) + AB \\
&= rr'b + ss'a + AB = \phi(rr' + A, ss' + B).
\end{aligned}$$

In order to show that ϕ is bijective, it is sufficient to define the inverse map $\psi : R/AB \longrightarrow R/A \oplus R/B$ by $\psi(r + AB) = (r + A, r + B)$ and verifying that $\psi \circ \phi$ and $\phi \circ \psi$ are the identity maps; this is again easily done. $\square$

Ideals as $\mathbb{Z}$-Modules

Clearly every ideal in $\mathcal{O}_K$ is a $\mathbb{Z}$-module (and therefore is generated by at most two elements); the converse is not true since e.g. $M = [1, 0] = \mathbb{Z}$ is a $\mathbb{Z}$-module in $\mathcal{O}_K$ but clearly not an ideal: the only ideal containing 1 is the unit ideal $(1) = \mathcal{O}_K$. A different way of looking at this is the following: ideals in $\mathcal{O}_K$ are $\mathcal{O}_K$-modules, and the fact that $\mathbb{Z} \subset \mathcal{O}_K$ implies that every ideal is a $\mathbb{Z}$-module.

Given a $\mathbb{Z}$-module $M = [n, c + m\omega]$, under what conditions on a, m, n is M an ideal? This question is answered by the next

[1] Actually this is a complete set of residue classes modulo $\mathfrak{a} = (2)$ in R. The ring $R/(2)$ has zero divisors because $(1 + \sqrt{-5}\,)^2 = -4 + 2\sqrt{-5} \equiv 0 \bmod (2)$; in particular, (2) is not a prime ideal in R.

Proposition 3.7. *A nonzero $\mathbb{Z}$-module $M = [n, c + m\omega]$ is an ideal if and only if $m \mid n$, $m \mid c$ (hence $c = mb$ for some $b \in \mathbb{Z}$) and $n \mid m \cdot N(b + \omega)$.*

Writing $n = ma$ for some integer a, this shows that ideals can be written in the form $\mathfrak{a} = m[a, b + \omega]$ for integers a, m such that $a \mid N(b + \omega)$.

Proof. Since M is an ideal, $n \in M \cap \mathbb{Z}$ implies $n\omega \in M$. Thus we have $n \in H$ (see the proof of Prop. 3.1) by definition of H. This shows that $n\mathbb{Z} = M \cap \mathbb{Z} \subseteq H = m\mathbb{Z}$, hence $m \mid n$ (if the multiples of n are contained in the multiples of m, then m must divide n; this instance of "to divide means to contain" will reoccur frequently in the following).

In order to show that $m \mid c$ we observe that $\omega^2 = x + y\omega$ for suitable $x, y \in \mathbb{Z}$. Since M is an ideal, $c + m\omega \in M$ implies $(c + m\omega)\omega = mx + (c + my)\omega \in M$, hence $c + my \in H$ by definition of H, and therefore $c + my$ is a multiple of m. This implies immediately that $m \mid c$, hence $c = mb$ for some $b \in \mathbb{Z}$.

In order to prove the last divisibility relation we put $\alpha = c + m\omega = m(b + \omega)$. Then $\alpha \in M$ implies $\alpha(b + \omega') \in M$. Since $\frac{1}{m} N\alpha = m(b + \omega)(b + \omega') \in M \cap \mathbb{Z}$, we conclude that $\frac{1}{m} N(b + \omega)$ is a multiple of n. $\square$

For $\mathbb{Z}$-modules M in $\mathcal{O}_K$ we have shown that $MM' = f\mathcal{O}$ for some module $\mathcal{O}$ containing 1. If M is an ideal, then so is $\mathcal{O}$, and since every ideal containing 1 is the unit ideal, we find

Proposition 3.8. *If $\mathfrak{a}$ is a nonzero ideal in $\mathcal{O}_K$, then there exists some integer $f > 0$ with $\mathfrak{a}\mathfrak{a}' = (f)$.*

In fact, this integer f is nothing but the norm of $\mathfrak{a}$, that is, the number of residue classes in $\mathcal{O}_K/\mathfrak{a}$:

Proposition 3.9. *Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$, and write $\mathfrak{a}\mathfrak{a}' = f\mathcal{O}_K$ for some natural number f. Then $f = \mathrm{N}(\mathfrak{a})$.*

Proof. By Prop. 3.1 we can write $\mathfrak{a} = m[a, b + \omega]$, and we have $\mathrm{N}(\mathfrak{a}) = m^2 a$. It remains to show that $\mathfrak{a}\mathfrak{a}' = (m^2 a)$. To this end, we compute

$$\begin{aligned}
\mathfrak{a}\mathfrak{a}' &= m^2[a, b + \omega][a, b + \omega'] \\
&= m^2[a^2, a(b + \omega), a(b + \omega'), N(b + \omega)] \\
&= m^2 a[a, b + \omega, b + \omega', \tfrac{1}{a} N(b + \omega)].
\end{aligned}$$

The last module is integral because of Proposition 3.7. We want to show that it is the unit ideal. Note that the ideal must be generated by a rational integer since $\mathfrak{a}\mathfrak{a}' = (f)$. But the only integers dividing $b + \omega$ are ± 1 (see the next lemma). $\square$

Lemma 3.10. *If g is an integer dividing $a + b\omega \in \mathcal{O}_K$, then $g \mid b$.*

Proof. We have $g \mid a + b\omega$ if and only if $\frac{a + b\omega}{g} \in \mathcal{O}_K$. But elements of $\mathcal{O}_K$ have the form $c + d\omega$ with integers c, d, hence we conclude that $g \mid a + b\omega$ if and only if $\frac{a}{g}$ and $\frac{b}{g}$ are integers, i.e., if and only if $g \mid a$ and $g \mid b$. $\square$

Proposition 3.9 implies in particular that $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\mathrm{N}(\mathfrak{b})$ because both sides generate the same ideal $\mathfrak{a}\mathfrak{b}\mathfrak{a}'\mathfrak{b}'$. Here are a few more useful properties:

- $N\mathfrak{a} = 1 \iff \mathfrak{a} = (1)$: if $N\mathfrak{a} = 1$, then $(1) = \mathfrak{a}\mathfrak{a}' \subseteq \mathfrak{a} \subseteq \mathcal{O}_K = (1)$, and the converse is clear.
- $N\mathfrak{a} = 0 \iff \mathfrak{a} = (0)$: if $\mathfrak{a}\mathfrak{a}' = (0)$, then $N\alpha = \alpha\alpha' = 0$ for all $\alpha \in \mathfrak{a}$.
- For principal ideals $\mathfrak{a} = (\alpha)$ we have $\mathrm{N}\mathfrak{a} = |\mathrm{N}(\alpha)|$. In fact, $\mathfrak{a}\mathfrak{a}' = (\alpha\alpha') = (\mathrm{N}\alpha)$.

3.3 Unique Factorization into Prime Ideals

We want to show that every ideal in the ring $\mathcal{O}_K$ of integers in a quadratic number field $K = \mathbb{Q}(\sqrt{d}\,)$ can be factored uniquely into prime ideals.

The Cancellation Law

Now we turn to the proof of unique factorization for ideals. The idea behind the proof is the same as in the proof of unique factorization for numbers: from equality of two products, conclude that there must be two equal factors, and then cancel. Now cancelling a factor is the same as multiplying with its inverse; the problem is that we do not have an inverse for ideals.

In the ring $R = \mathbb{Z}/6\mathbb{Z}$ we have $(2)(3) = (2)(0)$, but cancelling (2) yields nonsense. Similar examples exist in all rings with zero divisors. Are there examples in integral domains? Yes, there are. Simple calculations show that $(a,b)^3 = (a^2, b^2)(a,b)$ in arbitrary commutative rings; whenever $(a^2, b^2) \neq (a,b)^2$, we have a counter example to the cancellation law. For an example, take $R = \mathbb{Z}[X,Y]$ and observe that $XY \in (X,Y)^2$, but $XY \notin (X^2, Y^2)$.

The cancellation law even fails in subrings of $\mathcal{O}_K$: consider e.g. the ring $R = \mathbb{Z}[\sqrt{-3}\,]$; then a simple calculation shows that $(2)(2, 1+\sqrt{-3}\,) = (1+\sqrt{-3}\,)(2, 1+\sqrt{-3}\,)$, and cancelling would produce the incorrect statement $(2) = (1+\sqrt{-3}\,)$. It was Dedekind who realized that his ideal theory only works in rings $\mathcal{O}_K$:

Proposition 3.11. *If $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ are nonzero ideals in $\mathcal{O}_K$ with $\mathfrak{a}\mathfrak{b} = \mathfrak{a}\mathfrak{c}$, then $\mathfrak{b} = \mathfrak{c}$.*

Proof. The idea is to reduce the cancellation law for ideals to the one for numbers, or rather for principal ideals.

Thus assume first that $\mathfrak{a} = (\alpha)$ is principal. Then $\alpha\mathfrak{b} = \mathfrak{a}\mathfrak{b} = \mathfrak{a}\mathfrak{c} = \alpha\mathfrak{c}$. For every $\beta \in \mathfrak{b}$ we have $\alpha\beta \in \alpha\mathfrak{c}$, hence there is a $\gamma \in \mathfrak{c}$ such that $\alpha\beta = \alpha\gamma$. This shows $\beta = \gamma \in \mathfrak{c}$, hence $\mathfrak{b} \subseteq \mathfrak{c}$. By symmetry we conclude that $\mathfrak{b} = \mathfrak{c}$.

Now assume that $\mathfrak{a}$ is an arbitrary ideal. Then $\mathfrak{a}\mathfrak{b} = \mathfrak{a}\mathfrak{c}$ implies that $(\mathfrak{a}\mathfrak{a}')\mathfrak{b} = (\mathfrak{a}\mathfrak{a}')\mathfrak{c}$. Since $\mathfrak{a}\mathfrak{a}' = (N\mathfrak{a})$ is principal, the claim follows from the first part of the proof. $\square$

This shows that the ideals in $\mathcal{O}_K$ form a monoid with cancellation law, analogous to the natural numbers.

Divisibility of Ideals

We say that an ideal $\mathfrak{b}$ is divisible by an ideal $\mathfrak{a}$ if there is an ideal $\mathfrak{c}$ such that $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$. Since $\mathfrak{c} \subseteq \mathcal{O}_K$ we see $\mathfrak{b} = \mathfrak{a}\mathfrak{c} \subseteq \mathfrak{a}(1) = \mathfrak{a}$; this fact is often expressed by saying "to divide is to contain". As a matter of fact, the converse is also true:

Proposition 3.12. *If $\mathfrak{a}, \mathfrak{b}$ are nonzero ideals in $\mathcal{O}_K$, then $\mathfrak{a} \supseteq \mathfrak{b}$ if and only if $\mathfrak{a} \mid \mathfrak{b}$.*

Proof. From $\mathfrak{a} \supseteq \mathfrak{b}$ we deduce $\mathfrak{b}\mathfrak{a}' \subseteq \mathfrak{a}\mathfrak{a}' = (a)$, where $a = N\mathfrak{a}$. Then $\mathfrak{c} = \frac{1}{a}\mathfrak{b}\mathfrak{a}'$ is an ideal because of $\frac{1}{a}\mathfrak{a}'\mathfrak{b} \subseteq \mathcal{O}_K$ (the ideal axioms are easily checked) Now the claim follows from $\mathfrak{a}\mathfrak{c} = \frac{1}{a}\mathfrak{b}\mathfrak{a}\mathfrak{a}' = \mathfrak{b}$. □

We know that maximal ideals are always prime, as it is known that $\mathfrak{a}$ is maximal in a ring R if and only if $R/\mathfrak{a}$ is a field, and it is prime if and only if $R/\mathfrak{a}$ is an integral domain.

In the rings of integers in algebraic number fields all three notions coincide; irreducible and maximal ideals are the same:

- irreducible ideals are maximal: if $\mathfrak{a}$ were not maximal, then there were an ideal $\mathfrak{b}$ with $\mathfrak{a} \subsetneq \mathfrak{b} \subsetneq (1)$; this implies $\mathfrak{b} \mid \mathfrak{a}$ with $\mathfrak{b} \neq (1), \mathfrak{a}$.
- maximal ideals are irreducible: for $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ implies $\mathfrak{a} \subsetneq \mathfrak{b} \subsetneq (1)$.

It remains to show that, in our rings, prime ideals are maximal; note that this is not true in general rings. In fact we have to use Proposition 3.12 in the proof.

Proposition 3.13. *In rings of integers of qadratic number fields, prime ideals are maximal.*

Proof. Assume that $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ and $\mathfrak{a} \nmid \mathfrak{b}$; then $\mathfrak{a} \mid \mathfrak{c}$, and since $\mathfrak{c} \mid \mathfrak{a}$ (to divide is to contain) we have $\mathfrak{a} = \mathfrak{c}$ and therefore $\mathfrak{b} = (1)$. □

Observe that from $\mathfrak{a} \mid \mathfrak{c}$ and $\mathfrak{c} \mid \mathfrak{a}$ we cannot conclude equality $\mathfrak{a} = \mathfrak{c}$: we do get $\mathfrak{a} = \mathfrak{c}\mathfrak{d}$ and $\mathfrak{c} = \mathfrak{a}\mathfrak{e}$, hence $\mathfrak{a} = \mathfrak{d}\mathfrak{e}\mathfrak{a}$. But without the cancellation law we cannot conclude that $\mathfrak{d}\mathfrak{e} = (1)$.

In $R = \mathbb{Z}[X]$, the ideal (X) is prime since $\mathbb{Z}[X]/(X) \simeq \mathbb{Z}$ is an integral domain; it is not maximal, since $\mathbb{Z}$ is not a field, and in fact we have $(X) \subset (2, X) \subset R$.

Now we can prove

Theorem 3.14. *Every nonzero ideal $\mathfrak{a}$ in the ring of integers $\mathcal{O}_K$ of a quadratic number field K can be written uniquely (up to order) as a product of prime ideals.*

Proof. We start with showing the existence of a factorization into irreducible ideals. If $\mathfrak{a}$ is irreducible, we are done. If not, then $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$; if $\mathfrak{b}$ and $\mathfrak{c}$ are irreducible, we are done. If not, we keep on factoring. Since $N\mathfrak{a} = N\mathfrak{b}N\mathfrak{c}$ and $1 < N\mathfrak{b},\ N\mathfrak{c} < N\mathfrak{a}$ etc. this process must terminate, since the norms are natural numbers and cannot decrease indefinitely.

Now we prove uniqueness. Assume that $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$ are two decompositions of $\mathfrak{a}$ into prime ideals. We claim that $r = s$ and that we can reorder the $\mathfrak{q}_i$ in such a way that we have $\mathfrak{p}_i = \mathfrak{q}_i$ for $1 \le i \le r$. Since $\mathfrak{p}_1$ is prime, it divides some $\mathfrak{q}_j$ on the right hand side, say $\mathfrak{p}_1 \mid \mathfrak{q}_1$. Since $\mathfrak{q}_1$ is irreducible, we must have equality $\mathfrak{p}_1 = \mathfrak{q}_1$, and the cancellation law yields $\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_s$. The claim now follows by induction. □

3.4 Decomposition of Primes

Now that we know that ideals in $\mathcal{O}_K$ can be factored uniquely into prime ideals, we have to come up with a description of these prime ideals. For quadratic (and, as we will see, also for cyclotomic) fields this is not hard.

Lemma 3.15. *Let $\mathfrak{p}$ be a prime ideal; then there is a unique prime number p such that $\mathfrak{p} \mid (p)$.*

Proof. We have $\mathfrak{p} \mid \mathfrak{p}\mathfrak{p}' = (N\mathfrak{p})$; decomposing $N\mathfrak{p}$ in $\mathbb{Z}$ into prime factors and using the fact that $\mathfrak{p}$ is prime shows that $\mathfrak{p}$ divides (hence contains) some ideal (p) for prime p. If $\mathfrak{p}$ would divide (hence contain) prime ideals (p) and (q) for different primes p and q, it would also contain 1, since p and q are coprime: this implies, by Bezout, the existence of $x, y \in \mathbb{Z}$ with $px + qy = 1$. □

If p is the prime contained in $\mathfrak{p}$, then we say that the prime ideal $\mathfrak{p}$ lies above p. Since (p) has norm p^2, we find that $N\mathfrak{p}$ equals p oder p^2.

Lemma 3.16. *If $\mathfrak{p}$ is an ideal in $\mathcal{O}_K$ with norm p, then it is prime.*

Proof. The ideal is clearly irreducible ($\mathfrak{p} = \mathfrak{a}\mathfrak{b}$ implies $p = N\mathfrak{p} = N\mathfrak{a} \cdot N\mathfrak{b}$), hence prime. □

For describing the prime ideals in quadratic number fields it is useful to have the notion of the discriminant. If $K = \mathbb{Q}(\sqrt{m}\,)$ with m squarefree, let $\{1, \omega\}$ denote an integral basis. We then define

$$\operatorname{disc} K = \left|\begin{smallmatrix} 1 & \omega \\ 1 & \omega' \end{smallmatrix}\right|^2 = (\omega - \omega')^2 = \begin{cases} m & \text{if } m \equiv \ \ 1 \mod 4, \\ 4m & \text{if } m \equiv 2, 3 \mod 4. \end{cases}$$

Theorem 3.17. *Let p be an odd prime, $K = \mathbb{Q}(\sqrt{m}\,)$ a quadratic number field, and $d = \operatorname{disc} K$ its discriminant.*

- *If $p \mid d$, then $p\mathcal{O}_K = (p, \sqrt{m}\,)^2$; we say that p is ramified in K.*

- *If* $(d/p) = +1$*, then* $p\mathcal{O}_K = \mathfrak{p}\mathfrak{p}'$ *for prime ideals* $\mathfrak{p} \neq \mathfrak{p}'$*; we say that* p *splits (completely) in* K*.*
- *If* $(d/p) = -1$*, then* $p\mathcal{O}_K$ *is prime, and we say that* p *is inert in* K*.*

Proof. Assume first that $p \mid d$; since p is odd, we also have $p \mid m$. Now

$$(p, \sqrt{m}\,)^2 = (p^2, p\sqrt{m}, m) = (p)(p, \sqrt{m}, \tfrac{m}{p}) = (p),$$

since the ideal $(p, \sqrt{m}, \frac{m}{p})$ contains the coprime integers p and $\frac{m}{p}$, hence equals (1).

Next assume that $(d/p) = 1$; then $d \equiv x^2 \bmod p$ for some integer $x \in \mathbb{Z}$. Putting $\mathfrak{p} = (p, x + \sqrt{m}\,)$ we find

$$\begin{aligned}\mathfrak{p}\mathfrak{p}' &= (p^2, p(x + \sqrt{m}\,), p(x - \sqrt{m}\,), x^2 - m)\\ &= (p)(p, x + \sqrt{m}, x - \sqrt{m}, \tfrac{x^2-m}{p}).\end{aligned}$$

Clearly $2\sqrt{m} = x + \sqrt{m} - (x - \sqrt{m}\,)$ and therefore $4m = (2\sqrt{m})^2$ are contained in the last ideal; since p and $4m$ are coprime, this ideal equals (1), and we have $\mathfrak{p}\mathfrak{p}' = (p)$. If we had $\mathfrak{p} = \mathfrak{p}'$, then it would follow that $4m \in \mathfrak{p}$ and $\mathfrak{p} = (1)$: contradiction.

Finally assume that $(d/p) = -1$. If there were an ideal $\mathfrak{p}$ of norm p, Proposition 3.7 would show that it has the form $\mathfrak{p} = (p, b + \omega)$ with $p \mid N(b + \omega)$: in fact, we find $\mathfrak{p} = m[a, b + \omega]$ and $N\mathfrak{p} = m^2 a$. Since $N\mathfrak{p} = p$, this implies $m = 1$ and $a = p$, hence $\mathfrak{p} = [p, b + \omega]$ with $p \mid N(b + \omega)$.

If $\omega = \sqrt{m}$, this means $b^2 - m \equiv 0 \bmod p$, hence $(d/p) = (4m/p) = (m/p) = +1$ in contradiction to our assumption. If $\omega = \frac{1}{2}(1 + \sqrt{m}\,)$, then $(2b + 1)^2 \equiv m \bmod p$, and this again is a contradiction. □

The description of all prime ideals above 2 is taken care of by the following

Exercise. Let $K = \mathbb{Q}(\sqrt{m}\,)$ be a quadratic number field, where m is square-free.

- If $m \equiv 2 \bmod 4$ then $2\mathcal{O}_K = (2, \sqrt{m}\,)^2$.
- If $m \equiv 3 \bmod 4$ then $2\mathcal{O}_K = (2, 1 + \sqrt{m}\,)^2$.
- If $m \equiv 1 \bmod 8$ then $2\mathcal{O}_K = \mathfrak{a}\mathfrak{a}'$, where $\mathfrak{a} = (2, \frac{1+\sqrt{m}}{2}\,)$ and $\mathfrak{a} \neq \mathfrak{a}'$.
- If $m \equiv 5 \bmod 8$ then $2\mathcal{O}_K$ is prime.

The two cases p odd and $p = 2$ can be subsumed into one by introducing the *Kronecker-Symbol* (d/p). This agrees with the Legendre symbol for odd primes p and is defined for $p = 2$ and $d \equiv 1 \bmod 4$ by $(d/2) = (-1)^{(d-1)/4}$; for $d \not\equiv 1 \bmod 4$ we put $(d/2) = 0$.

Before we go on, let us recall a few notions from algebra. A domain R is called a principal ideal domain (PID) if every ideal in R is principal. Every Euclidean ring (such as $\mathbb{Z}$, $\mathbb{Z}[i]$, $K[X]$) is a PID, and every PID is a unique factorization domain (UFD). The knowledge that some ring $\mathcal{O}_K$ is a PID would allow us to prove results about the representation of primes by binary quadratic forms:

Proposition 3.18. *Assume that $\mathcal{O}_K$ is a PID, where $K = \mathbb{Q}(\sqrt{m}\,)$. Then every prime p with $(d/p) = +1$ can be written in the form $\pm p = x^2 - my^2$ if $m \equiv 2, 3 \bmod 4$, and in the form $\pm 4p = x^2 - my^2$ if $m \equiv 1 \bmod 4$.*

Proof. Assume that $(d/p) = +1$; then p splits in K, hence $p = \mathfrak{p}\mathfrak{p}'$ for prime ideals $\mathfrak{p}$, $\mathfrak{p}'$ of norm p. Since $\mathcal{O}_K$ is a PID, there is an $\alpha \in \mathcal{O}_K$ such that $\mathfrak{p} = (\alpha)$. Taking the norm show that $(N\alpha) = (p)$ as ideals, hence $N\alpha = \pm p$. The claim now follows by writing $\alpha = x + y\omega$, where $\{1, \omega\}$ is the standard integral basis of $\mathcal{O}_K$. $\square$

If we could show that the rings of integers in $\mathbb{Q}(\sqrt{m}\,)$ for $m = -1$ and $m = -2$ were PIDs (actually this is easy to prove by showing they are Euclidean), this would imply

$$p \equiv 1 \bmod 4 \Longrightarrow p = x^2 + y^2,$$
$$p \equiv 1, 3 \bmod 8 \Longrightarrow p = x^2 + 2y^2,$$

as well as many similar results.

This stresses the importance of finding a method for determining when $\mathcal{O}_K$ is a PID. We will present such a method in the next two chapters: the main ingredients are the unit group and the ideal class group of a number field K.

Exercises

3.1 Compute the matrix M_α for $\alpha = a + b\omega + c\omega^2$ in the cubic number field $\mathbb{Q}(\omega)$ with $\omega^3 = 2$.

3.2 Show that every subgroup A of $\mathbb{Z}$ is automatically a subring and even an ideal in $\mathbb{Z}$, and that there is an $a \in \mathbb{Z}$ such that $A = a\mathbb{Z}$.

3.3 Let $n \in \mathbb{N}$ be a natural number. Find a basis for the ideal (n) in $\mathcal{O}_K$, where $K = \mathbb{Q}(\sqrt{m}\,)$ is a quadratic number field.

3.4 Show that $(3, 1 + \sqrt{-5}\,) = [3, 1 + \sqrt{-5}\,]$ in $R = \mathbb{Z}[\sqrt{-5}\,]$, i.e., that every R-linear combination $3\alpha + (1 + \sqrt{-5}\,)\beta$ with $\alpha, \beta \in R$ can already be written in the form $3a + (1 + \sqrt{-5}\,)b$ with $a, b \in \mathbb{Z}$.

3.5 Show that in $R = \mathbb{Z}[\sqrt{-5}\,]$ we have $R/(\sqrt{-5}\,) \simeq \mathbb{Z}/5\mathbb{Z}$ and deduce that $(\sqrt{-5}\,)$ is a maximal ideal.

3.6 Show that all ideals of prime norm p in $\mathcal{O}_K$ have the form $[p, a + \omega]$, where $p \mid N(a + \omega)$.

3.7 Show that the set of upper triangular 2×2-matrices with coefficients in some ring R is a subring, but not an ideal of the ring of all 2×2-matrices.

3.8 Consider the space S of all sequences of rational numbers. This is a ring with respect to pointwise addition and multiplication:

$$(a_1, a_2, a_3, \ldots) + (b_1, b_2, b_3, \ldots) = (a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots),$$
$$(a_1, a_2, a_3, \ldots) \cdot (b_1, b_2, b_3, \ldots) = (a_1 b_1, a_2 b_2, a_3 b_3, \ldots).$$

Show that the the following subsets of S actually are subrings:

1. the set N of sequences converging to 0;
2. the set D of sequences converging in $\mathbb{Q}$;
3. the set C of Cauchy sequences;
4. the set B of bounded sequences.

Observe that $N \subset D \subset C \subset B \subset S$. Determine which of these subrings are ideals in B (resp. C, D). Show that all of these rings contain zero divisors, and that N is maximal in C (so C/N is a field; actually $C/N \simeq \mathbb{R}$: this is one possible way of constructing the field of real numbers).

4. Units

In this chapter we will determine the unit groups of the rings $\mathcal{O}_K$ of integers in quadratic number fields. We will also show how the knowledge of units allows us to test in finitely many steps whether a given ideal in $\mathcal{O}_K$ is principal.

4.1 The Pell Equation

The determination of the unit group of quadratic number fields is an important task; knowledge of units is needed for solving diophantine equations or for computing the ideal class group. For general commutative rings R, the units form a group $R^\times$; in the following we will determine the structure of the unit group for rings of integers $R = \mathcal{O}_K$ in quadratic number fields.

Lemma 4.1. *Let $K = \mathbb{Q}(\sqrt{m}\,)$ be a quadratic number field. Then an $\varepsilon \in \mathcal{O}_K$ is a unit if and only if $N\varepsilon = \pm 1$.*

Proof. Let $\varepsilon \in \mathcal{O}_k$ be a unit; then $\varepsilon\eta = 1$ for some $\eta \in \mathcal{O}_k$, and taking the norm shows that $N\varepsilon N\eta = N(1) = 1$. Since $N\varepsilon$ and $N\eta$ are integers, we either have $N\varepsilon = N\eta = 1$ or $N\varepsilon = N\eta = -1$.

Conversely, if $N\varepsilon = 1$, then $1 = N(\varepsilon) = \varepsilon\varepsilon'$ shows that ε is a unit. $\square$

Let us make this criterion explicit. If $m \equiv 2, 3 \bmod 4$, then $\varepsilon = t + u\sqrt{m}$, and $N\varepsilon = t^2 - mu^2$. Thus in this case, finding units is equivalent to solving the Pell[1] equation

$$t^2 - mu^2 = \pm 1.$$

For example, $1 + \sqrt{2}$ and $2 + \sqrt{3}$ are units in $\mathbb{Z}[\sqrt{2}\,]$ and $\mathbb{Z}[\sqrt{3}\,]$, respectively.

If $m \equiv 1 \bmod 4$, then we can write $\varepsilon = \frac{t+u\sqrt{m}}{2}$ and find that we have to solve the Pell equation

$$t^2 - mu^2 = \pm 4.$$

Alternatively, we can write $\varepsilon = r + s\omega$ with $\omega = \frac{1+\sqrt{m}}{2}$; since $N\varepsilon = r^2 + rs + \frac{1-m}{4}$, we have to solve the equation

[1] Named by Euler after the British mathematician John Pell, who had nothing to do with this equation first studied by the Indians, and then by Fermat, Wallis and Brouncker.

$$r^2 + rs + \frac{1-m}{4} = \pm 1,$$

which can be transformed into $t^2 - mu^2 = \pm 4$ by multiplying through by 4 and completing the square.

Note that solutions of $t^2 - mu^2 = \pm 4$ give us also solutions of the Pell equation $t^2 - mu^2 = \pm 1$ in the following way: if t and u are even, this is clear (just cancel 2). Assume therefore that $t \equiv u \equiv 1 \bmod 2$; we claim first that $m \equiv 5 \bmod 8$ in this case. In fact, we have $m \equiv 1 \bmod 4$ anyway; if $m \equiv 1 \bmod 8$, then $\pm 4 = t^2 - mu^2$ for odd values of t, u implies, in light of $t^2 \equiv u^2 \equiv 1 \bmod 8$, that $r \equiv \pm 4 \equiv t^2 - mu^2 \equiv 1 - m \bmod 8$, hence $m \equiv 5 \bmod 8$.

Now put $\varepsilon = \frac{t+u\sqrt{m}}{2}$. We claim that $\varepsilon^3 \in \mathbb{Z}[\sqrt{m}\,]$. This will follow from a brute force computation:

$$\begin{aligned}\varepsilon^3 &= \frac{1}{8}(t^3 + 3mtu^2) + \frac{1}{8}(3t^2m + mu^3)\sqrt{m} \\ &= \frac{t}{8}(t^2 + 3mu^2) + \frac{m}{8}(3t^2 + mu^2)\sqrt{m}.\end{aligned}$$

Now $t^2 + 3mu^2 \equiv 1 + 3 \cdot 5 \equiv 0 \bmod 8$ and $3t^2 + mu^2 \equiv 3 + 5 \equiv 0 \bmod 8$ show that the coefficients of ε^3 are integers.

As an example, observe that the unit $\varepsilon = \frac{1+\sqrt{5}}{2}$ corresponding to $1^2 - 5 \cdot 1^2 = -4$ gives $\varepsilon^3 = 2 + \sqrt{5}$, which corresponds to $2^2 - 5 \cdot 1^2 = -1$.

The structure of $\mathcal{O}_K^\times$ for complex quadratic fields is easily determined:

Theorem 4.2. *Assume that $m < 0$ is squarefree, let $K = \mathbb{Q}(\sqrt{m}\,)$, and let $R = \mathcal{O}_K$ denote the ring of integers in K. Then*

$$R^\times = \begin{cases} \langle i \rangle \simeq \mathbb{Z}/4\mathbb{Z} & \text{if } m = -1; \\ \langle -\rho \rangle \simeq \mathbb{Z}/6\mathbb{Z} & \text{if } m = -3; \\ \langle -1 \rangle \simeq \mathbb{Z}/2\mathbb{Z} & \text{otherwise.} \end{cases}$$

Here $i = \sqrt{-1}$ denotes a primitive fourth, and $\rho = \frac{1}{2}(-1 + \sqrt{-3}\,)$ a primitive cube root of unity.

Proof. Assume first that $m \equiv 2, 3 \bmod 4$ and let $\varepsilon = a + b\sqrt{m}$ be a unit. Since norms are positive in complex quadratic fields, this implies $1 = N\varepsilon = a^2 - mb^2$. If $m < -1$, this equation only has the trivial solutions $(a, b) = (\pm 1, 0)$; if $m = -1$, there are exactly four solutions, namely $(a, b) = (\pm 1, 0)$ and $(0, \pm 1)$; The corresponding units in this case are all powers of i. It is easily checked that the map sending i^a to $a \bmod 4$ is a well defined isomorphism between $\mathbb{Z}[i]^\times = \langle i \rangle$ and $\mathbb{Z}/4\mathbb{Z}$.

If $m \equiv 1 \bmod 4$, we put $\varepsilon = \frac{a+b\sqrt{m}}{2}$ and have to solve $4 = a^2 - mb^2$. Again there are only the trivial solutions $(\pm 2, 0)$ for $m < -3$, showing that the only units in this case are ± 1. If $m = -3$, on the other hand, there are six solutions $(\pm 2, 0)$, $(\pm 1, \pm 1)$, giving rise to the six units

$$\pm 1, \quad \pm\frac{-1+\sqrt{-3}}{2}, \quad \pm\frac{1+\sqrt{-3}}{2}.$$

Setting $\rho = \frac{-1+\sqrt{-3}}{2}$ (this is a cube root of unity since $\rho^3 = 1$), then E_K is generated by $-\rho$ (a sixth root of unity). $\square$

Solving the Pell equation for positive values of m is much more difficult. Fermat claimed he could show that $x^2 - my^2 = 1$ always is solvable for non-square positive values of m and challenged the British mathematicians to prove this; since the problem was not clearly formulated, the British mathematicians solved the equation in rational numbers, which is easy (we will get back to this problem later). After Fermat had clarified that he wanted integral solutions, Wallis and Brouncker showed how to solve the equation in finitely many steps, but Fermat later complained that they did not prove that the method always works. As usual, Fermat also did not give any proof thereof: it was Lagrange who first succeeded in giving a complete proof. In due time, the the theory of continued fractions became the standard approach to the Pell equation. Our approach will be different and goes back to Dirichlet.

The idea is to construct many elements with small norm in the hope of finding two elements α and β that not only have the same norm but that actually generate the same principal ideal. In fact, we have $(\alpha) = (\beta)$ if and only if $\frac{\alpha}{\beta}$ is a unit (possibly a trivial one, though).

Here's how it looks in practice: for finding a unit in $\mathbb{Z}[\sqrt{11}\,]$, we solve lots of equations $x^2 - 11y^2 = n$ for integers n with small absolute value. For $y = 1$, the expression $x^2 - 11y^2$ will be small if $x \approx \sqrt{11}$, i.e. for $x = 3$ and $x = 4$. We find

$$\begin{aligned} 3^2 - 11 &= -2, \\ 4^2 - 11 &= +5. \end{aligned}$$

We continue by trying $y = 2$ and $x \approx 2\sqrt{11}$, i.e.

$$\begin{aligned} 6^2 - 11 \cdot 2^2 &= -8, \\ 7^2 - 11 \cdot 2^2 &= +5. \end{aligned}$$

Thus $4 \pm \sqrt{11}$ and $7 \pm 2\sqrt{11}$ all have norm 5. Which of these elements generate the same ideal? One way to find out is by computing the quotients. We have

$$\frac{7+2\sqrt{11}}{4+\sqrt{11}} = \frac{(7+2\sqrt{11})(4-\sqrt{11})}{(4+\sqrt{11})(4-\sqrt{11})} = \frac{6+\sqrt{11}}{5},$$

which is not an algebraic integer, showing that $(7+2\sqrt{11}\,)$ and $(4+\sqrt{11}\,)$ generate different prime ideals above 5. Next

$$\frac{7+2\sqrt{11}}{4-\sqrt{11}} = \frac{(7+2\sqrt{11})(4+\sqrt{11})}{(4+\sqrt{11})(4-\sqrt{11})} = \frac{50+15\sqrt{11}}{5} = 10+3\sqrt{11},$$

and we have found a unit $\varepsilon = 10 + 3\sqrt{11}$.

Since trial and error is somewhat unsatisfactory, let us see how we could have predicted that $7+2\sqrt{11}$ and $4-\sqrt{11}$ were the right elements to consider. We know that these elements generate ideals of norm 5, i.e., they must all be prime ideals above 5. Now there are only two of these, namely $\mathfrak{5}_1 = (5, 1+\sqrt{11})$ and $\mathfrak{5}_2 = (5, 1+\sqrt{11})$. Thus $\sqrt{11} \equiv -1 \bmod \mathfrak{5}_1$ and $\sqrt{11} \equiv +1 \bmod \mathfrak{5}_2$, therefore

$$\begin{aligned} 7 + 2\sqrt{11} &\equiv 0 \bmod \mathfrak{5}_1, \\ 7 + 2\sqrt{11} &\equiv 4 \bmod \mathfrak{5}_2, \\ 4 + \sqrt{11} &\equiv 3 \bmod \mathfrak{5}_1, \\ 4 + \sqrt{11} &\equiv 0 \bmod \mathfrak{5}_2 \end{aligned}$$

etc., showing that $(7 + 2\sqrt{11}) = (4 - \sqrt{11}) = \mathfrak{5}_1$.

Another way we could have compute a nontrivial unit here is by observing that $(2) = \mathfrak{2}^2$ is ramified in K. Since $3 + \sqrt{2}$ has norm -2, we must have $\mathfrak{2} = (3 + \sqrt{11})$, and now $(2) = \mathfrak{2}^2 = (3 + \sqrt{11})^2 = (20 + 6\sqrt{11})$ shows that $\frac{20+6\sqrt{11}}{2} = 10 + 3\sqrt{11}$ is a unit.

4.2 Solvability of the Pell Equation

Here's a modernized version of Dirichlet's standard proof found in most textbooks. The idea is the following: there are only finitely many integral ideals of bounded norm in $\mathbb{Q}(\sqrt{m})$; if we can construct sufficiently many elements with bounded norm, then there must be two that generate the same ideal and therefore differ by a unit.

The idea is to construct a sequence of algebraic integers $\alpha_j = x_j + y_j\sqrt{m}$ (m a positive squarefree integer) with $|N\alpha_j| < B$. Eventually there will be two elements α_i and α_j generating the same ideal, and their quotient $\varepsilon = \alpha_i/\alpha_j$ will then be a unit. In order to make sure that $\varepsilon \neq \pm 1$ we construct the α_j in such a way that $\alpha_1 > \alpha_2 > \ldots > \alpha_k > \ldots$. (Our proof uses the fact that the number of ideals in $\mathcal{O}_K$ with bounded norm is finite; this can also be proved more generally for rings $\mathbb{Z}[\sqrt{m}]$ for general nonsquare m, and then the proof below shows the solvability of the Pell equation $x^2 - my^2 = 1$ for all nonsquare numbers $m < 0$).

This is achieved in exactly the same way as above for $m = 11$: we consider the sequence $y = 0, 1, \ldots, N$ and let x denote the smallest integer $> y\sqrt{m}$; then $0 < x - y\sqrt{m} \leq 1$ and $x + y\sqrt{m} < BN$ for $B = \lceil 2\sqrt{m} \rceil$. Since there are $N + 1$ such numbers $x - y\sqrt{m}$ in the interval $(0, 1)$, Dirichlet's box principle guarantees the existence of pairs (a, b) and (a', b') with $0 < (a - b\sqrt{m}) - (a' - b'\sqrt{m}) < \frac{1}{N}$. Putting $x = a - a'$ and $y = b - b'$ we find $0 < x - y\sqrt{m} < \frac{1}{N}$ and $0 < |x + y\sqrt{m}| < BN$. Thus we can find numbers $x - y\sqrt{m}$ with positive

absolute value as small as we wish, but in such a way that $\mathrm{N}(x - y\sqrt{m}\,) < B$ is bounded.

Now we can construct our sequence of α_j. We start with $\alpha_1 = 1$. Assume we have already found α_i for $i = 1, \ldots, k-1$ with

$$\alpha_1 > \alpha_2 > \ldots > \alpha_{k-1} > 0$$

and $|\mathrm{N}(\alpha_i)| < B$. By the argument above we can find $\alpha_k = x - y\sqrt{m}$ with $0 < \alpha_k < \alpha_{k-1}$ and $|\mathrm{N}(\alpha_k)| < B$.

Since there are only finitely many integral ideals with norm $< B$, there must exist $i < j$ with $(\alpha_i) = (\alpha_j)$. But then $\varepsilon = \alpha_i/\alpha_j > 1$ is a unit, and we have proved:

Theorem 4.3. *Every real quadratic field has units $\neq \pm 1$. In particular, the equation $X^2 - mY^2 = 1$, where $m > 1$ is an integer, has integral solutions with $y > 0$.*

Now that we know that the Pell equation is solvable, let us compute the structure of the unit group in $\mathcal{O}_K$:

Theorem 4.4. *Let $K = \mathbb{Q}(\sqrt{m}\,)$ be a real quadratic number field with $m > 0$ squarefree. Then*

$$E_K = \mathcal{O}_K^\times \simeq \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}.$$

In other words, every unit $\varepsilon \in E_K$ can be written uniquely in the form $\varepsilon = (-1)^a \eta^b$ with $a \in \mathbb{Z}/2\mathbb{Z}$ and $b \in \mathbb{Z}$, where η is the smallest unit in E_K with $\eta > 1$.

Proof. We first have to prove that there is a smallest unit $\eta > 1$. If not, then there is a sequence of units $\eta_1 > \eta_2 > \ldots > 1$; then $0 < |\eta_i'| = 1/\varepsilon_i < 1$, hence if we write $\eta_j = x_j + y_j\sqrt{m}$, we find $2|x_j| = |\eta_j + \eta_j'| \leq |\eta_j| + |\eta_j'| < \eta_1 + 1$: this shows that there are only finitely many choices for x, and the same argument with η_j' replaced by $-\eta_j'$ shows that the same holds for y_j. This is a contradiction.

Now let $\varepsilon > 1$ be any unit. If $\varepsilon = \eta^n$ for some integer n we are done; if not, then there is some $n \in \mathbb{N}$ with $\eta^n < \varepsilon < \eta^{n+1}$. But then $\upsilon = \varepsilon\eta^{-n}$ is a unit in $\mathcal{O}_K$ with $1 < \upsilon < \eta$, contradicting the choice of η.

Thus every unit > 1 has the form η^n for some $n \in \mathbb{N}$. If $0 < \varepsilon < 1$, then $1/\varepsilon > 1$, hence $\varepsilon = \eta^n$ for some integer $m < 0$. Finally, if $\varepsilon < 0$, then $-\varepsilon > 0$ has the form η^n. This proves that every unit can be written as $\pm\eta^n$. $\square$

Simple Consequences

We have just seen that the Pell equation $x^2 - my^2 = 1$ has nontrivial integral solutions whenever m is not a square; next we will give a few simple consequences of the solvability of the Pell equation.

Below, the following argument will be used repeatedly:

Lemma 4.5. *Let $a, b, c, m \in \mathbb{N}$ be integers, m squarefree, such that $ab = mc^2$, and assume that $d = \gcd(a, b)$. Then $a = rdx^2$, $b = sdy^2$ for $r, s, x, y \in \mathbb{N}$ with $rs = m$ and $dxy = c$.*

Proof. Put $\alpha = \frac{a}{m}$, $\beta = \frac{b}{m}$, and $\gamma = \frac{c}{m}$. Then $\alpha\beta = m\gamma^2$, and $\gcd(\alpha, \beta) = 1$. Next put $r = \gcd(\alpha, m)$ and $s = \gcd(\beta, m)$.

Then $rs = m$: in fact, write $m = p^a q^b \cdots$; then $p^a || \alpha\beta$, and since $\gcd(\alpha, \beta) = 1$ we conclude that either $p^a \mid \alpha$ or $p^a \mid \beta$, hence $p^a || \gcd(\alpha, m) \gcd(\beta, m) = rs$. The claim now follows from the observation primes dividing rs must divide m.

Now $\frac{\alpha}{r}\frac{\beta}{s} = \gamma^2$, and the factors on the left are coprime. Thus they are perfect squares, that is, $\alpha = rx^2$ and $\beta = sy^2$, and finally $a = d\alpha = rdx^2$ and $b = d\beta = sdy^2$. □

Now we claim

Proposition 4.6. *If $p \equiv q \equiv 3 \bmod 4$ are primes, then $px^2 - qy^2 = \pm 1$ is solvable in integers for some choice of signs.*

Proof. Consider $K = \mathbb{Q}(\sqrt{m})$ for $m = pq$, and let $\eta = t + u\sqrt{m}$ correspond to the minimal nontrivial solution of the Pell equation $t^2 - mu^2 = 1$. Since $m \equiv 1 \bmod 4$, we see that t is odd and u is even. Next $mu^2 = t^2 - 1 = (t-1)(t+1)$; we claim that $\gcd(t-1, t+1) = 2$. Clearly both numbers are even, hence it is sufficient to show that the gcd divides 2. But this is clear, since $\gcd(t-1, t+1)$ divides the difference $t + 1 - (t-1) = 2$.

Thus with $u = 2w$ we get $mw^2 = \frac{t-1}{2}\frac{t+1}{2}$, and since the factors on the right are coprime, unique factorization implies that we must have one of the following:

$$\begin{aligned} t+1 &= 2r^2 & \qquad t-1 &= 2ms^2, \\ t+1 &= 2pr^2 & \qquad t-1 &= 2qs^2, \\ t+1 &= 2qr^2 & \qquad t-1 &= 2ps^2, \\ t+1 &= 2mr^2 & \qquad t-1 &= 2s^2. \end{aligned}$$

Here r and s are nonzero integers that we may and will choose positive. Substracting these equations from each other and dividing through by 2 we find

$$\begin{aligned} 1 &= r^2 - ms^2, \\ 1 &= pr^2 - qs^2, \\ 1 &= qr^2 - ps^2, \\ 1 &= mr^2 - s^2. \end{aligned}$$

The first equation contradicts the minimality of (t, u) because $t = r^2 + ms^2$ shows that $0 < r < t$. The last equation is also impossible since it implies

$-s^2 \equiv 1 \bmod p$, i.e., $-1 \equiv s^2 \bmod p$: but $p \equiv 3 \bmod 4$, so -1 is a quadratic nonresidue modulo p. Thus the second or the third of these equations must be solvable, which is what we wanted to prove. □

This simple result implies a piece of the quadratic reciprocity law:

Corollary 4.7. *If $p \equiv q \equiv 3 \bmod 4$ are prime, then $(\frac{p}{q}) = -(\frac{q}{p})$.*

Proof. Consider a solution of the equation $pr^2 - qs^2 = \pm 1$. Switching the roles of p and q if necessary we may assume that the plus sign holds. But then we get the congruences $pr^2 \equiv 1 \bmod q$ and $-qs^2 \equiv 1 \bmod p$, which in turn imply $(p/q) = +1$ and $(q/p) = -1$. □

Here is another famous result:

Proposition 4.8. *If $p \equiv 1 \bmod 4$ is prime, then the negative Pell equation $t^2 - pu^2 = -1$ is solvable.*

Proof. Again, pick a minimal solution of $x^2 - py^2 = 1$ and write $py^2 = (x-1)(x+1)$. For the same reason as above, $\gcd(x+1, x-1) = 2$, and we find that one of the following two sets of equations hold:

$$\begin{aligned} x+1 &= 2r^2, & x-1 &= 2ps^2 \\ x+1 &= 2pr^2, & x-1 &= 2s^2. \end{aligned}$$

This implies as before either $r^2 - ps^2 = 1$ (contradicting the minimality of the chose solution) or $pr^2 - s^2 = 1$, thus proving the claim. □

For certain families of quadratic fields it is easy to write down units explicitly. In fact, assume that $m = n^2 - 1$ is squarefree for some even integer n. Then $m \equiv 3 \bmod 4$, hence the fundamental unit comes from the minimal solution of $t^2 - my^2 = 1$. But the minimal solution is clearly $(t, u) = (n, 1)$, and we see:

Proposition 4.9. *Let n be an even integer and assume that $m = n^2 - 1$ is squarefree. Then $\eta = n + \sqrt{m}$ is the fundamental unit of $\mathbb{Q}(\sqrt{m}\,)$.*

4.3 Principal Ideal Tests

Assume that we are given an ideal $\mathfrak{a}$ in $\mathcal{O}_K$; how can we tell whether it is principal?

This is a finite task for complex quadratic fields. Take, for example, the ideal $\mathfrak{z} = (2, \omega)$ for $\omega = \frac{1+\sqrt{-m}}{2}$ and some $m > 0$ with $-m \equiv 1 \bmod 8$. Then $\mathfrak{z}$ is an ideal of norm 2. If it were principal, there would exist elements with norm 2 (norm -2 is impossible in the complex quadratic case). But $N(\frac{a+b\sqrt{-m}}{2}) = 2$ is equivalent to $a^2 + mb^2 = 8$, and this equation only has solutions for $m = 7$.

Proposition 4.10. *Let $m \equiv 7 \bmod 8$ be a positive squarefree integer. Then the prime ideals above 2 in $\mathbb{Q}(\sqrt{-m})$ are principal if and only if $m = 7$.*

In real quadratic fields, testing whether a given ideal is principal is a much less trivial task. Consider e.g. the case $m = 79$ and the ideal $(3, 1+\sqrt{79})$. This ideal is principal if and only if one of the equations $x^2 - 79y^2 = \pm 3$ is solvable. As a matter of fact, $x^2 - 79y^2 = 3$ is impossible modulo 4 since $3 = x^2 - 79y^2 \equiv x^2 + y^2 \bmod 4$. Thus the question is: does $x^2 - 79y^2 = -3$ have a solution?

Here it is not obvious how to check this in finitely many steps. Just plugging in $y = 1, 2, 3 \ldots$ will not help since we do not know where to stop.

Here the unit group comes to our rescue. Consider a real quadratic number field $K = \mathbb{Q}(\sqrt{m})$ assume that $\alpha = a + b\sqrt{m}$ has norm $N\alpha = n$, and let $\varepsilon > 1$ be a nontrivial unit in $\mathcal{O}_K$. Then we can choose an integer m in such a way that $1 \le |\alpha\varepsilon^m| < \varepsilon$. Put $\beta = x + y\sqrt{m}$; since $N(\beta) = N(\alpha)N(\varepsilon) = \pm n$, we find $|\beta'| = \frac{|\beta\beta'|}{|\beta|} = \frac{|n|}{|\beta|} < |n|$. But then $2|x| = |\beta + \beta'| \le |\beta| + |\beta'| < \varepsilon + |n|$, and similarly $2|y|\sqrt{m} < \varepsilon + |n|$.

In our case $m = 79$, the fact that $N(9+\sqrt{79}) = 2$ implies that $\varepsilon = \frac{1}{2}(9+\sqrt{79})^2 = 80 + 9\sqrt{79}$ is a unit (actually ε is fundamental). Since $n = -3$, we find that $2|y|\sqrt{79} < \varepsilon + 3 < 163$ and therefore $|y| \le 9$. Checking all the values of y between 0 and 9 shows that $\mathfrak{z}_1 = (3, 1+\sqrt{79})$ is not principal. With a little bit more effort, we can prove in the same way that $\mathfrak{z}_1^2$ is not principal. On the other hand, the fact that $N(17+2\sqrt{79}) = -27$ shows that $\mathfrak{z}_1^3 = (17+2\sqrt{79})$ is principal. In fact, the relation $17 + 2\sqrt{79} = 3 \cdot 5 + 2(1+\sqrt{79})$ implies $17 + 2\sqrt{79} \in \mathfrak{z}_1$; moreover, $17 + 2\sqrt{79}$ is not divisible by 3, hence cannot be contained in $\mathfrak{z}_2$.

Actually we can easily improve our bounds by choosing m more cleverly. In fact, we might just as well pick m in such a way that

$$\frac{\sqrt{|n|}}{\sqrt{\varepsilon}} \le |\alpha\varepsilon^m| < \sqrt{|n|\varepsilon}.$$

Then, with $\beta = \alpha\varepsilon^m$, we get $|\beta'| = \frac{|\beta\beta'|}{|\beta|} < \sqrt{|n|\varepsilon}$, and this implies

$$|y| < \frac{\sqrt{|n|\varepsilon}}{\sqrt{m}}.$$

As a matter of fact, using the following lemma due to Cassels we can do still better:

Lemma 4.11. *Suppose that the positive real numbers x, y satisfy the inequalities $x \le s$, $y \le s$, and $xy \le t$. Then, $x + y \le s + t/s$.*

Proof. $0 \le (x-s)(y-s) = xy - s(x+y) + s^2 \le s^2 + t - s(x+y)$. □

Putting $x = |\alpha|$ and $y = |\alpha'|$ in Lemma 4.11 we find

$$|2a| \le |\alpha| + |\alpha'| < \sqrt{n\varepsilon} + \sqrt{n/\varepsilon},$$

and likewise

$$|2b\sqrt{m}\,| = |\beta - \beta'| \le |\beta| + |\beta'| < \sqrt{n\varepsilon} + \sqrt{n/\varepsilon}.$$

We have proved

Proposition 4.12. *Let $k = \mathbb{Q}(\sqrt{m}\,)$ be a real quadratic number field, $\varepsilon > 1$ a unit in k, and $0 \ne n = |N\xi\,|$ for $\xi \in k$. Then there is a unit $\eta = \varepsilon^j$ such that $\xi\eta = a + b\sqrt{m}$ and*

$$|a| < \frac{\sqrt{n}}{2}(\sqrt{\varepsilon} + 1/\sqrt{\varepsilon}\,), \qquad |b| < \frac{\sqrt{n}}{2\sqrt{m}}(\sqrt{\varepsilon} + 1/\sqrt{\varepsilon}\,).$$

Note that if $m \equiv 1 \bmod 4$, the number y may be half an integer!

4.4 Elements of small norms

After these preparations, it is an easy matter to prove the following result originally due to Davenport:

Proposition 4.13. *Let m, n, t be natural numbers such that $m = t^2 + 1$; if the diophantine equation $|x^2 - my^2| = n$ has solutions in $\mathbb{Z}$ with $n < 2t$, then n is a perfect square.*

Proof. Let $\xi = x + y\sqrt{m}$; then $|\,N\xi\,| = n$, and since $\varepsilon = t + u\sqrt{m} > 1$ is a unit in $\mathbb{Z}[\sqrt{m}\,]$, we can find a power η of ε such that $\xi\eta = a + b\sqrt{m}$ has coefficients $a,\ b$ which satisfy the bounds in Proposition 2.2. Since $2t < \varepsilon < 2\sqrt{m}$, we find

$$|b| \le \frac{\sqrt{n}}{2\sqrt{m}}\left(\sqrt{\varepsilon} + \frac{1}{\sqrt{\varepsilon}}\right) < 1 + \frac{1}{t}.$$

Since the assertion is trivial if $t = 1$, we may assume that $t \ge 2$, and now the last inequality gives $|b| \le 1$. If $b = 0, |N\xi| = a^2$ would be a square; therefore, $b = \pm 1$, and this yields $\alpha = \xi\eta = a \pm \sqrt{m}$. Now $|N\xi| = |N\alpha| = |a^2 - m|$ is minimal for values of a near $\sqrt{m}$, and we find

$$\begin{array}{lcll} |a^2 - m| & = & 2t & \text{if } a = t - 1; \\ |a^2 - m| & = & 1 & \text{if } a = t; \\ |a^2 - m| & = & 2t & \text{if } a = t + 1. \end{array}$$

This proves the claim.

Using the idea in the proof of Proposition 4.13 one can easily show more:

Proposition 4.14. *Let m, n, t be natural numbers such that $m = t^2 + 1$; if the diophantine equation $|x^2 - my^2| = n$ has solutions in $\mathbb{Z}$ with $n < 4t + 3$, then $n = 4t - 3$, $n = 2t$, or n is a perfect square.*

In [1], Proposition 4.13 was used to show that the ideal class group of $k = \mathbb{Q}(\sqrt{m}\,)$ has non-trivial elements (i.e. classes that do not belong to the genus class group) if $m = t^2 + 1$ and $t = 2lq$ for $l > 1$ and prime q: since $m \equiv 1 \bmod q$, q splits in k, i.e. we have $(q) = \mathfrak{p}\mathfrak{p}''$. If $\mathfrak{p}$ were principal, the equation $x^2 - my^2 = \pm 4q$ would have solutions in $\mathbb{Z}$; but since $4q < 2t = 4lq$ is no square, this contradicts Proposition 4.12.

4.5 Computing the Fundamental Unit

The computation of units in quadratic number rings is a difficult and very interesting task. Part of the interest in these calculations stems from the fact that knowing the fundamental unit of $\mathbb{Q}(\sqrt{m}\,)$ often allows us to factor m: from $x^2 - my^2 = 1$ we get $my^2 = x^2 - 1 = (x-1)(x+1)$, and $\gcd(m, x-1)$ is a (possibly trivial) factor of m. For example, the fundamental unit for $m = 91$ is $\varepsilon = 1574 + 165\sqrt{91}$, and $\gcd(91, 1573) = 13$. Thus, as a general rule, computing a solution of the Pell equation $x^2 - my^2 = 1$ is at least as hard as factoring m (and definitely harder if m happens to be a large prime).

Now let us see how our method for computing the fundamental unit works for some larger values of m, say $m = 3431$. We start by collecting elements with small norms:

α	$\mathrm{N}\alpha$
$55 + \sqrt{m}$	$-2 \cdot 7 \cdot 29$
$56 + \sqrt{m}$	$-5 \cdot 59$
$57 + \sqrt{m}$	$-2 \cdot 7 \cdot 13$
$58 + \sqrt{m}$	-67
$59 + \sqrt{m}$	$-2 \cdot 5^2$
$60 + \sqrt{m}$	13^2
$61 + \sqrt{m}$	$2 \cdot 5 \cdot 29$
$62 + \sqrt{m}$	$7 \cdot 59$
$63 + \sqrt{m}$	$2 \cdot 269$

By the way: by coincidence, $60^2 - m = 13^2$ is a square; this shows that $m = 60^2 - 13^2 = (60-13)(60+13) = 47 \cdot 73$. This is happens only rarely after so few computations, but is the basic idea in Fermat's method of facoring.

It seems that not all primes occur as factors: certainly none of these numbers is divisible by 3: this is because there is no ideal of norm 3 in $\mathcal{O}_K$. In fact, $x^2 - my^2 \equiv 0 \bmod p$ implies $(\frac{m}{p}) \neq -1$, so none of the primes with $(\frac{m}{p}) = -1$, such as $p = 3, 11, 17 \ldots$ can occur in these factorizations. The others do, but waiting until we find two elements with the same norm (let alone generating the same ideal) will take forever.

It is a better idea to construct elements generating the same ideal by multiplying together several elements of small norm. Here's how we do this: first we list all the prime ideals in $\mathcal{O}_K$ with small norm: $\mathfrak{2} = (2, 1 + \sqrt{m}\,)$, $\mathfrak{5}_1 = (5, 1 + \sqrt{m}\,)$, $\mathfrak{5}_2 = (5, 1 - \sqrt{m}\,)$, $\mathfrak{7} = (7, 1 + \sqrt{m}\,)$, $\mathfrak{7}_2 = (7, 1 - \sqrt{m}\,)$.

Then we factor the elements of small norm and keep only those which factor over our "factor base":

α	2	5_1	5_2	7_1	7_2
$1+\sqrt{m}$	1	1	0	3	0
$1-\sqrt{m}$	1	0	1	0	3
$41+\sqrt{m}$	1	3	0	0	1
$41-\sqrt{m}$	1	0	3	1	0
$59+\sqrt{m}$	1	0	2	0	0
$59-\sqrt{m}$	1	2	0	0	0

The first line in this table represents the ideal factorization

$$(1+\sqrt{m}\,) = 2^1 \cdot 5_1^1 \cdot 7_1^3.$$

Picking a factor base as small as ours is not a good idea in practice, since there will only be very few "smooth" elements, i.e. elements that factor over the factor base.

If we look carefully at this table we see that $(1+\sqrt{m}\,)(41+\sqrt{m}\,)^3$ has factorization $2^4 5_1^{10} 7_1^3 7_2^3$. Since $2^2 = (2)$ and $7_1 7_2 = (7)$, we find that the element

$$\frac{(1+\sqrt{m}\,)(41+\sqrt{m}\,)^3}{2^2 \cdot 7^3} = 21549 + 364\sqrt{m}$$

has the prime ideal factorization 5_1^{10}. But now $(59-\sqrt{m}\,)^5 = 2^5 5^{10}$ shows that

$$\alpha = \frac{(59-\sqrt{m}\,)^5}{21549+364\sqrt{m}} = 49316884 - 841948\sqrt{m}$$

is an integer with ideal factorization 2^5. Since this ideal is ramified, the element $\varepsilon = 2^5\alpha^{-2}$ must be a unit, and we find

$$\varepsilon = 152009690466840 + 2595140740627\sqrt{m}.$$

Now

$$\gcd(152009690466841, 3431) = 1,$$
$$\gcd(152009690466839, 3431) = 3431,$$

so the fundamental unit does not give us any factor of m.

Note that this method not only gave us a nontrivial unit, it also gave us what is called a "compact presentation":

$$\varepsilon = \frac{2(1+\sqrt{m}\,)^2(41+\sqrt{m}\,)^6}{7^6(59-\sqrt{m}\,)^{10}}.$$

Finally let us remark that it was our knowledge of prime ideal factorization in quadratic number fields that has allowed us to compute this unit.

Now that we know a nontrivial unit, how can we be sure it is the fundamental unit? In any case we know that $\varepsilon = \pm\eta^m$ for some $m \in \mathbb{Z}$, where $m \in \mathbb{Z}$. Since $\varepsilon > 1$, we see that the plus sign holds and that $m \geq 1$. Clearly ε is not a square (we can see this from the compact representation), which shows that ε is twice a square. Of course we can check by hand that ε is not a p-th power for $p = 3, 5, 7, 11, \ldots$, but we do not know how far we have to carry on with these tests.

Here is how to achieve this:

Lemma 4.15. *Let $\varepsilon > 1$ be the fundamental unit of a real quadratic number field with discriminant d. Then*

$$\log \varepsilon > \begin{cases} \log d^{1/2} & \text{if } N\varepsilon = +1, \\ \log(d^{1/2} - 1) & \text{if } N\varepsilon = +1. \end{cases}$$

Proof. Assume that $K = \mathbb{Q}(\sqrt{m}\,)$ with $m \equiv 2, 3 \bmod 4$ and $N\varepsilon = +1$. Then the minimal value of ε is $a + \sqrt{m}$ with $a \approx \sqrt{m}$; since $N\varepsilon = +1$, we must have $a \geq \sqrt{m}$, and this shows that $\varepsilon \geq 2\sqrt{m} = \sqrt{d}$.

The other cases are treated similarly. □

In our case, $m = 3431 = 47 \cdot 73$; since m is divisible by a prime $47 \equiv 3 \bmod 4$, we find that $N\varepsilon = +1$, and this gives us $\log \varepsilon \geq 4.763\ldots$, hence $m = \log \varepsilon / \log \eta \leq 33.3/4.763 = 6.991\ldots$; this shows that $m \leq 6$, and since we already know that ε is not a square, we even have $m \leq 5$.

Thus it remains to test whether ε is a cube or a fifth power. The easiest way to do this by hand is by looking for a prime ideal $\mathfrak{p}$ such that $\varepsilon \bmod \mathfrak{p}$ is not a cube etc. Now $\varepsilon \equiv 0 - 3\sqrt{m} \equiv 3 \bmod 5_1$ shows again that ε is not a square since 3 is not a square modulo 5_1 because $(\frac{3}{5}) = -1$.

For showing that ε is not a cube we need a prime ideal of norm $\equiv 1 \bmod 3$. Then $\varepsilon \equiv 3 + \sqrt{m} \equiv 2 \bmod 7_1$, and since 2 is not a cube modulo 7, it is not a cube modulo 7_1 since $\mathcal{O}_K/7_1 \simeq \mathbb{Z}/7\mathbb{Z}$. Thus ε is not a cube.

Finally, the prime ideal $\mathfrak{q} = (61, 25 + \sqrt{m}\,)$ of norm 61. We find $\varepsilon \equiv 40 - 3\sqrt{m} \equiv 54 \bmod \mathfrak{q}$. A tedious calculation shows that 54 is not a fifth power modulo 61.

If you prefer working with real numbers instead of residues modulo prime ideals, here's what you do. Compute the real numbers

$$\varepsilon \approx 304019380933680.00000, \qquad 1/\varepsilon \approx 3.289 \cdot 10^{-15}.$$

Clearly $\varepsilon + 1/\varepsilon$ is an integer: this follows from $\varepsilon = a + b\sqrt{m}$ and $1/\varepsilon = a - b\sqrt{m}$. Now assume that ε is a fifth power: taking fifth roots shows that we must have $\eta \approx 788.098052\ldots$ and $1/\eta \approx 0.0012688776\ldots$. Again, $\eta + 1/\eta$ must be an integer, but we find $\eta + 1/\eta \approx 788.0993\ldots$. Thus η is not a fifth power, and the cases $m = 2, 3$ can be treated analogously.

Remark. The calculations above were made using `pari`; note, however, that the computation of the compact presentation of ε could have easily done by

hand! `pari` is an absolutely indispensible tool for working with number fields. Here, the command

```
r = quadgen(4*3431)
```

defines the algebraic number $r = \sqrt{3431}$ (`quadgen` only takes discriminants as an input). Then typing in

```
(1+r)*(41+r)^3/(2^2*7^3)
```

produces the desired output

```
21549 + 364*r
```

For a list of all number theoretic commands, enter

```
?4
```

for a short description of a command, enter

```
?quadgen
```

As an exercise, find out what `quadunit(3431)` is doing. You can quit `pari` by typing in

```
\q
```

4.6 Factoring

The same idea we used for computing the fundamental unit can be used to directly factor n. As an example, take $n = 4469$ and factor the integers $a^2 - n$ for $a \approx \sqrt{4469} = 66.85\ldots$ into primes. Keep the factorizations that factor over some small number base of primes p with $(\frac{n}{p}) \neq 1$ such as $p = 2, 5, 11, \ldots$ (but including the "prime" -1):

a	-1	2	5
62	1	0	4
63	1	2	3
67	0	2	1

The first line represents the factorization $62^2 - n = -5^4$.

Here the idea is to find a solution to $a^2 \equiv b^2 \bmod n$; if we have such a pair of integers, then $\gcd(a-b, n)$ and $\gcd(a+b, n)$ are (possibly trivial) factors of n. Note that $(63^2 - n)(67^2 - n) = -2^4 5^4$ implies that $63^2 67^2 \equiv -2^4 5^4 \bmod n$; moreover, $62^2 \equiv -5^4 \bmod n$, hence $63^2 \cdot 67^2 \equiv 4^2 62^2 \bmod n$, and we find $\gcd(63 \cdot 67 - 4 \cdot 62, n) = 1$: no luck.

Increasing our factor base we find

a	-1	2	5	11	13
62	1	0	4	0	0
63	1	2	3	0	0
67	0	2	1	0	0
71	0	2	0	1	1
72	0	0	1	1	1
83	0	2	1	2	0

Now we see that $67^2 72^2 \equiv 71^2 \cdot 5^2 \bmod n$, but this gives us the trivial factorization again. Finally, $67^2 \cdot 11^2 \equiv 83^2 \bmod n$ gives us $\gcd(67 \cdot 11 - 83, n) = 109$, and in fact we have $n = 41 \cdot 109$.

Finding relations is of course just a matter of linear algebra: interpret the exponents in the factorizations as elements of an $\mathbb{F}_2$-vector space; then finding squares corresponds to finding linear dependences in the factorizations. For example, the $\mathbb{F}_2$-vectors of the factorizations of $67^2 - n$ and $83^2 - n$ both are $(0, 0, 1, 0, 0)$.

The factorization method based on this idea is called the quadratic sieve and was the best method for factoring integers without a small prime factors before the number field sieve was invented by Pollard.

Exercises

4.1 Find the elements of smallest nontrivial norm in other simplest quadratic number fields.

4.2 Use the results of the preceding exercise to find examples of quadratic fields with large class number.

4.3 Show that if $m = 2p$ for some prime $p \equiv 5 \bmod 8$, then the fundamental unit of $\mathbb{Q}(\sqrt{m}\,)$ has negative norm.

4.4 Show that if $m = 2p$ with $p \equiv 3 \bmod 4$ prime, then either $x^2 - my^2 = 2$ or $x^2 - my^2 = -2$ is solvable in integers.

4.5 Show that if $m = 2p$ with $p \equiv 3 \bmod 4$ prime, then 2ε is a square in $\mathcal{O}_K$, where ε is the fundamental unit of $K = \mathbb{Q}(\sqrt{m}\,)$.

4.6 Compute the fundamental units of $\mathbb{Q}(\sqrt{m}\,)$ for $m = 3$, 19, 43, 67, 131, 159, 199.

5. The Ideal Class Group

The two most important groups associated to number fields are the unit group $E_K = \mathcal{O}_K^\times$ studied in Chapter 4, and the ideal class group $\mathrm{Cl}(K)$ that we will study next. The intimate relation between these invariants will become clear through Dedekind's zeta function associated to K.

5.1 Class Group

Definition

We have seen that the set of nonzero ideals in $\mathbb{Z}_K$ form a monoid with cancellation law. Such monoids can be made into groups by imitating the construction of $\mathbb{Z}$ from $\mathbb{N}$ (or that of $\mathbb{Q}$) from $\mathbb{Z}$); the group I_K of these fractional ideals contains the group $H_K = \{(\alpha) : \alpha \in K^\times\}$ of principal ideals as a subgroup, and the quotient group $\mathrm{Cl}(K) = I_K/H_K$ is called the class group of K. This group is trivial if and only if $\mathbb{Z}_K$ is a PID. The order $h(K)$ of $\mathrm{Cl}(K)$ is called the class number of K.

We can avoid this formal procedure by introducing fractional ideals as actual sets: write $\mathfrak{a}\mathfrak{b}^{-1} = \mathfrak{a}\mathfrak{b}'(\mathfrak{b}\mathfrak{b}')^{-1} = \frac{1}{b}\mathfrak{a}\mathfrak{b}$, where $b = N\mathfrak{b}$ denotes the norm of $\mathfrak{b}$, and define $\frac{1}{\alpha}\mathfrak{c} := \{\frac{\gamma}{\alpha} : \gamma \in \mathfrak{c}\}$. The set of nonzero fractional ideals forms a group with respect to multiplication; note that the inverse of the integral ideal $\mathfrak{a}$ is the fractional ideal $\mathfrak{a}^{-1} = \frac{1}{a}\mathfrak{a}'$ with $a = N\mathfrak{a}$.

In these notes, we choose a third possibility: we define an equivalence relation on the set of all integral ideals and then make the equivalence classes into a group.

To this end, let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals; they are called equivalent ($\mathfrak{a} \sim \mathfrak{b}$) if there exist $\alpha, \beta \in \mathbb{Z}_K$ such that $\alpha\mathfrak{a} = \beta\mathfrak{b}$. Checking the usual axioms (symmetry, reflexivity, transitivity) is left as an exercise.

On the set of equivalence classes of ideals we define a multiplication as follows: given classes c und d, we pick representatives $\mathfrak{a} \in c$ and $\mathfrak{b} \in d$, and then put $c \cdot d = [\mathfrak{a}\mathfrak{b}]$. This definition does not depend on the choice of representatives; the class of the unit ideal is the neutral element; and finally the fact that $\mathfrak{a}\mathfrak{a}' = (a)$ shows that $[\mathfrak{a}]^{-1} = [\mathfrak{a}']$.

Thus the ideal classes $[\mathfrak{a}]$ form an abelian group $\mathrm{Cl}(K)$. If this group is trivial, then every ideal is equivalent to (1), that is, every ideal is principal.

Since the converse is also clear, we see that $\mathbb{Z}_K$ is a PID if and only if K has class number 1.

Consider e.g. the ring $R = \mathbb{Z}[\sqrt{-5}\,]$; here we have the classes $1 = [(1)]$ und $c = [\mathfrak{a}]$ mit $\mathfrak{a} = (2, 1+\sqrt{-5}\,)$. We have $c^2 = 1$ since $\mathfrak{a}^2 = (2)$ ist $c^2 = 1$. Putting $\mathfrak{b} = (3, 1+\sqrt{-5}\,)$ we find $\mathfrak{a} \sim \mathfrak{b}$: in fact, $\mathfrak{a}\mathfrak{b} = (1+\sqrt{-5}\,)$ implies $\mathfrak{a}\mathfrak{b} \sim (1)$, hence $[\mathfrak{b}] = [\mathfrak{a}]^{-1} = [\mathfrak{a}]$. More calculations seem to suggest that there are only two classes, that is, the class number of R seems to be 2.

The goal of this section is to show that $\mathrm{Cl}(K)$ is finite and to give an algorithm for computing it. The finiteness of the class group is one of three important finiteness theorems in algebraic number theory:

- $\mathrm{Cl}(K)$ is finite;
- $E_K = \mathbb{Z}_K^\times$ is a finitely generated abelian group;
- given a $B > 0$, the set of number fields with discriminant $< B$ is finite.

Finiteness of the Class Number

We now show that every ideal class in $\mathrm{Cl}(k)$ contains an integral ideal with norm bounded by a constant depending only on k; this immediately implies the finiteness of the class number.

Let us call an ideal in $\mathbb{Z}_K$ primitive if it is not divisible by a rational integer $m > 1$. Clearly every ideal class is represented by a primitive ideal.

According to Proposition 3.7, every ideal $\mathfrak{a}$ has a $\mathbb{Z}$-basis of the form $\{n, m(b+\omega)\}$ with $m \mid n$; Thus $\mathfrak{a}$ is primitive if and only if $m = 1$. In other words: if $\mathfrak{a}$ is primitive, then there exist $n \in \mathbb{N}$ and $b \in \mathbb{Z}$ such that $\mathfrak{a} = n\mathbb{Z} \oplus (b+\omega)\mathbb{Z}$, and we have $N\mathfrak{a} = n$. Now we claim:

Theorem 5.1. *Let $m \in \mathbb{Z}$ be squarefree, $K = \mathbb{Q}(\sqrt{m}\,)$ a quadratic field with ring of integers $\mathcal{O}_K = \mathbb{Z}[\omega]$ and discriminant d. Define the Gauss bound*

$$\mu_K = \begin{cases} \sqrt{d/5}, & \text{if } d > 0, \\ \sqrt{-d/3}, & \text{if } d < 0. \end{cases}$$

Then every ideal class in $\mathrm{Cl}(K)$ contains an integral nonzero ideal with norm $\leq \mu_K$; in particular, the number $h = \#\,\mathrm{Cl}(K)$ of ideal classes is finite.

The bounds are clearly best possible: for $d = 5$ and $d = -3$ they are sharp. If $\mu_K \leq 2$, then every ideal class contains a nonzero integral ideal with norm < 2; but then the norm must be 1, hence every ideal class contains the unit ideal, and we deduce that $h = 1$ and that $\mathcal{O}_K$ is a PID. Theorem 5.1 says that this is true for $-12 \leq d \leq 20$, i.e. for $m \in \{-11, -7, -3, -2, -1, 2, 3, 5, 13, 17\}$.

Exercise. If $d \equiv 5 \bmod 8$, then (2) is inert, hence there are no ideals of norm 2 in $\mathcal{O}_K$. Show that this implies that the fields with $d = -19, 21, 29, 37$ have class numberl 1. Which fields do you get by demanding in addition that (3) be inert (that is, $d \equiv 2 \bmod 3$)?

Now consider $R = \mathbb{Z}[\sqrt{-5}\,]$, where $d = -20$; according to Theorem 5.1, every ideal class contains a nonzero ideal with norm $< \sqrt{20/3}$, hence ≤ 2. Since there are only two such ideals, namely the unit ideal (1) and the nonprincipal ideal $(2, 1+\sqrt{-5}\,)$, we deduce that R has class number 2.

Actually we can show more: we have seen that $\mathrm{Cl}(K)$ is generated by the classes of (1) and $\mathfrak{a} = (2, 1+\sqrt{-5}\,)$. Now let p be a prime with $(-20/p) = +1$; then $p\mathbb{Z}_K = \mathfrak{p}\mathfrak{p}'$ for some prime ideal $\mathfrak{p}$ with norm p. Then $\mathfrak{p}$ is either principal, say $\mathfrak{p} = (a + b\sqrt{-5}\,)$ and thus $p = a^2 + 5b^2$, or $\mathfrak{p} \sim \mathfrak{a}$, and then $\mathfrak{a}\mathfrak{p} = (C + d\sqrt{-5}\,)$ is principal. In the latter case we get $2p = C^2 + 5d^2$; since C and d are both odd, we can write $C = 2c + d$ for some $c \in \mathbb{Z}$ and find $2p = (2c+d)^2 + 5d^2 = 4c^2 + 4cd + 6d^2$, that is, $p = 2c^2 + 2cd + 3d^2$. In other words: if $(-5/p) = +1$, then $p = a^2 + 5b^2$ or $p = 2c^2 + 2cd + 3d^2$.

Since $p = a^2 + 5b^2 \equiv a^2 + b^2 \equiv 1 \bmod 4$, this can only happen if $p \equiv 1 \bmod 20$. Similarly, $p = 2c^2+2cd+3d^2 \equiv 3 \bmod 4$, that is, $p \equiv 11, 19 \bmod 20$. We have proved:

Theorem 5.2. *Primes $p \equiv 1, 9 \bmod 20$ are represented by the quadratic form x^2+5y^2, whereas primes $p \equiv 11, 19 \bmod 20$ are represented by $2x^2+2xy+3y^2$.*

An important consequence of Theorem 5.1 is the following observation:

Corollary 5.3. *Let $K = \mathbb{Q}(\sqrt{m}\,)$ be a quadratic number field with class number h, and assume that $p\mathcal{O}_K = \mathfrak{p}\mathfrak{p}'$ splits completely in $\mathcal{O}_K$. Then there exist $x, y \in \mathbb{N}$ such that $\pm 4p^h = x^2 - my^2$.*

Proof. The h-th power of any ideal in $K = \mathbb{Q}(\sqrt{m}\,)$ is principal. In particular, $\mathfrak{p}^h = (\frac{x+y\sqrt{m}}{2})$ for suitable integers x, y, and taking the norm yields $p^h = |\frac{x^2-my^2}{4}|$. $\square$

Proof of Theorem 5.1. Let $c = [\mathfrak{a}]$ be an ideal class represented by an ideal $\mathfrak{a}$. We may and will assume that $\mathfrak{a}$ is primitive. Therefore $\mathfrak{a} = (a, \alpha)$ with $a = N\mathfrak{a}$ and $\alpha = b + \omega = s + \frac{1}{2}\sqrt{d}$ for some $s \in \mathbb{Q}$ with $2s \in \mathbb{Z}$. If $a \leq \mu_K$, we are done; if not, we apply the Euclidean algorithm to the pair (s, a) and find $q \in \mathbb{Z}$ such that $s - qa = r$ and

$$|r| \leq \frac{a}{2} \text{ if } d < 0,$$
$$\frac{a}{2} \leq |r| \leq a \text{ if } d > 0.$$

Setting $\alpha_1 = r + \frac{1}{2}\sqrt{d}$ we find $\alpha_1 \in \mathfrak{a}$, $|N\alpha_1| \leq \frac{1}{4}(a^2 - d) < a^2$, and $\mathfrak{a}_1 := \frac{1}{a}\alpha_1'\mathfrak{a} \sim \mathfrak{a}$ is an integral ideal with $[\mathfrak{a}_1] = [\mathfrak{a}]$ and $N\mathfrak{a}_1 < N\mathfrak{a}$. We repeat this step until we find an ideal of norm $\leq \mu_K$; since the norm decreases with each step, the algorithm terminates.

The proof of the inequality $|N\alpha_1| \leq \frac{1}{4}(a^2 - d) < a^2$ is simple: if $d < 0$, then $|N\alpha_1| = |r^2 - \frac{d}{4}| \leq \frac{a^2+|d|}{4} < 1$ since $a^2 > \mu_K^2 = \frac{|d|}{3}$, and if $d > 0$, we have $-a^2 = \frac{a^2-5a^2}{4} < r^2 - \frac{d}{4} < a^2$.

It remains to show that the ideal $\mathfrak{a}_1$ is integral; but this is clear in light of $\frac{1}{a}\alpha_1'\mathfrak{a} \subseteq \mathcal{O}_K \iff \alpha'\mathfrak{a} \subseteq (a) = \mathfrak{a}\mathfrak{a}' \iff (\alpha') \subseteq \mathfrak{a}'$. $\square$

5.2 Computation of Class Groups

With a little practice, computing class groups of quadratic number fields can be fun (Gauss computed class groups of fields with discriminant down to $-10,000$). Here we will indicate how to proceed for small discriminants.

We will also use the notation $(a, b, \ldots)$ for the abelian group $\mathbb{Z}/a\mathbb{Z} \oplus \mathbb{Z}/b\mathbb{Z} \oplus \cdots$.

$d = -23$

Here every ideal class contains an ideal of norm 1 or 2, hence all ideal classes are given by 1, $[\mathfrak{2}_1]$ and $[\mathfrak{2}_2]$, where $\mathfrak{2}_1 = (2, \omega)$ and $\mathfrak{2}_2 = (2, \omega')$. As usual, $\{1, \omega\}$ denotes the standard integral basis, i.e., we have $\omega = \frac{-1+\sqrt{-23}}{2}$. The ideal $\mathfrak{2}_1$ is not principle since $\mathcal{O}_K$ does not contain elements of norm 2 (the equation $a^2+23b^2 = 8$ is not solvable in integers). Of course $\mathfrak{2}_1 \cdot \mathfrak{2}_2 = (2) \sim (1)$, so $[\mathfrak{2}_2] = [\mathfrak{2}_1]^{-1}$. This shows us that the class group is generated by $[\mathfrak{2}_1]$; since there are exactly three classes, we conclude that $\mathrm{Cl}(K) \simeq \mathbb{Z}/3\mathbb{Z}$. In fact, $\mathfrak{2}_1^3 = (\frac{3-\sqrt{-23}}{2}) = (2-\omega)$; note that $(2-\omega) \subset (2, \omega)$, so $(2-\omega)$ really is $\mathfrak{2}_1^3$ and not $\mathfrak{2}_2^3$.

Now let us consider primes p with $(\frac{-23}{p}) = +1$. They split into $(p) = \mathfrak{p}\mathfrak{p}'$; for some primes p, the ideals $\mathfrak{p}$ and $\mathfrak{p}'$ will be principal, and for others not. Can these primes be characterized? The answer is yes, but actually lies quite deep. In fact, consider the polynomial $f(x) = x^3 - x + 1$. Factoring this polynomial over $\mathbb{F}_p$ shows e.g. that $f(X)$ is irreducible modulo 13, but that $f(x) \equiv (x+4)(x+13)(x-17) \bmod 59$. On the other hand, the prime ideals above 13 are not principal, whereas $(6 + \sqrt{-23})$ has norm 59.

This is no accident; in fact we have

Proposition 5.4. *Let $K = \mathbb{Q}(\sqrt{-23})$, and let p be a prime such that $(\frac{-23}{p}) = +1$. Then the polynomial $f(x) = x^3 - x + 1$ of discriminant -23 splits into three linear factors over $\mathbb{F}_p$ or is irreducible according as there are elements of norm p in $\mathcal{O}_K$ or not.*

Actually, this is a consequence of class field theory, the theory of abelian extensions of number fields. In any case it shows that for understanding quadratic extensions, we also have to study number fields of higher degree.

$d = -30$

We know that the class group is generated by ideals with norm ≤ 6. The only prime ideals with norm ≤ 6 are $\mathfrak{2} = (2, \sqrt{-30})$, $\mathfrak{3} = (3, \sqrt{-30})$, and

$5 = (5, \sqrt{-30})$. These are all ramified: $2^2 = (2)$, $3^2 = (3)$, $5^2 = (5)$, hence $2^2 \sim 3^2 \sim 5^2 \sim 1$. The factorization $\sqrt{-30} = 2 \cdot 3 \cdot 5$ provides us with the additional relation $2 \cdot 3 \cdot 5 \sim 1$, i.e., $5 \sim 2 \cdot 3$. Thus every ideal class contains one of the following ideals: (1), 2, 3, $2 \cdot 3$, and $K = \mathbb{Q}(\sqrt{-30})$ has class number ≤ 4.

We now show that these classes are all different. In fact, none of 2, 3, $2 \cdot 3$ can be principle since there are no elements of norm 2, 3 or 6 in $\mathcal{O}_K$. Moreover, $2 \sim 3$ would imply $2 \cdot 3 \sim 3^2 \sim (1)$, which is wrong. Thus the class number is exactly 4.

The only group with 4 elements and exponent 2 is Klein's four group $V_4 = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, hence we have $\mathrm{Cl}(K) \simeq (2, 2)$.

$d = 4 \cdot 478$

Here $\mathcal{O}_K = \mathbb{Z}[\sqrt{478}]$, and every ideal class contains an ideal of norm ≤ 9. Here are the prime ideals of norm ≤ 7:

- $2 = (2, \sqrt{478})$, $2^2 = (2)$;
- $3_1 = (3, 1 + \sqrt{478})$, $3_2 = (3, 1 + \sqrt{478})$, $3_1 \cdot 3_2 = (3)$;
- $7_1 = (7, 3 + \sqrt{478})$, $7_2 = (7, 3 - \sqrt{478})$, $7_1 \cdot 7_2 = (7)$.

Now we need relations between these ideals. We find some by factoring elements of small norm; such elements can be found in the vicinity of $a + \sqrt{478}$ with $a \approx \sqrt{478} = 21.863\ldots$. In fact, we find

- $(22+\sqrt{478})$ has norm 6, so it must be equal to 3_1 or 3_2. Now $22 \equiv 1 \bmod 3$ shows that $22 + \sqrt{478} \in (3, 1 + \sqrt{478})$, and we conclude that $3_1 \sim 2^{-1}$ (and therefore $3_2 \sim 2$).
- We need relations involving 2 and 7_j. We find elements with norm divisible by 7 as follows: any $\alpha = a + \sqrt{478}$ with $a \equiv 3 \bmod 7$ is divisible by 7_1, and by 7_2 if $a \equiv 4 \bmod 7$. The norm of such elements is small around $a = 22$, and we easily find that $(17 + \sqrt{478}) = 3_2^3 7_1$.
  Similarly we get $(24 + \sqrt{478}) = 2 \cdot 7_1^2$ and $(25 + \sqrt{478}) = 3_1 \cdot 7_2^2$, as well as $(10 + \sqrt{478}) = 2 \cdot 3_1^3 7_1$.

Now the real computation begins. From $2 \cdot 3_1 \sim 1$ and $2^2 = (2)$ we conclude that $3_1^2 \sim 1$, and we can actually compute generators: from $2^2 3_1^2 = (22 + \sqrt{478})^2$ we deduce that $3_1^2 = \frac{1}{2}(22 + \sqrt{478})^2 = (481 + 22\sqrt{478})$.

Also, $3_2^3 7_1 \sim (1)$ implies $7_1 \sim 3_1^3 \sim 3_1$. Next $3_1 \cdot 7_2^2 \sim (1)$ implies $7_1^2 \sim 3_1$, and we conclude that $7_1 \sim 7_2 \sim (1)$. But now everything collapses, and we find that K has class number 1.

Repeating the above reasoning with actual numbers will give us a generator α for 2, and then $\frac{1}{2}\alpha^2 = \varepsilon$ will be a nontrivial unit. Let's do this now.

We have $(25 + \sqrt{478}) = 3_1 \cdot 7_2^2$, $(17 - \sqrt{478}) = 3_1^3 7_2$, as well as $3_1^2 = \frac{1}{2}(22 + \sqrt{478})^2$. Thus $3_1 \cdot 7_2 = \left(\frac{2(17-\sqrt{478})}{(22+\sqrt{478})^2}\right)$, and therefore

$$\mathfrak{7}_2 = \left(\frac{(25+\sqrt{478}\,)(22+\sqrt{478}\,)^2}{2(17-\sqrt{478}\,)}\right).$$

The actual calculation shows that $\mathfrak{7}_2$ is generated by the element $4635 + 212\sqrt{478}$.

Finally, $(24-\sqrt{478}\,) = \mathfrak{2}\mathfrak{7}_2^2$ gives $\mathfrak{2} = \left(\frac{24-\sqrt{478}}{(4635+212\sqrt{478}\,)^2}\right)$, hence

$$\varepsilon = \left(\frac{(24-\sqrt{478}\,)^2}{2(4635+212\sqrt{478}\,)^4}\right)$$

is a unit; in fact, $\varepsilon = 1617319577991743 - 73974475657896\sqrt{478}$ is the inverse of the fundamental unit of K.

5.3 The Bachet-Mordell Equation

Let us now see what we can say about the integral solutions of the diophantine equation $y^2 = x^3 - d$ (named after Bachet and Mordell, who studied them). We will start with arbitrary d, but will impose conditions on d as we go along.

We start by factoring the equation over $K = \mathbb{Q}(\sqrt{d}\,)$:

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

What can we say about the gcd of the ideals $\mathfrak{a} = (y+\sqrt{-d}\,)$ and $\mathfrak{a}'$? Any common prime factor $\mathfrak{p}$ (with $\mathfrak{p} \mid p$) also divides $2\sqrt{-d}$; since $\mathfrak{p} \mid \sqrt{-d}$ (and $p \neq 2$) implies $p \mid d$, $p \mid y$, $p \mid x$ and finally $p^2 \mid d$, we can exclude this possibility by demanding that d be $\boxed{\text{squarefree}}$.

We now have to discuss the remaining possibility $\mathfrak{p} \mid 2$:

- $d \equiv 2 \bmod 4$: then $\mathfrak{p} \mid (\sqrt{-d}\,)$ (since $\mathfrak{p} = (2, \sqrt{-d}\,)$), hence $\mathfrak{p} \mid y$, $p \mid y$ and finally $x^3 = y^2 + d \equiv 2 \bmod 4$: contradiction, since cubes cannot be divisible exactly by 2.
- $d \equiv 1 \bmod 4$: here $\mathfrak{p} = (2, 1+\sqrt{-d}\,)$, hence $\mathfrak{p} \mid (y+\sqrt{-d}\,)$ if and only if y is odd. This implies $x^3 = y^2 + d \equiv 1 + 1 \equiv 2 \bmod 4$, which again is a contradiction.
- $d \equiv 3 \bmod 4$: here $y+\sqrt{-d}$ is divisible by $\mathfrak{p}$ (even by 2) if y is odd. Then $d = x^3 - y^2$ implies that x is even, hence $d \equiv -y^2 \equiv -1 \bmod 8$. Thus if we assume that $\boxed{d \not\equiv 7 \bmod 8}$, find that no $\mathfrak{p} \mid 2$ can be a common divisor of $\mathfrak{a}$ and $\mathfrak{a}'$.

Thus $\mathfrak{a}$ and $\mathfrak{a}'$ are coprime. Since their product is a cube, there exists an ideal $\mathfrak{b}$ such that $\mathfrak{a} = \mathfrak{b}^3$; conjugation then shows that $\mathfrak{a}'^3 = \mathfrak{b}'^3$.

Now let h denote the class number of $\mathbb{Q}(\sqrt{-d}\,)$. Since both $\mathfrak{b}^3$ as well as $\mathfrak{b}^h$ are principal, we can conclude that $\mathfrak{b}$ is principal if we assume that $\boxed{3 \nmid h}$. Thus $\mathfrak{b} = \left(\frac{r+s\sqrt{-d}}{2}\right)$ with $r \equiv s \bmod 2$.

In the case $\boxed{d > 0,\ d \neq 1, 3}$ the only units are ± 1, hence the ideal equation yields the equation of numbers

$$y + \sqrt{-d} = \left(\frac{r + s\sqrt{-d}}{2}\right)^3,$$

where we have subsumed the sign into the cube. Comparing coefficients now yields $1 = \frac{1}{8}(3r^2 s - ds^3)$, hence $8 = 3r^2 s - ds^3 = s(3r^2 - ds^2)$.

This implies $s \mid 8$, hence $s = \pm 1$ or $r \equiv s \equiv 0 \bmod 2$. In the first case we get $\pm 8 = 3r^2 - d$, hence $d = 3r^2 \mp 8$; in the second case we put $r = 2t$, $s = 2u$ and find $1 = u(3t^2 - du^2)$, that is $u = \pm 1$ and $d = 3t^2 \mp 1$.

Thus we have shown: if d, under the above assumptions, does not have the form $3t^2 \pm 1$ or $3t^2 \pm 8$, then the diophantine equation $y^2 = x^3 - d$ does not have an integral solution.

What happens if d has this form? Assume e.g. that $d = 3r^2 - 8$; then comparing coefficients (using $s = 1$) yields $8y = r^3 - 3dr = r^3 - 9r^3 + 24r = 24r - 8r^3$, that is $y = (3 - r^2)r$, as well as $y^2 + d = r^6 - 6r^4 + 12r^2 - 8 = (r-2)^3$, hence $x = r - 2$. Thus $d = 3r^2 - 8$ yields the solution $(r^2 - 2, \pm(3 - r^2)r)$ of our diophantine equation. Similarly, other representations yield other solutions: $d = 3r^2 + 8$, $3t^2 + 1$, $3t^2 - 1$ gives rise to the solutions $(r^2 + 2, \pm r(r^2 + 3))$, $(4t^2 + 1, \pm t(8t^2 + 3))$, $(4t^2 - 1, \pm t(8t^2 - 3))$.

The only question that remains is: can d have more than one of these representations? The answer is: $d = 11$ has exactly two representations, all other d have at most one. The proof is simple: equations such as $3r^2 - 8 = 3t^2 - 1$ are impossible modulo 3; $3r^2 - 8 = 3t^2 + 1$ leads to $3(r^2 - t^2) = 9$, hence $r^2 - t^2 = (r - t)(r + t) = 3$, whose only solution is $r = \pm 2$, $t = \pm 1$, which leads to $d = 4$, but this is not squarefree; the possibility $3r^2 + 8 = 3t^2 - 1$ yields $3 = t^2 - r^2$, hence $t = \pm 2$, $r = \pm 1$ and thus $d = 3 + 8 = 3 \cdot 2^2 - 1 = 11$).

We have proved:

Theorem 5.5. *Let $d \neq 1, 3$ be a squarefree natural number, and assume that $d \not\equiv 7 \bmod 8$. If the class number of $\mathbb{Q}(\sqrt{-d}\,)$ is not divisible by 3, then the diophantine equation $y^2 = x^3 - d$ has*

1. *exactly two pairs of integral solutions $(3, \pm 4)$ and $(15, \pm 58)$ for $d = 11$;*
2. *exactly one pair of integral solutions if $d \neq 11$ has the form $d = 3t^2 \pm 1$ or $d = 3t^2 \pm 8$, namely:*

$$(x, y) = \begin{cases} (4t^2 - 1, \pm t(8t^2 - 3)) & \text{if } d = 3t^2 - 1, \\ (4t^2 + 1, \pm t(8t^2 + 3)) & \text{if } d = 3t^2 + 1, \\ (t^2 - 2, \pm t(3 - t^2)) & \text{if } d = 3t^2 - 8, \\ (t^2 + 2, \pm t(t^2 + 3)) & \text{if } d = 3t^2 + 8. \end{cases}$$

3. *no integral solutions otherwise.*

Consider the case $d = 26 = 3 \cdot 3^2 - 1$: the equation $y^2 = x^3 - 26$ has the predicted solution $(207, \pm 42849)$ as well as $(3, \pm 1)$. The theorem implies that the class number of $\mathbb{Q}(\sqrt{-26})$ must be divisible by 3; in fact we have $h = 6$.

This can be generalized:

Proposition 5.6. *Let u be an odd integer, and put $d = 27u^6 - 1$. If d is squarefree, then $\mathbb{Q}(\sqrt{-d})$ has class number divisible by 3.*

Proof. We have $d = 3t^2 - 1$ for $t = 3u^3$, and Thm. 5.5 predicts the integral solutions $(4t^2 - 1, \pm t(8t^2 - 3))$ of $y^2 = x^3 - d$. In addition, there is the solution $(3u^2, 1)$, hence one of the conditions of the theorem is not satisfied. Since $d \not\equiv 7 \bmod 8$, we conclude that the class number of $\mathbb{Q}(\sqrt{-d})$ must be divisible by 3. $\square$

Similarly it can be proved that the integral solutions of $x^p + y^p = z^p$ are only the trivial solutions if p does not divide the class number of $\mathbb{Q}(\zeta_p)$ – this is Kummer's approach to Fermat's problem.

5.4 Quadratic Reciprocity

Genus theory of quadratic number fields K is an elementary special case of class field theory that predicts the structure of $\mathrm{Cl}(K)/\mathrm{Cl}(K)^2$, that is, the 2-rank of $\mathrm{Cl}(K)$. We do not have time to develop this beautiful theory here (see e.g. Chapter 2 in my Reciprocity Laws), but I want to give you at least an idea of what is going on.

Proposition 5.7. *For an odd prime p, put $p^* = (-1)^{(p-1)/2}p$, that is, $p^* = p$ for $p \equiv 1 \bmod 4$ and $p^* = -p$ if $p \equiv 3 \bmod 4$. Then the quadratic number field $K = \mathbb{Q}(\sqrt{p^*})$ with discriminant p^* has odd class number.*

If K has even class number, then by Cauchy's theorem there must exist an element of order 2. Thus for proving that h_K is odd we need to show that any ideal $\mathfrak{a}$ with $\mathfrak{a}^2 \sim (1)$ is principal.

From $\mathfrak{a}^2 \sim (1)$ and $\mathfrak{a}\mathfrak{a}' = (N\mathfrak{a}) \sim (1)$ we deduce that $\mathfrak{a} \sim \mathfrak{a}'$. Thus there exists some $\alpha \in K^\times$ with $\alpha\mathfrak{a} = \mathfrak{a}'$. If $N\alpha < 0$ (this can only happen if K is real), we replace α by $\alpha\varepsilon$, where ε is the fundamental unit (we know it has norm -1). Thus we may assume that $N\alpha > 0$. Taking the norm of $\alpha\mathfrak{a} = \mathfrak{a}'$ then shows that $N\alpha = +1$ (of course this does not imply that α is a unit – in general, α will not even be an algebraic integer).

Now we invoke

Lemma 5.8 (Hilbert's Satz 90)**.** *Let $K = \mathbb{Q}(\sqrt{m})$ be a quadratic number field, and assume that $N\alpha = +1$ for some $\alpha \in K^\times$. Then there is a $\beta \in K^\times$ such that $\alpha = \beta/\beta'$.*

Proof. If $\alpha = -1$, take $\beta = \sqrt{m}$. If $\alpha \neq -1$, put $\beta = \frac{\alpha}{\alpha+1}$; then $\frac{\beta}{\beta'} = \frac{\alpha(\alpha'+1)}{(\alpha+1)\alpha'} = \frac{\alpha(\alpha'+1)}{1+\alpha'} = \alpha$. □

Hilbert's Satz 90 provides us with some $\beta \in K$ such that $\alpha = \beta/\beta'$; this shows that $\beta\mathfrak{a} = \beta'\mathfrak{a}'$. In other words: the ideal $\mathfrak{b} = \beta\mathfrak{a}$ has the property that $\mathfrak{b} = \mathfrak{b}'$ (such ideals are called ambiguous). Ambiguous ideals have a very special form:

Lemma 5.9. *Let $\mathfrak{b}$ be an ambiguous ideal in a quadratic number field. Then $\mathfrak{b} = (b)\mathfrak{d}$ for some integer $b \in \mathbb{N}$ and some ideal $\mathfrak{d}$ whose prime ideal factorization only contains distinct ramified prime ideals.*

Proof. It is clear that such ideals are ambiguous: clearly the nontrivial automorpism σ fixes $b \in \mathbb{Z}$ and therefore (b); moreover, all ramified prime ideals $\mathfrak{p}$ have the property that $\mathfrak{p} = \mathfrak{p}^\sigma$.

Assume now that $\mathfrak{b} = \mathfrak{b}'$, and let b be the maximal natural number dividing $\mathfrak{b}$. Then $\mathfrak{b} = b\mathfrak{d}$ for some ambiguous ideal $\mathfrak{d}$. We claim that $\mathfrak{d}$ is not divisibly by split or inert primes, and this will prove our claim.

Clearly $\mathfrak{d}$ is not divisible by inert primes, because these are generated by integers $p \in \mathbb{N}$, contradicting the choice of b. Assume therefore that $(p) = \mathfrak{p}\mathfrak{p}'$ splits and that $\mathfrak{p} \mid \mathfrak{d}$. Then $\mathfrak{p}' \mid \mathfrak{d}' = \mathfrak{d}$, hence $(p) = \mathfrak{p}\mathfrak{p}'$ divides $\mathfrak{d}$, and again this contradicts our choice of b. □

In the case at hand, there is only one ramified prime ideal, namely $(\sqrt{p^*}\,)$, which happens to be principal. Thus all ambiguous ideals are principal, and in particular we conclude that $\mathfrak{a} \sim \mathfrak{b} \sim 1$. Thus Prop. 5.7 is proved.

Proof of the Quadratic Reciprocity Law

The basic idea behind the following proof of the quadratic reciprocity law goes back to Kummer. Since we already know that $(\frac{p}{q}) = -(\frac{q}{p})$ for primes $p \equiv q \equiv 3 \bmod 4$, we may assume that p or q is $\equiv 1 \bmod 4$.

Let us start with the first supplementary law:

a) $\left(\frac{-1}{p}\right) = +1 \iff p \equiv 1 \bmod 4$.
If $p \equiv 1 \bmod 4$, then $k = \mathbb{Q}(\sqrt{p}\,)$ has a unit ε with $N\varepsilon = -1$. Writing $\varepsilon = \frac{1}{2}(x + y\sqrt{p}\,)$, we get $x^2 - py^2 = -4$, and this implies $\left(\frac{-1}{p}\right) = +1$. Now assume that $\left(\frac{-1}{p}\right) = +1$; then p splits in the Euclidean field $\mathbb{Q}(\sqrt{-1}\,)$, which implies $p = a^2 + b^2$. Hence, $p \equiv 1 \bmod 4$.

b) If $p \equiv 1 \bmod 4$, then $\left(\frac{p}{q}\right) = +1 \iff \left(\frac{q}{p}\right) = +1$.
First note that $\left(\frac{p}{q}\right) = +1$ implies that q splits in $k = \mathbb{Q}(\sqrt{p})$, i.e. $q\mathcal{O}_k = \mathfrak{q}\mathfrak{q}'$; from Proposition 5.7 we know that h is odd. Therefore $\mathfrak{q}^h$ is principal, and there exist $x, y \in \mathbb{Z}$ such that $\pm 4q^h = x^2 - py^2$. This yields the congruence $\pm 4q^h \equiv x^2 \bmod p$, and $\left(\frac{-1}{p}\right) = +1$ shows that $\left(\frac{q}{p}\right) = +1$ as claimed.

Now suppose that $\left(\frac{q}{p}\right) = +1$; then $k = \mathbb{Q}(\sqrt{q^*}\,)$ has odd class number, where $q^* = (-1)^{(q-1)/2}q$, and p splits in k. Hence there exist $x, y \in \mathbb{Z}$ such that $\pm 4p^h = x^2 - q^*y^2$, and this implies $\left(\frac{\pm p}{q}\right) = +1$. But since the negative sign can hold only if $q^* \geq 0$, i.e., if $q \equiv 1 \bmod 4$, we get in fact $\left(\frac{p}{q}\right) = +1$.

c) If $p \equiv q \equiv 3 \bmod 4$, then $\left(\frac{p}{q}\right) = +1 \iff \left(\frac{q}{p}\right) = -1$.
We have already proved this.

d) $\left(\frac{2}{p}\right) = +1 \iff p \equiv \pm 1 \bmod 8$.
Put $p^* = (-1)^{(p-1)/2}p$; then $p^* \equiv 1 \bmod 4$, and $k = \mathbb{Q}(\sqrt{p^*}\,)$ has odd class number h. If $p \equiv \pm 1 \bmod 8$, then 2 splits in $k/\mathbb{Q}$, and this implies that $x^2 - p^*y^2 = \pm 4 \cdot 2^h$; we may actually assume that the positive sign holds: if $p \equiv 1 \bmod 4$, the fundamental unit has norm -1, and in case $p \equiv 3 \bmod 4$, we have $x^2 - p^*y^2 > 0$ anyway. Now we get $\left(\frac{2}{p}\right) = +1$.
For the proof of the other direction, assume that $(2/p) = 1$. Then p splits in $\mathbb{Q}(\sqrt{2}\,)$ and we get $\pm p = x^2 - 2y^2 \equiv \pm 1 \bmod 8$, since p is odd.

Exercises

5.1 Show that $K = \mathbb{Q}(\sqrt{-17}\,)$ has class group $\mathrm{Cl}(K) \simeq \mathbb{Z}/4\mathbb{Z}$.

5.2 Show that $K = \mathbb{Q}(\sqrt{-41}\,)$ has class group $\mathrm{Cl}(K) \simeq \mathbb{Z}/8\mathbb{Z}$.

5.3 Show that $K = \mathbb{Q}(\sqrt{-47}\,)$ has class group $\mathrm{Cl}(K) \simeq \mathbb{Z}/5\mathbb{Z}$.

5.4 Show that $K = \mathbb{Q}(\sqrt{-65}\,)$ has class group $\mathrm{Cl}(K) \simeq (2, 4)$.

5.5 Show that $K = \mathbb{Q}(\sqrt{-195}\,)$ has class group $\mathrm{Cl}(K) \simeq (2, 2)$.

5.6 Show that $\mathbb{Q}(\sqrt{79}\,)$ has class number 3.

5.7 Compute the class group and the fundamental unit of $\mathbb{Q}(\sqrt{195}\,)$.

5.8 Find families $\mathbb{Q}(\sqrt{-d}\,)$ of complex quadratic number fields with class numbers divisible by 3 for integers d of the form $d = 3t^2 + 1$ and $d = 3t^2 \pm 8$.

5.9 Consider the diophantine equation $y^2 = x^3 - d$ for squarefree $d \equiv 7 \bmod 8$. Show:

1. If $y^2 = x^3 - d$ has a solution with y even, then $d = 3t^2 - 1$ for some integer $t \equiv 0 \bmod 4$, and the only such solution is $(4t^2 - 1, \pm t(8t^2 - 3))$.
2. If $y^2 = x^3 - d$ has a solution with y odd, then the ideals $(\frac{y+\sqrt{-d}}{2})$ and $(\frac{y-\sqrt{-d}}{2})$ are coprime.
3. Use unique factorization into prime ideals to deduce that $(\frac{y+\sqrt{-d}}{2}) = \mathfrak{p}\mathfrak{b}^3$ and $(\frac{y-\sqrt{-d}}{2}) = \mathfrak{p}'\mathfrak{b}'^3$, where $\mathfrak{p}$ is a prime ideal above 2.
4. Assume first that $\mathfrak{p}$ is principal. Show that this happens if and only if $d = 7$, and solve the equation in this case.
5. Assume that the class number h of $\mathbb{Q}(\sqrt{-d}\,)$ is exactly divisible by 3, i.e., that $3 \mid h$ and $9 \nmid h$. Assume in addition that the ideal class $[\mathfrak{p}]$ has order divisible by 3. Then $y^2 = x^3 - d$ does not have any integral solution with y odd.

6. Show that if a is an odd integer such that $d = 2^{3m+2} - a^2$ is squarefree, then the ideal class $[\mathfrak{p}]$ has order divisible by 3. Now solve $y^2 = x^3 - d$ for $d = 23$ and $d = 31$.

5.10 Now consider the case $d < 0$, i.e. $y^2 - m = x^3$ for $m > 0$. Assume that $m \equiv 5 \bmod 8$, and that the fundamental unit of $\mathbb{Q}(\sqrt{m}\,)$ has the form $\varepsilon = \frac{1}{2}(t + u\sqrt{m}\,)$ for odd integers t, u. Assume finally that the class number of $\mathbb{Q}(\sqrt{m}\,)$ is not divisible by 3. If $y^2 - m = x^3$ has an integral solution, then show:

1. y is even.
2. $(y + \sqrt{m}\,) = (\alpha)^3$ for some $\alpha \in \mathcal{O}_K$.
3. $y + \sqrt{m} = \eta\alpha^3$ for some unit $\eta \in \mathcal{O}_K^\times$.
4. $\alpha^3 \equiv 1 \bmod 2$.
5. $y + \sqrt{m} \equiv 1 \bmod 2$.
6. $\eta \equiv 1 \bmod 2$.
7. Show that if η is a unit $\equiv 1 \bmod 2$, then η is a cube.
8. Deduce that $y + \sqrt{m} = \beta^3$ for some $\beta = \frac{r+s\sqrt{m}}{2}$. Solve the equation.

Observe that $(3, 8)$ is a solution of $y^2 - 37 = x^3$. What does this tell us about the fundamental unit of $\mathbb{Q}(\sqrt{37}\,)$? If you like solving diophantine equations such as the one above, an excellent resource is Mordell's book "Diophantine Equations".

5.11 Find an analog of Theorem 5.5 for the equation $y^2 + d = x^5$.

6. Cryptographic Applications

In this chapter I would like to discuss some applications of algebraic number theory to cryptography. For a better understanding, let me first briefly describe some applications of elementary number theory.

6.1 Protocols based on Elementary Number Theory

RSA

RSA is an acronym for Rivest, Shamir and Adleman, who developed and published this system. For sending secure messages over insecure channels, one first transforms messages into strings of numbers (for example by replacing each letter or symbol with its asci code). From now on, all messages m will be positive integers.

Now assume that Alice wants others to send her messages that cannot be read by anyone who's listening in. She picks two large prime numbers p and q (in practice, these should have about 150 digits each); she computes the product N and chooses an integer $1 < e < (p-1)(q-1)$ coprime to $\phi(N) = (p-1)(q-1)$. Then she publishes the pair (N, e), her "public key" (for example by listing it on her homepage).

Alice	Eve	Bob
picks two large primes p, q computes $N = pq$ picks random $e \in (\mathbb{Z}/\phi(N)\mathbb{Z})^\times$ publishes public key (N, e)	$\xrightarrow{(N,e)}$	
		computes $c \equiv m^e \bmod N$ sends c to Alice
	$\xleftarrow{c}$	
solves $de \equiv 1 \bmod (p-1)(q-1)$ computes $m \equiv c^d \bmod N$		

Fig. 6.1. The RSA protocol

Since the primes p and q have to remain secret, it is important to know how many such primes there are. By the prime number theorem, the number $\pi(x)$ of primes $\leq x$ is about $\pi(x) \sim \frac{x}{\log x}$. Thus the number of primes with 150 digits is approximately equal to $2.6 \cdot 10^{147}$, and hence there are lots of primes for everyone.

Now suppose Bob wants to send a message to Alice. He breaks up this message into little pieces $m < N$ ("little" means that m has less digits than N, i.e., up to 300), and encrypts each m by computing $c \equiv m^e \bmod N$. Then he sends the messages c to Alice.

The point is that even though everybody knows the public key (N, e) and, possibly, the messages c that Bob sent, only Alice can decrypt them. Here's what she does: she applies the Euclidean algorithm to e and $(p-1)(q-1)$ (only Alice knows p and q); since their gcd is 1, she finds a Bezout representation $1 = de + (p-1)(q-1)f$ for integers d and f; then I claim that $m \equiv c^d \bmod N$: thus to decode the message c, Alice only has to raise it to the dth power modulo N. In fact,

$$c^d \equiv (m^e)^d = m^{de} = m^{1-(p-1)(q-1)f} \bmod N.$$

The theorem of Euler-Fermat predicts that $m^{(p-1)(q-1)} \equiv 1 \bmod N$ (except in the unlikely case where m has a factor in common with N – even if you send a gazillion messages, this will not happen within the lifetime of the universe), and the claim follows.

If Alice wants to communicate with Bob, he will also have to pick his own public and secret key. In this case, Alice and Bob can even sign their messages in such a way that Alice can verify whether a message she receives really did come from Bob or from someone else.

Alice	Eve	Bob
Alice and Bob agree upon a hash function h		
picks two large primes p_A, q_A computes $N_A = p_A q_A$ picks random $e_A \in (\mathbb{Z}/\phi(N_A)\mathbb{Z})^\times$ solves $d_A e_A \equiv 1 \bmod \phi(N_A)$ publishes public key (N_A, e_A)	$\xleftrightarrow{(N,e)}$	picks two large primes p_B, q_B computes $N_B = p_B q_B$ picks random $e_B \in (\mathbb{Z}/\phi(N_B)\mathbb{Z})^\times$ solves $d_B e_B \equiv 1 \bmod \phi(N_B)$ publishes public key (N_B, e_B) computes $c \equiv m^{e_A} \bmod N_A$ computes $h \equiv h(m)^{d_B} \bmod N_B$
	$\xleftarrow{(c,h)}$	sends (c, h) to Alice
computes $m \equiv c^{d_A} \bmod N_A$ computes $H \equiv h^{e_B} \bmod N_B$ verifies that $H = h(m)$		

Fig. 6.2. The RSA signature protocol

The basic idea is the following: assume that Alice and Bob have chosen her public keys (N_A, e_A) and (N_B, e_B) as well as her private keys d_A and d_B. Then instead of sending $c \equiv m^{e_A} \bmod N_A$ to Alice, Bob encrypts his messages twice by computing $c' \equiv c^{d_B} \bmod N_B$.

When Alice receives c', she first computes $c \equiv (c')^{e_B} \bmod N_B$ using Bob's public key, and then decrypts c with her own private key. Since only Bob knows d_B, and since this knowledge was necessary for the successful encryption, Alice concludes that the message must have come from Bob.

Unfortunately, this does not work in practice: although Alice can compute $c \equiv (c')^{e_B} \bmod N_B$, the smallest positive integer c representing this residue class is not necessarily the same as the one representing $c \equiv m^{e_A} \bmod N_A$; all we know is that they are congruent modulo N_B. If N_A has 302 digits and N_B only 300, then $m^{e_A} \bmod N_A$ will almost always be represented by an integer $> N_B$. The solution to this problem: do not double encrypt the messages m themselves, but only the values $h(m)$ for a suitably chosen hash function h with values $< \min\{N_A, N_B\}$.

Hash Functions

A hash function adds some sort of "signature" to a message; a very simple hash function would be the map sending a message m to its parity: add up the bits of m modulo 2. In practice, hash functions h are often required to have the following properties:

- a hash function should map a message of arbitrary lenght to a number with a fixed number of bits;
- for any message m, the value $h(m)$ is easy to compute;
- collision resistant: it is difficult to find two messages m, m' with $h(m) = h(m')$;
- preimage resistant: given a hash value y, it is difficult to find some m with $h(m) = y$.

Diffie-Hellman Key Exchange

The RSA method described above solves two problems in cryptography: *confidential message transmission* and *authentication*. A third important problem is *key exchange*: here Alice and Bob have to agree on some large random number known only to them, to be used as a key for encrypting information.

Here's what they do: they pick a large prime number p and a primitive root g modulo p (by definition, the powers of $g \bmod p$ generate the whole group $(\mathbb{Z}/p\mathbb{Z})^\times$; for example, 3 is a primitive root modulo 7, but 2 is not). The numbers p and g are public. Then Alice picks a random number a from $\{0, 1, \ldots, p-2\}$ (of course, the choices $a = 0$ or $a = 1$ will be catastrophic, but this happens with probability $\approx \frac{1}{p}$, that is, never at all), and Bob similarly

Alice	Eve	Bob
Alice and Bob agree upon a prime p and a primitive root g mod p		
Alice picks a random $a \in (\mathbb{Z}/p\mathbb{Z})^\times$		Bob picks a random $b \in (\mathbb{Z}/p\mathbb{Z})^\times$
Alice computes $A \equiv g^a \bmod p$		Bob computes $B \equiv g^b \bmod p$
Alice sends A to Bob	$\xrightarrow{A}$	
	$\xleftarrow{B}$	Bob sends B to Alice
Alice computes $K \equiv B^a \bmod p$		Bob computes $K \equiv A^b \bmod p$

Fig. 6.3. The Diffie-Hellman key exchange

chooses some number b from the same interval; the numbers a and b are kept secret.

Now Alice computes $A \equiv g^a \bmod p$ and sends A to Bob; Bob computes $B \equiv g^b \bmod p$ and sends B to Alice. Then Alice computes $K \equiv B^a \bmod p$, and Bob $K \equiv A^b \bmod p$; the common key then is $K \equiv g^{ab} \bmod p$.

If Eve was eavesdropping, she knows p, g, $A \equiv g^a \bmod p$ and $B \equiv g^b \bmod p$. If she could solve the discrete logarithm problem (DLP), that is, compute the exponent a from the knowledge of p, g and $g^a \bmod p$, then she could easily compute K. As of today, no fast way of solving this problem is known, except in special cases, such as for primes p such that $p-1$ is divisible by small primes only. It is also unknown whether computing $g^{ab} \bmod p$ from g, $g^a \bmod p$ and $g^b \bmod p$ is as hard as solving the DLP, in other words, whether there is a faster method of computing $g^{ab} \bmod p$ from these data than solving the DLP.

ElGamal

There is also an encryption protocol based on DLP, called ElGamal. It works like this: Alice picks p, g, a as before, and computes $A \equiv g^a \bmod p$. Her public key is (p, g, A), her secret key is a. In order to encode messages $m < p$, Bob picks some b as before, computes $B \equiv g^b \bmod p$, and publishes (p, g, B); then he computes $c \equiv A^b m \bmod p$ and sends c to Alice.

At this point, Alice knows B and c, as well as her secret key a. She puts $f = p - 1 - a$ and computes

$$B^f c \equiv g^{b(p-1-a)} A^b m \equiv g^{-ab} g^{ab} m \equiv m \bmod p$$

since $g^{p-1} \equiv 1 \bmod p$.

Shamir's no-key protocol

The following cryptosystem was described in an unpublished mansucript of Shamir; it is also called Massey-Omura, and was first discovered by M.

Alice	Eve	Bob
Alice and Bob agree upon a prime p and a primitive root g mod p		
Alice picks $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ at random, and then computes $A \equiv g^a \bmod p$.		
Alice sends A to Bob	$\xrightarrow{A}$	
		Bob picks $b \in (\mathbb{Z}/p\mathbb{Z})^\times$ at random, computes $B \equiv g^b \bmod p$ and encrypts his message m via $c \equiv mA^b \bmod p$
	$\xleftarrow{B,c}$	Bob sends B and c to Alice
Alice decrypts the message by computing $m \equiv cB^{p-1-a} \bmod p$		

Fig. 6.4. The ElGamal public key cryptosystem

Williamson but not published since he was working at the GCHQ (a British Intelligence service) at the time.

Alice	Eve	Bob
Alice and Bob agree upon a prime p		
Alice picks a pair of integers a, a^{-1} with $aa^{-1} \equiv 1 \bmod p-1$		Bob picks a pair $b, b^{-1} \in (\mathbb{Z}/(p-1)\mathbb{Z})^\times$ with $bb^{-1} \equiv 1 \bmod p-1$.
Alice computes $x \equiv m^a \bmod p$ and sends x to Bob.	$\xrightarrow{x}$	
	$\xleftarrow{y}$	Bob computes $y \equiv x^b \bmod p$ and sends y to Alice
Alice computes $z \equiv y^{a^{-1}} \bmod p$ and sends z to Bob	$\xrightarrow{z}$	
		Bob decrypts the message by computing $m \equiv z^{b^{-1}} \bmod p$

Fig. 6.5. Shamir's no-key protocol

Figure 6.5 explains the protocol, which is clearly based on the difficulty of the discrete log problem.

Goldwasser-Micali probabilistic encryption

In the encryption protocols we have discussed so far, a message m was always encrypted in a unique way. In probabilistic encryption schemes, a message m is in general encrypted in different ways.

A simple example of a probabilistic encryption protocol (though one that does not work well in practice: every bit is encrypted as a number with about 600 bits! Here's how it works:

Alice	Eve	Bob
Alice picks large random primes p, q, puts $n = pq$, and randomly chooses an integer $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ such that $(\frac{a}{p}) = (\frac{a}{q}) = -1$. She publishes her public key (n, a).		
	$\xrightarrow{(n,a)}$	
		To encrypt a bit $b \in \{0, 1\}$, Bob picks a random $r \in (\mathbb{Z}/n\mathbb{Z})^\times$, and computes $c \equiv r^2 a^b \bmod n$.
	$\xleftarrow{c}$	Bob sends c to Alice.
Alice computes $(-1)^b = (\frac{c}{p})$		

Fig. 6.6. Goldwasser-Micali probabilistic encryption

In fact, if $b = 0$, then $(\frac{c}{p}) = (\frac{r^2}{p}) = +1$, and if $b = 1$, then $(\frac{c}{p}) = (\frac{r^2 a}{p}) = -1$.

The security of the Goldwasser-Micali rests on the difficulty of factoring: if Eve manages to factor $n = pq$, she clearly can compute $b = (\frac{c}{p})$.

FLIP

FLIP stands for Fast Legendre Identification Protocol and was suggested recently by Banks, Lieman & Shparlinski. Assume Alice has an RSA-modulus $n = pq$ and wants to convince Bob that she knows p and q (without telling him what p and q are, of course). The protocol I will present in Fig. 6.7 is a radically simplified version of the original one, and very close to the Fiat-Shamir protocol that I will discuss next.

After running through the protocol, Bob will be convinced that Alice knows a prime factor p of her n, because there is no other way of computing the b_i except the one involving p.

The Fiat-Shamir zero-knowledge protocol

Consider the following problem: Alice knows a secret number s (some password, for example) and needs to convince Bob that she knows s without revealing any information about s. This is indeed possible; Fig. 6.8 explains how it works.

Note that $y^2 \equiv (rs^b)^2 \equiv r^2 v^b \equiv x \cdot v^b \bmod n$. Now assume that Malice wants to cheat by pretending she is Alice. She does not know Alice's secret s,

Alice	Eve	Bob
Alice picks large random primes p, q, puts $n = pq$, and randomly chooses an integer a such that $(\frac{a}{p}) = (\frac{a}{q}) = -1$. She then publishes her public key (n, a)	$\xrightarrow{(n,a)}$	
		Bob picks $c_1, \ldots, c_r \in (\mathbb{Z}/n\mathbb{Z})^\times$ and some random vector $(b_1, \ldots, b_r) \in \mathbb{F}_2^r$. He computes $C_i \equiv c_i^2 a^{b_i} \bmod n$.
	$\xleftarrow{C_1, \ldots, C_r}$	Bob sends $C_1, \ldots, C_r$ to Alice.
Alice computes $(C_i/p) = (-1)^{b_i}$ and sends $(b_1, \ldots, b_r)$ to Bob.	$\xrightarrow{(b_1, \ldots, b_r)}$	Bob verifies that Alices vector $(b_1, \ldots, b_r)$ coincides with his.

Fig. 6.7. FLIP

Alice	Eve	Bob
Alice picks two large primes p and q, and computes $n = pq$ and $v \equiv s^2 \bmod n$. She keeps p, q and s secret and publishes n and v.		
Now Alice picks a random $r \in (\mathbb{Z}/n\mathbb{Z})^\times$, computes $x \equiv r^2 \bmod n$, and sends x to Bob.	$\xrightarrow{x}$	
		Bob picks a bit $b \in \{0, 1\}$ at random.
	$\xleftarrow{b}$	
Alice computes $y \equiv r \cdot s^b \bmod n$.		
Alice sends y to Bob.	$\xrightarrow{y}$	
		Bob verifies that $y^2 \equiv x \cdot v^b \bmod n$.

Fig. 6.8. Fiat-Shamir zero-knowledge protocol

nor her primes p and q, but she knows the public $n = pq$ and $v \equiv s^2 \bmod n$. If Malice knew in advance that Bob would send the bit $b = 0$, she could send $x \equiv r^2 \bmod n$, and, after Bob sent back $b = 0$, send $y \equiv r \cdot s^0 \equiv r \bmod n$. If, however, she knew that Bob would send $b = 1$, Malice would send $x \equiv r^2 v^{-b} \bmod n$ and, after Bob sent back $b = 1$, send $y \equiv r \bmod n$ since then $y^2 \equiv x \cdot v^b \bmod n$. This shows that, with probability $p = \frac{1}{2}$, Malice can successfully pretend she knows the secret s. Repeating the procedure with 20 random choices of r and b, however, will reduce the probability of cheating to 2^{-20}, which is less than 10^{-6}.

It can be shown that anyone who can factor n can break this system. In fact we have

Lemma 6.1. *Let p be an odd prime, and assume that $(\frac{a}{p}) = +1$. Then there is an algorithm that solves $x^2 \equiv a \bmod p$ efficiently.*

Proof. We cannot explain here how to make the term "efficiently" precise. In addition, we only give the proof if $p \equiv 3 \bmod 4$ or if $p \equiv 5 \bmod 8$ (the difficulty of the problem grows with the power of 2 dividing $p-1$).

Assume that $p = 4m - 1$ and put $x = a^m$; then $x^2 \equiv a^{2m} = a^{\frac{p-1}{2}} a \equiv (\frac{a}{p})a = a \bmod p$, hence x is a square root of $a \bmod p$.

The case $p \equiv 5 \bmod 8$ was a problem in last semester's first midterm in elementary number theory; it said:

> Assume that $p \equiv 5 \bmod 8$ is prime, and that a is a quadratic residue modulo p.
> 1. Show that if $a^{(p-1)/4} \equiv 1 \bmod p$, then $x = a^{(p+3)/8}$ solves the congruence $x^2 \equiv a \bmod p$.
> 2. If $a^{(p-1)/4} \equiv -1 \bmod p$, then $x \equiv 2a(4a)^{(p-5)/8} \bmod p$ works.

You should have no problems checking this. □

This does not work if $N = p$ is not prime; in fact, if we had an algorithm that could extract square roots modulo N, then we could factor N quickly. The basic idea is this: first observe that if $N = pq$ is the product of two odd primes, and if a is a square modulo N, then there exist four different square roots. In fact, let $b^2 \equiv a \bmod p$ and $c^2 \equiv a \bmod q$; then a square root of $a \bmod N$ can be computed using the Chinese Remainder Theorem as $y \equiv b \bmod p$, $y \equiv c \bmod q$. Replacing b and c by their negatives gives three more roots.

Now pick a random integer s and compute $a \equiv s^2 \bmod N$. Compute one of its square roots r; then $r^2 - s^2 \equiv 0 \bmod N$, and unless $r \equiv \pm s \bmod N$ (this will happen with probability $\frac{1}{2}$), the number $\gcd(r - s, N)$ is a nontrivial factor of N.

Security and Efficiency

If you have a cryptographic protocol based on a difficult mathematical problem (such as factoring or DLP), you can crudely measure the security by demanding that breaking the code with the fastest known algorithms will take approximately 10 years (this is good enough for me and you, since no one will spend 10 years just to read one of our emails, but it might not be good enough for others). This allows you to compare the security of two different systems. The next problem is efficiency: how long does it take to encrypt and decrypt information for systems with the same "security"? By carefully studying the complexity of the algorithms used for en- and decrypting it is

possible to compare different systems. This is of course terribly important if you plan to use one of these systems in practice. In addition to security and efficiency it is also important to estimate how much data you have to store during the encryption.

On the other hand, studying the complexity of algorithms is not really part of number theory, which is why I will be content with presenting only the ideas behind some cryptographic protocols.

6.2 OSS

Already in 1984, Ong, Schnorr & Shamir presented a signature protocol based on quadratic forms. Here's how it works.

Alice	Eve	Bob
Alice picks two large primes p and q, and computes $n = pq$. She picks a random number $u < n$ coprime to n and computes $k \equiv -1/u^2 \bmod n$.		
Alice publishes (n, k)	$(n, k) \longrightarrow$	
For signing a message m, Alice chooses a random $r < n$ and computes $s_1 \equiv \frac{1}{2}(\frac{m}{r} + r) \bmod n$ and $s_2 \equiv \frac{u}{2}(\frac{m}{r} - r) \bmod n$.		
Alice sends (m, s_1, s_2) to Bob.	$(m, s_1, s_2) \longrightarrow$	Bob verifies the congruence $s_1^2 + ks_2^2 \equiv m \bmod n$.

Fig. 6.9. OSS signature protocol

Alice wants to sign her messages; she picks a large composite number n that is difficult to factor (for example the product of two large primes). Next she picks a random number $u < n$ coprime to n and computes $k \equiv -1/u^2 \bmod n$ using the Euclidean algorithm. Alice publishes (n, k) and keeps u secret. For signing a message $m < n$, Alice chooses a random $r < n$ coprime to n and computes

$$s_1 \equiv \tfrac{1}{2}(\tfrac{m}{r} + r) \bmod n, \qquad s_2 \equiv \tfrac{u}{2}(\tfrac{m}{r} - r) \bmod n.$$

Then she sends (m, s_1, s_2) to Bob. Note that with this choice of s_1, s_2, we have

$$\begin{aligned} s_1^2 + ks_2^2 &\equiv \tfrac{r^2}{4}\{(\tfrac{m}{r^2} + 1)^2 + ku^2(\tfrac{m}{r^2} - 1)^2\} \\ &\equiv \tfrac{r^2}{4}\{(\tfrac{m}{r^2} + 1)^2 - (\tfrac{m}{r^2} - 1)^2\} \equiv m \bmod n. \end{aligned}$$

Upon receiving Alice's message, Bob then checks whether $s_1^2 + ks_2^2 \equiv m \bmod n$; if this holds, he concludes that the message was sent by Alice because computing numbers s_1, s_2 with this property without the knowledge of u seems to be extremely difficult.

Alice has to choose one r for each message: should she sign two messages m, m' with the same r, then u can be computed from the signatures: in fact, Eve can compute the differences $\frac{1}{2}(\frac{m}{r} + r) - \frac{1}{2}(\frac{m'}{r} + r) \equiv \frac{m-m'}{2r} \bmod n$ and $\frac{u}{2}(\frac{m}{r} - r) - \frac{u}{2}(\frac{m'}{r} - r) \equiv \frac{m-m'}{2r}\, u$, and then compute u from these.

For forging the signature of an arbitrary message m, Eve has to solve the congruence $s_1^2 + ks_2^2 \equiv m \bmod n$, where k, m and n are given. This was believed to be a difficult problem, but OSS showed that such a solution may be achieved if the class number of $\mathbb{Q}(\sqrt{-k}\,)$ has only small prime factors.

What made Ong, Schnorr & Shamir think that their scheme would be hard to break? They knew that solving $x^2 - Dy^2 \equiv m \bmod n$ was easy if n is prime, or if the factorization of n is known. In fact, write the congruence in the form $x^2 \equiv Dy^2 + m \bmod p$. Then the right hand side, as a polynomial, attains exactly $\frac{p+1}{2}$ values. Since there are only $\frac{p-1}{2}$ nonsquares modulo p, at least one of them must be a square. Actually it can be proved that $Dy^2 + m \bmod p$ is a square with probability $\approx \frac{1}{2}$. Thus we can pick values for y at random, compute the Legendre symbol $(Dy^2 + m/p)$, and then extract the square root as soon as this symbol is $+1$.

Shortly after OSS was presented in 1984, Pollard found a fast way of solving the congruence $s_1^2 + ks_2^2 \equiv m \bmod n$. In the following I will present his solution; we will need the following

Lemma 6.2. *The substitution $u = \frac{x}{y}$, $v = \frac{1}{y}$ transforms the congruences $x^2 - Dy^2 \equiv k \bmod N$ into $u^2 - kv^2 \equiv D \bmod N$.*

This shows that for solving one congruence it is sufficient to solve the other. This was already realized by OSS in their original publication, and it turned out to be the key to breaking their system. The proof is trivial.

Ong, Schnorr and Shamir assumed that solving $x^2 - Dy^2 \equiv m \bmod N$ was as hard as factoring; Pollard showed that they were wrong. Here's what he did:

1. Find a prime $p \equiv m \bmod N$ with $(D/p) = +1$. The fact that this can be done quickly uses some analytic number theory. Actually, I might say a few things about the proof that there are infinitely many primes $p \equiv m \bmod N$ whenever $\gcd(m, N) = 1$ when we come across Dirichlet L-series.
2. Solve $x^2 \equiv D \bmod p$. We have shown that this can be done efficiently (at least if $p \equiv 1 \bmod 4$; but demanding in addition that $p \equiv 3 \bmod 4$ in step (1) does not make the running time much longer).
3. Solve $u^2 - Dv^2 = \lambda p$ for some integer λ with $|\lambda| \le \mu_K$, where μ_K is the Gauss bound for $K = \mathbb{Q}(\sqrt{D}\,)$. This is done as follows: $(p, x + \sqrt{D}\,)$ is a prime ideal above $\mathfrak{p}$ with norm p. Then there is some ideal $\mathfrak{b}$ in the class

$[\mathfrak{p}]^{-1}$ with norm $\lambda := N\mathfrak{b} < \mu_K$. But $\mathfrak{b}\mathfrak{p}$ is principal, say $\mathfrak{b}\mathfrak{p} = (u + v\sqrt{D})$ (actually u and v might be half integers), hence $|u^2 - Dv^2| = N\mathfrak{b}N\mathfrak{p} = p\lambda$.

4. Now assume that we can solve $w^2 - Dz^2 \equiv \lambda \bmod N$. Then we would have $\frac{u^2-Dv^2}{w^2-Dz^2} \equiv p \equiv k \bmod N$; note that the left hand side has the form $r^2 - Ds^2$ since it is the norm of $\frac{u+v\sqrt{D}}{w+z\sqrt{D}} = \frac{(u+v\sqrt{D})(w-z\sqrt{D})}{w^2-Dz^2}$. This shows that we are done once we have solved $w^2 - Dz^2 \equiv \lambda \bmod N$.
   Now we solve this congruence recursively: exchanging the roles of D and λ shows that it is sufficient to solve $w^2 - \lambda z^2 \equiv D \bmod N$. Now go back to step 1, find a prime $p \equiv D \bmod N$ with $(\lambda/p) = +1$, and repeat. Since $\lambda < D$ (in fact $\lambda < \mu_K$), we will be working in fields with smaller and smaller discriminants, and eventually end up with a congruence $x^2 - y^2 \equiv k \bmod N$. But this congruence has the obvious solutions $x \equiv \frac{k+1}{2}$, $y \equiv \frac{k-1}{2} \bmod N$.

After Pollard had broken their first signature protocol, Ong, Schnorr & Shamir (1984) came up with a modified signature scheme, which was then broken by Estes, Adleman, Kompella, McCurley & Miller (1998). Naccache (1994) gave yet another modification of OSS in 1993, but I don't know the verdict on this one.

6.3 Cryptography using Quadratic Number Fields

ElGamal

There is a variant of ElGamal with $(\mathbb{Z}/p\mathbb{Z})^\times$ replaced by class groups. It works as follows: Alice picks a random discriminant Δ (negative, large and squarefree; for example, she could pick $\Delta = -pq$ for two large primes $p \equiv 1$ and $q \equiv 3 \bmod 4$), a random class $C \in \mathrm{Cl}(\Delta)$, the class group of the quadratic number field $\mathbb{Q}(\sqrt{\Delta})$, and some random $a \leq \sqrt{|\Delta|}$. She computes $B = C^a$; her public key then is (Δ, C, B), and her private key is a.

For encrypting some message m, Bob randomly selects a $k \leq \sqrt{|\Delta|}$ and computes $D = C^k$, $E = B^k$, and $c = m \oplus f(E)$ for some preimage and collision resistant hash function f, where $\oplus$ denotes bitwise addition. The cipher text is then (D, c).

For decrypting, Alice computes $E = D^a$ and $m = c \oplus f(E)$.

Eve knows Δ, C, B, D and c, as well as the hash function f. For recovering m she needs to compute $f(E)$. If she could solve the DLP in $\mathrm{Cl}(\Delta)$, she could compute a from $B = C^a$ and then compute m as Alice.

It is clear that for answering questions of security and efficiency of this algorithm one needs to study carefully

- how fast $\mathrm{Cl}(\Delta)$ grows with $|\Delta|$;
- how the factors of the class number are distributed;
- how fast class groups can be computed;

Alice	Eve	Bob
Alice and Bob agree upon a hash function f		
Alice picks two large primes $p \equiv 1 \bmod 4$ and $q \equiv 3 \bmod 4$, and puts $\Delta = -pq$. She picks a random class $C \in \mathrm{Cl}(\Delta)$ and some random $a \leq \sqrt{\lvert\Delta\rvert}$, and computes $B = C^a$. Alices publishes her public key (Δ, C, B)	(Δ, C, B) $\longrightarrow$	
		Bob encrypts m by selecting a random $k \leq \sqrt{\lvert\Delta\rvert}$ and computing $D = C^k$, $E = B^k$, and $c = m \oplus f(E)$. Bob sends (D, c) to Alice
Alice computes $E = D^a$ and $m = c + f(E)$	(D, c) $\longleftarrow$	

Fig. 6.10. El Gamal via Class Groups

- whether the DLP in $\mathrm{Cl}(\Delta)$ can be solved;

as well as dozens of other problems.

The primary motivation for studying number field cryptography is not so much possible application in practice (at least for now, elementary protocols like RSA or elliptic curve cryptography perform better) but rather advancing problems in algebraic number theory: see the cryptography group in Darmstadt (Germany) around Buchmann
`http://www.informatik.tu-darmstadt.de/TI/Forschung/nfc.html`

Another cryptography group with related interests (but slightly slanted towards cryptography using function fields) resides in Calgary (Canada):
`http://cisac.math.ucalgary.ca/`

7. Binary Quadratic Forms

The representation of primes by binary quadratic forms was a topic of central importance for the first number theorists, in particular for Fermat, Euler, Lagrange and Legendre. It was Gauss who melded their isolated results into a coherent theory by studying composition of quadratic forms, introduced the notion of equivalence, and studied the group of equivalence classes of binary quadratic forms, which he called the class group of quadratic forms.

Manjul Bhargava came up with a new way of looking at Gauss composition and discovered some new composition laws; whenever something like this happens, it doesn't take long before traces of the new interpretation are found in classical works. Here it turns out that Bhargava's composition of quadratic forms was hinted at by Gauss, and anticipated by Dedekind, Speiser, Riss, and Shanks, to name a few; Shanks, one of the very few mathematicians after 1950 who was fluent in quadratic forms, also basically discovered the connection to binary cubic forms.

We can only barely scratch the surface of Bhargava's theory of general composition laws that he discussed in his thesis and which were published in four parts in the Annals of Mathematics.

7.1 Groups Acting on Sets

Let G be a group and S a set. We say that G acts on S from the left if there is a map $G \times S \longrightarrow S : (g, s) \longmapsto gs$ such that

1. $(gh)s = g(hs)$
2. $1s = s$

for all $g, h \in G$ and all $s \in S$. The orbit of some $s \in S$ under the action of G is the set $\{gs : g \in G\}$. The stabiliser of $s \in S$ is the subgroup(!) $\{g \in G : gs = s\}$.

Examples.

1. If V is a K-vector space, then the multiplicative group $K^\times$ acts on V via $(r, v) \longmapsto rv$.
2. The groups $\mathrm{GL}_n(K)$ act on the the K-vector spaces K^n in a natural way. The special case $n = 1$ gives us back the first example.

3. The Galois group of a normal extension $K/\mathbb{Q}$ acts on almost everything: it acts on the field K, its multiplicative group $K^\times$, the ring of integers $\mathcal{O}_K$, the unit group $\mathcal{O}_K^\times$, the semigroup of ideals in $\mathcal{O}_K$, and on the class group of K.
4. Every group acts on itself via $g \cdot h = gh$.

The Group $\Gamma = \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$

The group $\mathrm{SL}_2(\mathbb{Z})$ of all 2×2-matrices with integral entries and determinant $+1$ occurs in many areas of mathematics, such as number theory, hyperbolic geometry, or complex analysis. It is well known that this group acts on the upper half plane $\mathcal{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ via $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z = \frac{az+b}{cz+d}$. In fact, a simple calculation shows that $\mathrm{Im}(Mz) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$, so if $\mathrm{Im}(Mz) > 0$ if $\mathrm{Im}\, z > 0$. Another simple calculation shows that $(MN)z = M(Nz)$ for $M, N \in \mathrm{SL}_2(\mathbb{Z})$, so we really do have a group action.

Now observe that $Mz = (-M)z$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$; in other words: $-I$ acts trivially on $\mathcal{H}$. This shows that we actually have $\Gamma = \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$ acting on the upper half plane. The elements of Γ are represented by matrices in $\mathrm{SL}_2(\mathbb{Z})$, with M and $-M$ being regarded as equal.

In the following, the matrices $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ will play a prominent role. T represents a shift by 1 to the right since $T(z) = z + 1$, and S is the composition of a reflection at the unit circle and a reflection at the imaginary axis ($S(z) = -\frac{1}{z}$). It is easily checked that $S^2 = (ST)^3 = I$. Note that although S and ST have finite order, the product $S \cdot ST = T$ has infinite order.

We now define an equivalence relation on $\mathcal{H}$ by setting $z' \sim z$ for $z, z' \in \mathcal{H}$ if $z' = M(z)$ for some $M \in \Gamma$. The following result is fundamental in the theory of modular forms:

Theorem 7.1. *The set*

$$F = \{z \in \mathcal{H} : |z| \geq 1,\ -\tfrac{1}{2} \leq \mathrm{Re}(z) < \tfrac{1}{2}, \quad \textit{and}\ \mathrm{Re}(z) \leq 0\ \textit{if}\ |z| = 1\}$$

is a complete set of representatives for $\mathcal{H}/\sim$, i.e., for every $z \in \mathcal{H}$ there is a unique $z' \in F$ with $z' \sim z$.

Moreover, if $gz = z$ for some $z \in D$ and $g \in \Gamma$, then

$$\begin{cases} z = i & \textit{and}\ g = S; \\ z = \rho & \textit{and}\ g = ST\ \textit{or}\ g = (ST)^2. \end{cases}$$

The first picture in Fig. 7.1 displays the fundamental domain F; the second one shows how F is transformed under small powers of T and S.

Proof. Let G be the subgroup of Γ generated by S and T, and consider some $z \in \mathcal{H}$.

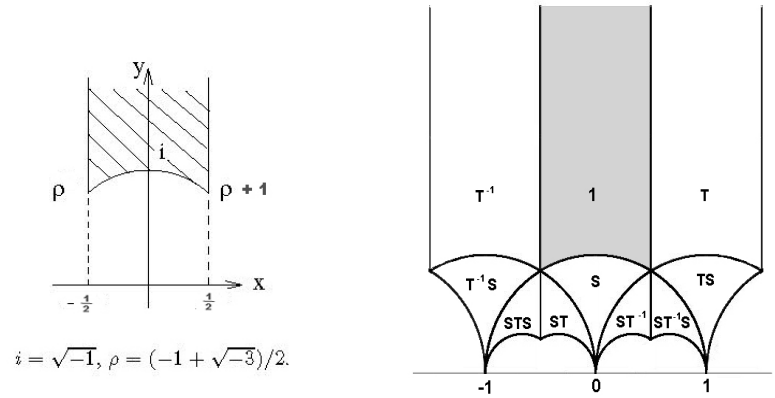


Fig. 7.1. Fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$

1. There exists an element $g \in G$ with $gz \in F$.
   Recall that for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ we have $\operatorname{Im} gz = \frac{\operatorname{Im} z}{|cz+d|^2}$; since c and d are integers, there exist only finitely many pairs (c, d) such that $|cz + d|$ is smaller than some bound C. In fact, $\operatorname{Im}(cz + d) = \operatorname{Im}(cz) = |c| \operatorname{Im} z$, and there are only finitely many integers c for which this is bounded by C. For each of these finitely many values of c, the real part of $cz + d$ is bounded by C, and now the claim follows.
   A different way of seeing this is by observing that the elements of the form $cz + d$ form a lattice $\mathbb{Z} \oplus z\mathbb{Z}$ in $\mathbb{C}$, and for each $C > 0$ there are only finitely many lattice points inside the circle of radius C around the origin.
   This implies that there exists some $g \in G$ such that $\frac{\operatorname{Im} z}{|cz+d|^2}$ is maximal. Next, there is an integer n such that $-\frac{1}{2} \leq \operatorname{Re} T^n gz < \frac{1}{2}$. Note that $z' = T^n gz$ satisfies $\operatorname{Im} z' = \operatorname{Im} gz$ since T is just a horizontal translation. We claim that $|z'| \geq 1$. But if we had $|z'| < 1$, then $Sz' = -\frac{1}{z'}$ would have imaginary part strictly larger than $\operatorname{Im} z'$, which contradicts the choice of g.
   Thus $z' \in F$ except when $|z'| = 1$ and $\operatorname{Re} z > 0$. But in this case, $Sz' \in F$.
2. If $z, gz \in F$ for some $g \in G$, then $g = 1$ except when
   a) $z = i$ and $g = S$;
   b) $z = 1 + \rho$ and $g = ST$ or $g = (ST)^2$.
   In fact, assume we have $z \in F$ and $gz \in F$ for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Replacing (z, g) by (gz, g^{-1}) if necessary we may assume without loss of generality that $\operatorname{Im} gz \geq \operatorname{Im} z$. But this implies $|cz + d| \leq 1$. Now $z \in F$ implies $\operatorname{Im} z \geq \frac{1}{2}\sqrt{3}$, hence $|cz+d| > 1$ whenever $|c| \geq 2$. This shows that $c \in \{-1, 0, 1\}$.

Assume first that $c = 0$. Then $d = \pm 1$, and g is a translation by ± 1. But it is not possible to have both z and $z \pm 1$ in F.
Thus $c = \pm 1$. Since g and $-g$ represent the same element in Γ, we may assume that $c = 1$. Then $|z + d| \leq 1$ implies $d \in \{0, \pm 1\}$ (look at the real parts).

a) If $d = 1$, then $|z+1| \leq 1$ and $|z| \geq 1$ imply $z = \rho = \frac{-1+\sqrt{-3}}{2}$: in fact, we find $1 \geq |1+z|^2 = (1+z)(1+\overline{z}) = 1 + z + \overline{z} + z\overline{z} \geq 2 + z + \overline{z}$, hence $2 \operatorname{Re} z \leq -1$. But this is only possible if $\operatorname{Re} z = -\frac{1}{2}$. But then $|z| = 1$, hence $z = \rho$ as claimed.
Now $g\rho = \frac{a\rho+b}{\rho+1} = a + (a-b)\rho$. Clearly $g\rho \in F$ if and only if $a - b = 1$ and $a = 0$, i.e., if and only if $g = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right) = ST$.
b) If $d = -1$, then $|z - 1| \leq 1$ and $|z| \geq 1$ imply $\operatorname{Re} z = \frac{1}{2}$ as above, hence $z = \rho + 1$. But $\rho + 1 \notin F$.
c) If $d = 0$, then $ad - bc = 1$ implies $b = -1$, hence $gz = \frac{az+b}{z} = a - \frac{1}{z}$. From $|z| \leq 1$ and $z \in F$ we get $|z| = 1$; thus z and $\frac{1}{z}$ lie on the unit circle, and we conclude that $-\frac{1}{2} < \operatorname{Re} \frac{1}{z} \leq \frac{1}{2}$. But then $gz \in F$ shows $a \in \{-1, 0, 1\}$.
If $a = 0$, then z and $-\frac{1}{z}$ lie on the unit circle and in F. With $z = s + it \in F$ this shows that $-\frac{1}{z} = -s + it \in F$. A look at the definition of F will convince you that this is only possible if $s = 0$ and $t = 1$, i.e., if $z = i$; in this case, $g = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) = S$ fixes z.
If $a = 1$, then z and $1 - \frac{1}{z}$ lie in F. There is, hoever, no point $w = -\frac{1}{z}$ on the unit circle for which $1 + w \in F$.
Finally, if $a = -1$, then z and $gz = \frac{-z-1}{z} = -1 - \frac{1}{z}$, then as above we find that the only possibility is $z = \rho$, which gives $gz = \rho$. Note that $g = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right) = (ST)^2$.

□

Theorem 7.2. *The group Γ is generated by S and T.*

If you know a little group theory, then this result can be stated in the form $\Gamma = \langle S, T | S^2 = (ST)^3 = 1 \rangle$; in more fancy language, this means that Γ is the free product of $\langle S \rangle \simeq \mathbb{Z}/2\mathbb{Z}$ and $\langle ST \rangle \simeq \mathbb{Z}/3\mathbb{Z}$.

Proof. Let γ be an element of Γ and put $w = 2i \in F$. Let G be the group generated by S and T. In the proof of Theorem 7.1 we have seen that there is some $g \in G$ such that $g(\gamma z) \in F$. On the other hand, w and $g\gamma w$ are both in F, and $w \neq i, \rho$: this implies $g\gamma = 1$, that is, $\gamma = g^{-1} \in G$. □

The Action of $\mathbf{SL_2}(\mathbb{Z})$ on binary quadratic forms

We are interested in the set of integral binary quadratic forms $Q = (A, B, C) = Ax^2 + Bxy + Cy^2$. A typical problem is: which integers n are represented by a given quadratic form Q, that is, can be written in the form $n = Q(x, y)$ for integers x, y?

Examples: we have already seen that odd primes p are represented by $Q(x,y) = x^2 + y^2$ if and only if $p \equiv 1 \bmod 4$, or by $Q(x,y) = x^2 - 2y^2$ if and only if $p \equiv \pm 1 \bmod 8$.

This question is clearly related to questions we have studied before: the equation $n = Ax^2 + Bxy + Cy^2$ can also be written in the form $4An = (2Ax+By)^2 - (B^2-4AC)y^2$, and whether this equation is solvable in integers depends on whether there is an ideal of norm $4An$ in the ring of integers of the quadratic number field $\mathbb{Q}(\sqrt{\Delta}\,)$, where $\Delta = B^2 - 4AC$ is the **discriminant** of Q. Note that we will only consider quadratic forms whose discriminant is not a square.

In this chapter we will describe a more direct method of attacking this question and present the beginnings of the theory of binary quadratic forms, which goes back to Gauss.

The group $\mathrm{SL}_2(\mathbb{Z})$ acts on the set of binary quadratic forms in the following way: given a quadratic form $Q = (A,B,C)$ and a matrix $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ we define the quadratic form $Q' = Q|_M = (A',B',C')$ by $Q|_M(x,y) = Q(rx+sy, tx+uy)$. A simple calculation shows that

$$\begin{aligned} A' &= Ar^2 + Brt + Ct^2, \\ B' &= 2(Ars + Ctu) + B(ru+st), \\ C' &= As^2 + Bsu + Cu^2. \end{aligned}$$

It is easily verified that $\Delta' = B'^2 - 4A'C' = (ru-st)^2\Delta$, and now the fact that $M \in \mathrm{SL}_2(\mathbb{Z})$ implies that $\Delta' = \Delta$.

As an example, consider the form $Q(x,y) = x^2 + y^2$ of discriminant $\Delta = -4$, and the matrix $M = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Then $Q|_M(x,y) = (2x+y)^2 + (x+y)^2 = 5x^2 + 6xy + 2y^2$, and $Q|_M$ has discriminant $6^2 - 4 \cdot 5 \cdot 2 = -4$.

We can bring in some more linear algebra in the following way. To every binary quadratic form (A,B,C) we associate the matrix $P = \left(\begin{smallmatrix} A & B/2 \\ B/2 & C \end{smallmatrix}\right)$ (the occurrence of half-integers made Gauss look only at binary quadratic forms whose middle coefficient B is even); then a simple calculation shows that $Ax^2 + Bxy + Cy^2 = (x,y)P\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Moreover we find that $\operatorname{disc} Q = B^2 - 4AC = -4 \det P$. Another simple calculation shows that if the matrix P corresponds to the quadratic form Q, then the matrix corresponding to $Q|_M$ is M^tPM. This again shows that

$$\operatorname{disc} Q|_M = -4 \det M^tPM = -4 \det P (\det M)^2 = (\det M)^2 \operatorname{disc} Q.$$

Moreover we see that $Q|_{MN}$ corresponds to $(MN)^tP(MN) = N^tM^tPMN$, hence $Q|_{MN} = (Q|_M)|_N$: this means that $\mathrm{SL}_2(\mathbb{Z})$ acts on quadratic forms from the right.

Thus $\mathrm{SL}_2(\mathbb{Z})$ acts on the set of binary quadratic forms of discriminant Δ. In order to become familiar with this action, we now prove

Lemma 7.3. *If Q represents n, then so does $Q|_M$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$.*

Proof. Assume that $n = Q(x, y)$ for integers x, y. If P is the associated matrix, then $n = (x, y)P\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $Q|_M$ is associated to M^tPM, we have $n = Q|_M(u, v) = (u, v)M^tPM\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ for the vector $\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = M^{-1}\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $M \in \mathrm{SL}_2(\mathbb{Z})$, we have $M^{-1} \in \mathrm{SL}_2(\mathbb{Z})$ as well, and this means that u and v are integers. $\square$

Thus Q and $Q|_M$ represent exactly the same numbers.

Example. $Q(x, y) = x^2 + y^2$ represents $5 = 2^2 + 1^2$. With $M = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ we have found $Q|_M(x, y) = 5x^2 + 6xy + 2y^2$. Now $M^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 2 \end{smallmatrix}\right)$, hence $M^{-1}\left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, and indeed we have $Q|_M(1, 0) = 5$.

We say that Q represents n properly if $n = Q(x, y)$ for coprime integers x, y. Thus $4 = 1^2 + 3 \cdot 1^2$ shows that $x^2 + 3y^2$ represents 4 properly, whereas $4 = 2^2 + 3 \cdot 0^2$ is not a proper representation.

Lemma 7.4. *If Q represents a prime p, then Q is equivalent to a form Q' with p as its first coefficient.*

Proof. Assume that $Q = (A, B, C)$ and that $Q(r, t) = p$ for integers r, t. Since $\gcd(r, t)^2 \mid p$, we have $\gcd(r, t) = 1$, so by Bezout there are integers s, t with $ru - st = 1$. Setting $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right)$ we find that $Q|_M = Q'$ with $Q' = (A', B', C')$ and $A' = Ar^2 + Brt + Ct^2 = Q(r, t) = p$. $\square$

7.2 Reduction of Binary Quadratic Forms

From now on, all our forms will be *primitive*, that is, we assume that $\gcd(A, B, C) = 1$ for all our forms (A, B, C).

Now let $Q = (A, B, C)$ be a binary quadratic form with **negative** discriminant $\Delta = B^2 - 4AC$; let us also assume that $A > 0$: in this case, Q is a positive definite quadratic form since $4AQ(x, y) = (2AX + BY)^2 - \Delta Y^2$. To each such form we associate a point

$$z = \frac{-B + i\sqrt{|\Delta|}}{2A} \in \mathcal{H}.$$

Note that from every such point we can get back the form Q: the coefficients A and B can be read off directly, and then C can be determined from $\Delta = B^2 - 4AC$.

Lemma 7.5. *Let Q be a positive definite quadratic form, and let $z \in \mathcal{H}$ its associated point. Then for $M \in \mathrm{SL}_2(\mathbb{Z})$, the quadratic form $Q|_M$ is associated to $M^{-1}z$.*

This is left as an Exercise.

We say that a quadratic form Q with negative discriminant is *reduced* if Q corresponds to a point in F. Since every $z \in \mathcal{H}$ is equivalent to one in

F, we conclude that every quadratic form Q with negative discriminant is equivalent to exactly one reduced form.

Now we have to study when $z = \frac{-B+i\sqrt{|\Delta|}}{2A}$ lies in F. The condition $|z| \geq 1$ is equivalent to $C \geq A$. Next $-\frac{1}{2} \leq \frac{-B}{2A} < \frac{1}{2}$ is equivalent to $-A < B \leq A$. Finally, if $|z| = 1$ we need to make sure that $\operatorname{Re} z \leq 0$; this translates into the condition $B \geq 0$ if $A = C$.

Thus we have seen:

Lemma 7.6. *A binary quadratic form $Q = (A, B, C)$ with negative discriminant is reduced if and only if $-A < B \leq A \leq C$ or $0 \leq B \leq A = C$.*

It is easy to find a bound for the coefficient A of a reduced form:

Lemma 7.7. *If $Q = (A, B, C)$ is a reduced binary quadratic form with negative discriminant Δ, then $|A| \leq \sqrt{-\Delta/3}$.*

Proof. We know $B^2 \leq A^2$ and $A \leq C$, hence $-\Delta = 4AC - B^2 \geq 4A^2 - A^2 = 3A^2$. □

As a corollary we observe that there are only finitely many reduced forms of given discriminant $\Delta < 0$: there are only finitely many A by Lemma 7.7, hence only finitely many B with $|B| \leq A$. Finally, for each pair (A, B) there is at most one C because $\Delta = B^2 - 4AC$ is fixed.

The number of reduced forms of discriminant $\Delta < 0$ is denoted by $h(\Delta)$ and is called the class number. It is quite easy to compute all reduced forms of small discriminant:

Δ	$h(\Delta)$	reduced forms
-3	1	$x^2 + xy + y^2$
-4	1	$x^2 + y^2$
-7	1	$x^2 + xy + 2y^2$
-8	1	$x^2 + 2y^2$
-11	1	$x^2 + xy + 3y^2$
-12	1	$x^2 + 3y^2$
-15	2	$x^2 + xy + 4y^2, 2x^2 + xy + 2y^2$
-16	1	$x^2 + 4y^2$
-19	1	$x^2 + xy + 5y^2$
-20	2	$x^2 + 5y^2, 2x^2 + 2xy + 3y^2$
-23	3	$x^2 + xy + 6y^2, 2x^2 \pm xy + 3y^2$
-24	2	$x^2 + 6x^2, 2x^2 + 3y^2$
-27	1	$x^2 + xy + 7y^2$

If you compare this table with the class numbers you computed for the fields K with discriminants $-3 \geq \Delta \geq -23$, then you will notice that we have exactly $h(\Delta) = h_K$ reduced forms if Δ is a field discriminant; note also that the results for $\Delta = -12, -16, -27$ cannot be explained with our results on class groups of quadratic fields.

This is exactly what we will prove; in fact, the set $\mathrm{Cl}(\Delta)$ of binary quadratic forms with discriminant Δ can be given a natural group structure, and then we have $\mathrm{Cl}(K) \simeq \mathrm{Cl}(\Delta)$ for $K = \mathbb{Q}(\sqrt{\Delta}\,)$, at least if $\Delta < 0$.

Before we continue, let us explicitly compute the class number for $\Delta = -4 \cdot 65$. We know that $|a| \le \sqrt{-\Delta/3} < 10$. Thus we have $-9 \le A < B \le 9 \le C$ and $-\Delta = 260 = 4AC - B^2$. Clearly $B = 2b$ is even, and we have $65 = AC - b^2$. Now we go through the individual cases; the congruence $65 \equiv b^2 \bmod A$ will occasionally help us to save work.

- $A = 1$: since B is even, we have $B = 0$ and therefore $C = 65$. We find the form $(1, 0, 65)$.
- $A = 2$: then $B = 0$ is impossible, so we must have $B = \pm 2$. Now $65 = 2C - 1$ gives $C = 33$, and we get the forms $(2, \pm 2, 33)$.
- $A = 3$: clearly $b \ne 0$; $b = \pm 1$ leads to $C = 22$ and to the form $(3, \pm 2, 22)$.
- $A = 4$: this is again impossible since $65 \equiv -b^2 \bmod 4$ is not solvable.
- $A = 5$: For $B = 0$ we find $(5, 0, 13)$. From $65 = 5C - b^2$ we see that b must be divisible by 5, hence B must be divisible by 10, and this only works for $B = 0$.
- $A = 6$: here we find $b = 1$ and $C = 11$, that is, the forms $(6, \pm 2, 11)$. The cases $b = 2$ and $b = 3$ lead to contradictions.
- $A = 7$: this is impossible since $(\frac{-65}{7}) = -1$.
- $A = 8$: this contradicts $65 \equiv -b^2 \bmod 4$.
- $A = 9$: here we check that $65 = 9C - b^2$ is not solvable for integers b with $|b| \le 4$.

Thus the set of reduced forms of discriminant $-4 \cdot 65$ is

$$\{(1, 0, 65), (2, \pm 2, 33), (3, \pm 2, 22), (5, 0, 13), (6, \pm 2, 11)\}.$$

The Reduction Algorithm for Definite Forms

Given a quadratic form (A, B, C) with negative discriminant, how can we find an equivalent reduced form? The algorithm below is a consequence of several simple observations.

First, for $M = \left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$ we find $Q|_M(x, y) = A(x + by)^2 + B(x + by)y + Cy^2$, hence

$$Q|_M = (A, B + 2Ab, Ab^2 + Bb + C). \tag{7.1}$$

Thus we can use such a transformation to decrease the size of B while keeping A fixed.

Next, for $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ we get $Q|_S(x, y) = A(-y)^2 + B(-xy) + Cx^2$:

$$Q|_S = (C, -B, A). \tag{7.2}$$

Thus S can be used to exchange A and C.

Here's the algorithm:

`input: a primitive quadratic form (A,B,C) with` $\Delta < 0$ `and` $A > 0$.
`output: an equivalent reduced form` (A'', B'', C'').

1. If $|B| > A$, find $b \in \mathbb{Z}$ with $|B + 2Ab| \le A$, and put $(A', B', C') = Q|_M$ for $M = \left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$. Then $|B'| \le A' = A$ (see (7.1)).
2. If $A' \le C'$ goto step 3. If $A' > C'$, use S to replace the form (A', B', C') by $(C', -B', A')$. If $|B'| > C'$, goto step 1.
3. Now we have a quadratic form (A'', B'', C'') with $|B''| \le A'' \le C''$. If $C'' \ge A''$, this form is reduced unless $C'' = A''$ and $B'' < 0$; in this case, replace the form by $(C'', -B'', A'')$.

Note that this algorithm also can compute the matrix M for which $Q' = Q|_M$: all you have to do is keep track of the matrices used in each step and multiply them together.

Here's an example: start with the form $(3, 9, 7)$ with discriminant $\Delta = 8^2 - 4 \cdot 3 \cdot 7 = -3$. From $|9 + 6b| \le 3$ we find that we may take $b = -1$ or $b = -2$. With $b = -2$ we get $(A', B', C') = (3, -3, 1)$. Since $3 > 1$, we switch and get $(1, 3, 3)$. Now we repeat step 1: we find $|3 + 2b| \le 1$ for $b = -1$, and get $(1, 1, 1)$. Thus $(3, 9, 7) \sim (1, 1, 1)$, and this form is reduced.

Representations by Quadratic Forms

Let me also briefly explain how to prove that primes $p \equiv 1 \bmod 4$ can be written in the form $p = x^2 + y^2$. We have already seen how to do this using ideals; now we will give (the same) proof using the language of quadratic forms.

We start with

Lemma 7.8. *If Δ is a nonsquare discriminant and $(\frac{\Delta}{p}) = +1$, then there is a quadratic form $Q = (p, B, C)$ with discriminant Δ.*

Proof. Assume for simplicity that p is odd (the case $p = 2$ is left as an exercise). Write $\Delta \equiv B^2 \bmod p$. Since p is odd, we may assume that Δ and B have the same parity (otherwise replace B by $p - B$); then $\Delta \equiv B^2 \bmod 4p$, hence there is an integer C such that $\Delta = B^2 - 4pC$. But then $Q = (p, B, C)$ has the desired properties. □

In the ideal language, this means that $(p) = \mathfrak{p}\mathfrak{p}'$ since forms with first coefficients p get mapped to ideals of norm p under i.

Now observe that $Q = (p, A, C)$ represents p since $p = Q(1, 0)$. Since Q and $Q|_M$ represent the same numbers, p is also represented by $Q|_M$. In particular, p is represented by some reduced form of discriminant Δ.

Corollary 7.9. *If $(\frac{\Delta}{p}) = +1$ for some prime p, then p is represented by some reduced form of discriminant Δ.*

If $\Delta = -4$, there is only one reduced form, and we conclude that primes $p \equiv 1 \bmod 4$ have the form $p = x^2 + y^2$.

If $\Delta = -3$, the only reduced form is $x^2 + xy + y^2$, hence every prime $p \equiv 1 \bmod 3$ has the form $p = x^2 + xy + y^2$.

If $\Delta = -20$, there are two reduced forms, namely $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$. Every prime $p \equiv 1, 3, 7, 9 \bmod 20$ is represented by one of these forms.

Summary.

We have seen that $\mathrm{SL}_2(\mathbb{Z})$ acts on quadratic forms with discriminant Δ; equivalent forms Q and $Q|_M$ represent the same primes. If $\Delta < 0$, there are only finitely many reduced forms.

In the following, we will show that equivalence classes of quadratic forms of discriminant Δ can be composed; for $\Delta < 0$, this gives the set $\mathrm{Cl}(\Delta)$ of reduced forms of discriminant Δ a group structure, and we will show that $\mathrm{Cl}(\Delta) = \mathrm{Cl}(K)$, where $K = \mathbb{Q}(\sqrt{\Delta}\,)$.

7.3 The Class Number

The bijection is easy to define (showing that the bijection is an isomorphism, that is, respects the group laws, will require at lot more effort). We will define maps $i : \mathrm{Cl}(\Delta) \longrightarrow \mathrm{Cl}(K)$ and $f : \mathrm{Cl}(K) \longrightarrow \mathrm{Cl}(\Delta)$ and then show that $f \circ i$ and $i \circ f$ are the identity maps on $\mathrm{Cl}(\Delta)$ and $\mathrm{Cl}(K)$, respectively.

The map i sending quadratic forms $Q = (A, B, C)$ of discriminant $\Delta = B^2 - 4AC$ to integral ideals $\mathfrak{a}_Q$ in $\mathcal{O}_K$ is easily defined: we just put

$$i(Q) = [A, \tfrac{-B+\sqrt{\Delta}}{2}]. \tag{7.3}$$

This is an ideal in $\mathcal{O}_K$ since $A \mid N\big(\frac{-B+\sqrt{\Delta}}{2}\big) = AC$.

Example 1. The principal form

$$Q(x, y) = \begin{cases} x^2 - my^2 & \text{if } \Delta = 4m, \\ x^2 + xy - my^2 & \text{if } \Delta = 4m + 1 \end{cases}$$

has image $i(Q) = (1)$.

Example 2. The two reduced forms $(1, 0, 5)$ and $(2, 2, 3)$ of discriminant -20 get mapped to (1) and $(2, -1 + \sqrt{-5}\,)$, respectively.

We now want to show that i is induces a map on the classes, i.e., that $i(Q|_M) \sim i(Q)$ for $M \in \mathrm{SL}_2(\mathbb{Z})$. To this end, observe that $i(Q) = [A, \frac{-B+\sqrt{\Delta}}{2}]$ for $Q = (A, B, C)$. Now we write $i(Q) = A[1, \tau]$, where $\tau = \frac{-B+\sqrt{\Delta}}{2A}$ is the point in the upper half plane associated to Q. But with $Q' = (A', B', C') = Q|_M$ we now find $i(Q|_M) = A'[1, \tau'] = A'[1, M^{-1}\tau]$. It remains to show that the two ideals $A[1, \tau]$ and $A'[1, M^{-1}\tau]$ are equivalent.

Now let $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$; then $\{\tau, 1\}$ is the basis of a $\mathbb{Z}$-module $[1, \tau]$ in K if and only if $\{r\tau + s, t\tau + u\}$ is. But

$$[t\tau+u, r\tau+s] = (t\tau+u)[1, \tfrac{r\tau+s}{t\tau+u}] \sim [1, M\tau],$$

hence $[1,\tau] = (t\tau+u)[1, M\tau]$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$. Since $M^{-1} = \left(\begin{smallmatrix} u & -s \\ -t & r \end{smallmatrix}\right)$, we see that $[1,\tau] = (-t\tau+r)[1, M^{-1}\tau]$ and therefore

$$A'i(Q) = AA'[1,\tau] = AA'(-t\tau+r)[1, M^{-1}\tau] = A(-t\tau+r)i(Q|_M).$$

Thus equivalent quadratic forms correspond to equivalent ideals, and we have proved:

Proposition 7.10. *The map i defined in (7.3) satisfies $i(Q|_M) \sim i(Q)$, and therefore maps classes of forms to ideal classes.*

Now let us define the inverse map: given an ideal $\mathfrak{a}$ in the ring of integers $\mathcal{O}_K$ of the complex quadratic number field K with discriminant $\Delta = \operatorname{disc} K$, we write $\mathfrak{a} = [\alpha, \beta]$ and put

$$f(\mathfrak{a}) = Q_{\mathfrak{a}}(x,y) = \frac{N(\alpha x - \beta y)}{N(\mathfrak{a})}. \tag{7.4}$$

For this to make sense we must prove that $Q_{\mathfrak{a}}$ has integral coefficients and discriminant Δ. In fact, we have

$$\begin{aligned} N(\alpha x+\beta y) &= (\alpha x - \beta y)(\alpha' x - \beta' y) \\ &= \alpha\alpha' x^2 - (\alpha\beta' + \alpha'\beta)xy + \beta\beta' y^2 \\ &= Ax^2 + Bxy + Cy^2. \end{aligned}$$

Clearly, A, B and C are integers, since they are norms and traces of elements in $\mathcal{O}_K$. Moreover, $\alpha\alpha' \in \mathfrak{a}\mathfrak{a}' = (N\mathfrak{a})$, hence A is divisible by $N\mathfrak{a}$. But for the very same reason we have $B, C \in \mathfrak{a}\mathfrak{a}'$, hence $Q_{\mathfrak{a}} = \frac{N(\alpha x - \beta y)}{N\mathfrak{a}}$ also has integral coefficients.

Next, the discriminant of $Q_{\mathfrak{a}}$ is $\operatorname{disc} Q_{\mathfrak{a}} = \frac{B^2-4AC}{N\mathfrak{a}^2} = \frac{(\alpha\beta'-\alpha'\beta)^2)}{N\mathfrak{a}^2} = \Delta$. The last equality follows from the observation that an ideal $\mathfrak{a} = [\alpha,\beta]$ satisfies (Exercise!)

$$\begin{vmatrix} \alpha & \alpha' \\ \beta & \beta' \end{vmatrix} = \Delta \cdot N\mathfrak{a}^2.$$

Finally, the equivalence class of $f(\mathfrak{a})$ does not depend on the choice of the basis of $\mathfrak{a}$: in fact, let $\{\gamma,\delta\}$ denote another basis of $\mathfrak{a}$; then $\left(\begin{smallmatrix}\gamma\\ \delta\end{smallmatrix}\right) = M\left(\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right)$ for some $M \in \mathrm{SL}_2(\mathbb{Z})$. Since the norm of $\mathfrak{a}$ does not depend on the basis, we only have to study the effects of M on $N(\alpha\beta' - \alpha'\beta)$. Now $\alpha\beta' - \alpha'\beta = \det\left(\begin{smallmatrix}\alpha & \beta \\ \alpha' & \beta'\end{smallmatrix}\right)$; but then

$$\gamma\delta' - \gamma'\delta = \det\left(\begin{smallmatrix}\gamma & \delta \\ \gamma' & \delta'\end{smallmatrix}\right) = \det M\left(\begin{smallmatrix}\alpha & \beta \\ \alpha' & \beta'\end{smallmatrix}\right) = \alpha\beta' - \alpha'\beta.$$

Moreover, $f(\gamma\mathfrak{a}) = \frac{N\gamma \cdot N(\alpha x - \beta y)}{N(\gamma\mathfrak{a})} = \frac{N\gamma}{N(\gamma)} f(\mathfrak{a}) = f(\mathfrak{a})$ because $N\gamma = N(\gamma)$ for complex quadratic fields (note that the definition of f involves the norm

of an element in the numerator and the norm of an ideal in the denominator; if $\Delta < 0$, then $N\gamma > 0$; if $\Delta > 0$, then it can happen that $N(\gamma) = -N\gamma$, and this is exactly the reason why there sometimes are more equivalence classes of forms than ideal classes in this case).

This shows

Proposition 7.11. *The map f defined in (7.4) maps ideal classes to equivalence classes of quadratic forms.*

Now we can state

Theorem 7.12. *Consider the ideal class group $\mathrm{Cl}(K)$ of $K = \mathbb{Q}(\sqrt{\Delta})$, where $\Delta < 0$ is a field discriminant, and the class group $\mathrm{Cl}(\Delta)$ of primitive quadratic forms of discriminant Δ. Then the maps $i : \mathrm{Cl}(\Delta) \longrightarrow \mathrm{Cl}(K)$ and $f : \mathrm{Cl}(K) \longrightarrow \mathrm{Cl}(\Delta)$ are inverses of each other. In particular, the number of reduced forms of discriminant Δ is just the class number of K.*

Proof. Write $\mathfrak{a} = [A, \frac{1}{2}(-B + \sqrt{\Delta})]$. Then $Q = f(\mathfrak{a}) = (A, B, C)$ since $\frac{AA'}{N\mathfrak{a}} = \frac{A^2}{A} = A$ and $-\frac{1}{N\mathfrak{a}}(\alpha\beta' + \alpha'\beta) = \frac{1}{2}(B + \sqrt{\Delta}) + \frac{1}{2}(-B + \sqrt{\Delta}) = B$. But then $i(f(\mathfrak{a})) = i(Q) = \mathfrak{a}$.

Showing that $f \circ i$ is the identity map is left as an exercise. □

Consider e.g. the three reduced forms of discriminant $\Delta = -23$; they are $Q_0 = x^2 + xy + 6y^2$, $Q_2 = 2x^2 + xy + 3y^2$ and $Q_3 = 2x^2 + xy + 3y^2$. These correspond to the ideals $I_1 = [1, \omega] = (1)$, $I_2 = (2, \omega)$ and $I_3 = (2, -1 + \omega) = I_2'$, where $\omega = \frac{-1+\sqrt{-23}}{2}$. The fact that $[Q_2] + [Q_3] = [Q_1]$ corresponds to the ideal equation $I_2 I_3 = (2) \sim (1)$.

7.4 Gauss Composition

Fermat observed that if n is an integer that is properly represented by a quadratic form such as $x^2 + y^2$ or $x^2 + 2y^2$, and if p is a divisor of n, then p is also represented by this form. For example, $65 = 1^2 + 8^2$ implies that 5 and 13 are the sum of two squares.

Fermat also realized that this fails for the form $Q(x, y) = x^2 + 5y^2$: here $21 = 3 \cdot 7 = 1^2 + 5 \cdot 2^2$, but neither 3 nor 7 are represented. Fermat, however, conjectured that primes $p \neq 2, 5$ are represented by q if and only if $p \equiv 1, 9 \bmod 20$, and that the products of two primes $p \equiv 3, 7 \bmod 20$ are also represented.

Euler saw that primes $p \equiv 3, 7 \bmod 20$ have the property that $2p$ is represented by q; this implies Fermat's claim. In fact, assume that $2p = x^2 + 5y^2$. Then x and y are odd, and we may put $x = 2u + v$ and $y = v$ for integers u, v; this gives us $2p = (2u + v)^2 + 5v^2 = 4u^2 + 4uv + 6v^2$, or $p = 2u^2 + 2uv + 3v^2$; conversely, if $p = 2u^2 + 2uv + 3v^2$, then $2p = (2u + v)^2 + 5v^2$.

Thus the primes $p \equiv 3, 7 \bmod 20$ are not represented by the principal form $(1, 0, 5)$ of discriminant 20, but by the form $(2, 2, 3)$ of discriminant 20.

Now assume that $p = 2u^2 + 2uv + 3v^2$ and $q = 2s^2 + 2st + 3t^2$. Then Lagrange realized that

$$\begin{aligned} pq &= (2u^2 + 2uv + 3v^2)(2s^2 + 2st + 3t^2) \\ &= (2us + ut + vs + 3vt)^2 + 5(tu - sv)^2 \end{aligned}$$

Thus if $2p$ and $2q$ are represented by $x^2 + 5y^2$, then so is pq.

This observation was generalized by Legendre, and finally Gauss defined a composition of classes of forms that made the equivalence classes of forms into an abelian group (he did not know the concept of a group, and his statement of associativity is quite complicated, to say the least).

In the special case where the discriminant Δ is fundamental (i.e., the discriminant of a quadratic number field), Gauss said that a form Q_3 is the composition of forms Q_1, Q_2 with discriminant Δ if

$$Q_1(u, v)Q_2(s, t) = Q_3(x, y), \tag{7.5}$$

where x and y are $\mathbb{Z}$-linear combinations of us, ut, vs and vt. In an identity as (7.5) above, we can always replace Q_3 by an equivalent form (all we have to do is substitue $x = rX + sY$, $y = tX + uY$ for $\left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$), and this means that composition is only defined up to equivalence. In other words, coposition is defined on classes, not on forms. In the situation of (7.5), we write $[Q_1] \oplus [Q_2] = [Q_3]$. We will eventually show that this defines a group law on the set of equivalence classes of primitive binary quadratic forms of discriminant Δ.

As a more general example, consider the quadratic forms

$$Q_1 = (1, 0, fg), \quad Q_2 = (g, 0, f), \quad Q_3 = (f, 0, g)$$

of discriminant $\Delta = -4fg$. As already Euler, Lagrange and Legendre have observed, there is a connection between these forms:

Proposition 7.13. *If m is represented by Q_2 and n is represented by Q_3, then mn is represented by Q_1.*

Note that $5 = 2 \cdot 1^2 + 3 \cdot 1^2$ and $11 = 3 \cdot 1^2 + 2 \cdot 2^2$ are represented by $(2, 0, 3)$ and $(3, 0, 2)$, respectively, but not by $(1, 0, 6)$; yet their product is represented by $(1, 0, 6)$ since $55 = 7^2 + 6 \cdot 1^2$.

Proof. Assume that $m = gr^2 + fs^2$ and $n = ft^2 + gu^2$. Then

$$\begin{aligned} mn &= (gr^2 + fs^2)(ft^2 + gu^2) \\ &= fg(r^2t^2 + s^2u^2) + f^2s^2t^2 + g^2r^2u^2 \\ &= (fst + gru)^2 + fg(rt - su)^2, \end{aligned}$$

and now the claim follows. □

If you look at this identity carefully, you will notice that it can be stated as $[Q_2] \oplus [Q_3] = [Q_1]$.

Actually, similar calculations show that if m is represented by Q_i and n by Q_j, then mn is represented by Q_k for any permutation (i, j, k) of $(1, 2, 3)$.

Dirichlet Composition

If you actually have to compose two given classes of quadratic forms, Gauss's definition of composition is not very helpful. Dirichlet came up with an algorithm that can be performed by hand.

The basic idea is the following: since we are only interested in the equivalence class of the composition of $[Q_1]$ and $[Q_2]$, we may use the action of $\mathrm{SL}_2(\mathbb{Z})$ to replace Q_1 and Q_2 by equivalent forms before we compose them.

Let us now call quadratic forms (A, B, C) and (A', B', C') with nonsquare discriminant Δ *concordant* if the coefficients have the following properties:

1. $B = B'$;
2. $A' \mid C$ and $A \mid C'$.

The composition of concordant forms turns out to be extremely simple:

Proposition 7.14. *If Q and Q' are concordant, then $Q = (A, B, A'C)$ and $Q' = (A', B, AC)$ for integers A, A', B, C, and we have $[Q] \oplus [Q'] = [Q'']$ for $Q'' = (AA', B, C)$.*

Proof. This is basically a consequence of the following identity:

$$(Ar^2 + Brs + A'Cs^2)(A't^2 + Btu + ACu^2) = AA'x^2 + Bxy + Cy^2,$$

where $x = rt - Csu$ and $y = Aru + A'st + Bsu$.

The fact that Dirichlet composition defines a composition of classes will be proved below. □

Example 1. The forms $(1, 0, -m)$ and $(1, 1, m)$ of discriminant $\Delta = 4m$ and $\Delta = 1 - 4m$, respectively, generate the classes that serve as the neutral elements in the respective class groups.

In fact, $Q = [(1, 0, -m)]$ and Q are concordant, hence $2[Q] = [(1, 0, -m)] = Q$ in the first case, and $2[Q] = [(1, 1, -m)]$ in the second case.

Moreover, the inverse of $[(A, B, C)]$ is $[(A, -B, C)]$ because $[(A, B, C)] \oplus [(A, -B, C)] = [(A, B, C)] \oplus [(C, B, A)] = [(AC, B, 1)] = [(1, -B, AC)]$. The last form $(1, -B, AC)$ is equivalent to the principal form (see Exercise 7.9).

Example 2. Consider the form $Q = (2, 1, 2)$ with discriminant $\Delta = -15$. Here $2[Q] = [Q] + [Q] = [Q']$ for $Q' = (4, 1, C'')$, and $-15 = 1^2 - 4 \cdot 4C''$ gives $C'' = 1$. Thus $2[Q] = [(4, 1, 1)]$, and since $(4, 1, 1) \sim (1, 1, 4)$, $[Q]$ has order 2 in $\mathrm{Cl}(-15)$.

Example 3. The forms (A, B, C) and (C, B, A) are concordant, and we find $[(A, B, C)] \oplus [(C, B, A)] = [(AC, B, 1)] = [(1, -B, AC)]$.

Example 4. Consider the form $Q = (2, 1, 3)$ with discriminant $\Delta = -23$. Here Q and Q are not concordant, so we have to replace them by equivalent forms before we can compose them. Now $Q' = Q|_M = (2, -3, 4)$ for $M = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, hence $2[Q] = 2[Q'] = [Q'']$ for $Q'' = (4, -3, 2)$. Now $(4, -3, 2) \sim (2, 3, 4) \sim (2, -1, 3)$.

By Example 3, we also have $3[Q] = [(2, 1, 3)] \oplus [(2, -1, 3)] = [(2, 1, 3)] \oplus [(3, 1, 2)] = [(6, 1, 1)] = [1, -1, 6] = [1, 1, 6]$.

Now we need to show that, given forms Q_1, Q_2, we can always find equivalent forms $Q_1' \sim Q_1$ and $Q_2' \sim Q_2$ such that Q_1' and Q_2' are concordant. To this end we need

Lemma 7.15. *Let $Q = (A, B, C)$ be a primitive quadratic form. Then for any $N \in \mathbb{N}$ there are $r, s \in \mathbb{Z}$ such that $Q(r, s)$ is coprime to N.*

Proof. Write $N = rst$, where $(r, C) = 1$, and where the primes $p \mid s$ and $q \mid t$ satisfy $p \mid C$, $p \nmid A$, $q \mid A$ and $q \mid C$. Then we find

$$\begin{aligned} \gcd(Q(r, s), r) &= \gcd(Cs^2, r) = 1, \\ \gcd(Q(r, s), s) &= \gcd(Ar^2, s) = 1, \\ \gcd(Q(r, s), t) &= \gcd(Brs, t) = 1, \end{aligned}$$

where we have used that $\gcd(B, t) = 1$ because t is primitive. □

Now we can construct concordant forms:

Lemma 7.16. *Let Q_1 and Q_2 be quadratic forms of discriminant Δ. Then for any $N \in \mathbb{N}$ there exist concordant forms $Q_1' = (A, B, C)$ and $Q_2' = (A', B, C')$ equivalent to Q_1 and Q_2, respectively, and such that* $\gcd(A, A') = \gcd(AA', N) = 1$.

Proof. First we show that we can choose a form $R_1 = (A_1, B_1, C_1) \sim Q_1$ with $\gcd(A_1, N) = 1$. In fact, pick $r, s \in \mathbb{Z}$ coprime with $A_1 = Q_1(r, s)$ and $\gcd(A_1, N) = 1$. By Bezout there are $t, u \in \mathbb{Z}$ with $ru - st = 1$; then $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Now $R_1 = Q_1|_M = (A_1, B_1, C_1) \sim Q_1$.

Similarly we choose $R_2 = (A_2, B_2, C_2) \sim Q_2$ with $\gcd(A_2, A_1 N) = 1$.

Now we find integers n_1, n_2 such that $B_1 + 2A_1 n_1 = B_2 + 2A_2 n_2$. This equation can be written in the form $A_1 n_1 - A_2 n_2 = (b_1 - b_2)/2$, ans this has an integral solution because $b_1 \equiv \Delta \equiv b_2 \bmod 2$ and $\gcd(A_1, A_2) = 1$. Now put $M_i = \left(\begin{smallmatrix} 1 & n_i \\ 0 & 1 \end{smallmatrix}\right)$ and $B = B_i + 2A_i n_i$; then the forms $Q_i' = R|_{M_i} = (A_j, B, C_j)$ have the desired properties. □

Our next result shows that composition respects equivalence classes:

Proposition 7.17. *If Q_1, Q_2, R_1, R_2 are quadratic forms of discriminant Δ, and if $Q_i \sim R_i$, then $Q_1 * Q_2 \sim R_1 * R_2$.*

Here we have written $Q_1 * Q_2$ for the form obtained through Dirichlet composition from Q_1 and Q_2. Once we have shown that this composition only depends on equivalence classes, we will denote composition again by $[Q_1] \oplus [Q_2]$.

Proof. Later. See Flath's book for now. □

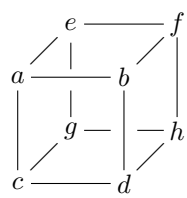
Now we finally can show

Theorem 7.18. *Let Δ be a nonsquare discriminant. Then composition makes the equivalence classes of binary quadratic forms of discriminant Δ into a finite abelian group.*

Proof. Later. See Flath's book for now. □

7.5 Bhargava's Cubes

Manjul Bhargava found a new interpretation of Gauss composition using cubes of integers such as



.

Each such cube can be sliced in three different ways, producing three pairs of 2×2-matrices

$$M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad N_1 = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

$$M_2 = \begin{pmatrix} a & c \\ e & g \end{pmatrix}, \qquad N_2 = \begin{pmatrix} b & d \\ f & h \end{pmatrix},$$

$$M_3 = \begin{pmatrix} a & e \\ b & f \end{pmatrix}, \qquad N_3 = \begin{pmatrix} c & g \\ d & h \end{pmatrix}.$$

To each cube A we can associate three binary quadratic forms $Q_i = Q_i^A$ by putting[1]

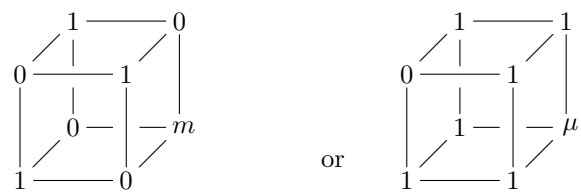
$$Q_i(x, y) = -\det(M_i x + N_i y).$$

A simple calculation shows that $\operatorname{disc} Q_1 = \operatorname{disc} Q_2 = \operatorname{disc} Q_3$; thus we can define

[1] The following formulas differ from those found in Bhargava's work; his conventions might possibly be the better choice if you go on to study cubic forms etc.

$$\begin{aligned}\operatorname{disc}(A) &= a^2h^2 + b^2g^2 + c^2f^2 + d^2e^2 \\ &\quad - 2(abgh + cdef + acfh + bdeg + aedh + bfcg) + 4(adfg + bceh).\end{aligned}$$

Now let d be the discriminant of a quadratic number field. Then we define the principal cube A_Δ by



according as $d = 4m$ or $d = 4m + 1 = 4\mu - 3$, with $\mu = m + 1$. Note that these cubes are "triply symmetric": rotation by 120° about the long diagonal containing m and μ, respectively.

In order to become more familiar with these cubes, let us compute the quadratic forms associated to the cube A_Δ. In the case $\Delta = 4m$ we find

$$M_1 = M_2 = M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad N_1 = N_2 = N_3 = \begin{pmatrix} 1 & 0 \\ 0 & m \end{pmatrix},$$

hence $Q_1 = Q_2 = Q_3$. Moreover,

$$M_1x + N_1y = \begin{pmatrix} y & x \\ x & my \end{pmatrix},$$

therefore $Q_1 = -\det(M_1x + N_1y) = x^2 - my^2$.

For $d = 4m + 1$ we similarly get

$$M_1 = M_2 = M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad N_1 = N_2 = N_3 = \begin{pmatrix} 1 & 1 \\ 1 & \mu \end{pmatrix},$$

hence

$$M_1x + N_1y = \begin{pmatrix} y & x+y \\ x+y & x+\mu y \end{pmatrix},$$

hence $Q_1 = Q_2 = Q_3 = -\det(M_1x + N_1y) = x^2 + xy - my^2$. This is a quadratic form of discriminant $\Delta = 4m + 1$.

Note that

$$\begin{aligned} x^2 - my^2 &= N(x + y\sqrt{m}\,), \qquad\qquad \text{and} \\ x^2 + xy - my^2 &= N\big(x + y\tfrac{1+\sqrt{m}}{2}\big) \end{aligned}$$

7.6 Bhargava Composition

Let us consider the cubes

$$\begin{tikzcd}
 & 0 \arrow[rr, dash] \arrow[dl, dash] \arrow[dd, dash] & & -f \arrow[dl, dash] \arrow[dd, dash] \\
1 \arrow[rr, dash] \arrow[dd, dash] & & 0 \arrow[dd, dash] & \\
 & -g \arrow[rr, dash] \arrow[dl, dash] & & 0 \arrow[dl, dash] \\
0 \arrow[rr, dash] & & -1 &
\end{tikzcd} \tag{7.6}$$

and compute the associated forms:

$$Q_1 = (1, 0, fg), \quad Q_2 = (g, 0, f), \quad Q_3 = (f, 0, g).$$

We have already seen, using Gauss composition, that $[Q_2] \oplus [Q_3] = [Q_1]$. Since $[Q_1] = 0$ is the neutral element in the class group, we can also write this in the form $[Q_1] \oplus [Q_2] \oplus [Q_3] = 0$. This suggests the following definition: for binary quadratic forms Q_1, Q_2, Q_3 of discriminant Δ we say that $[Q_1] \oplus [Q_2] \oplus [Q_3] = 0$ if there is a cube A such that $Q_j = Q_j^A$ for $j = 1, 2, 3$.

Let us now explain why it is important to talk about composition of classes (and not of forms) here. In the cube (7.6), add the back face to the front face. Here's what you get:

$$\begin{tikzcd}
 & 0 \arrow[rr, dash] \arrow[dl, dash] \arrow[dd, dash] & & -f \arrow[dl, dash] \arrow[dd, dash] \\
1 \arrow[rr, dash] \arrow[dd, dash] & & -f \arrow[dd, dash] & \\
 & -g \arrow[rr, dash] \arrow[dl, dash] & & 0 \arrow[dl, dash] \\
-g \arrow[rr, dash] & & -1 &
\end{tikzcd}$$

Computing the associated quadratic forms we find

$$Q_1' = (1 + fg, -2fg, fg), \quad Q_2' = (g, 0, f), \quad Q_3' = (f, 0, g).$$

Since $Q_2 = Q_2'$ and $Q_3 = Q_3'$, a composition of forms using cubes would imply the relations $Q_1 \oplus Q_2 \oplus Q_3 = 0$ and $Q_1' \oplus Q_2 \oplus Q_3 = 0$. If $\oplus$ would define a group law, we could conclude that $Q_1 = \ominus Q_2 \ominus Q_3 = \ominus Q_2' \ominus Q_3' = Q_1'$, yet these two forms are definitely not equal.

Let us now analyze where the relation $Q_1' \oplus Q_2 \oplus Q_3 = 0$ is coming from. The original cube had the slicings

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad N_1 = \begin{pmatrix} 0 & -f \\ -g & 0 \end{pmatrix}, \quad M_1 x + N_1 y = \begin{pmatrix} x & -fy \\ -gy & -x \end{pmatrix},$$

$$M_2 = \begin{pmatrix} 1 & 0 \\ 0 & -g \end{pmatrix}, \qquad N_2 = \begin{pmatrix} 0 & -f \\ -1 & 0 \end{pmatrix}, \quad M_2 x + N_2 y = \begin{pmatrix} x & -fy \\ -y & -gx \end{pmatrix},$$

$$M_3 = \begin{pmatrix} 1 & 0 \\ 0 & -f \end{pmatrix}, \qquad N_3 = \begin{pmatrix} 0 & -g \\ -1 & 0 \end{pmatrix}, \quad M_3 x + N_3 y = \begin{pmatrix} x & -gy \\ -y & -fx \end{pmatrix}.$$

After adding the back to the front face, we have

$$M_1' = \begin{pmatrix} 1 & -f \\ -g & -1 \end{pmatrix}, \quad N_1' = \begin{pmatrix} 0 & -f \\ -g & 0 \end{pmatrix}, \quad M_1'x + N_1'y = \begin{pmatrix} x & fx+fy \\ gx+gy & -x \end{pmatrix},$$
$$M_2' = \begin{pmatrix} 1 & 0 \\ -g & -g \end{pmatrix}, \quad N_2' = \begin{pmatrix} -f & -f \\ 0 & -1 \end{pmatrix}, \quad M_2'x + N_2'y = \begin{pmatrix} x+fy & fy \\ y-gx & -gx \end{pmatrix},$$
$$M_3' = \begin{pmatrix} 1 & 0 \\ -f & -f \end{pmatrix}, \quad N_3' = \begin{pmatrix} -g & -g \\ 0 & -1 \end{pmatrix}, \quad M_3'x + N_3'y = \begin{pmatrix} x-gy & -gy \\ -y-fx & -fx \end{pmatrix}.$$

Thus adding the back to the front face results in elementary row operations on the matrices $M_ix + N_iy$ for $i = 2, 3$, hence $Q_2' = Q_2$ and $Q_3' = Q_3$.

We can describe these changes using matrices if we agree that $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right)$ acts on (M_1, N_1) by replacing it with $(rM_1 + tN_1, sM_1 + uN_1)$. Then adding the back to the front face means applying $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ to the cube.

Lemma 7.19. *Let A be a cube, $M \in \mathrm{SL}_2(\mathbb{Z})$, and let $A' = A|_M$ be the cube we get by letting M act on A; then $\operatorname{disc} A' = \operatorname{disc} A$. If the associated quadratic forms are denoted by Q_i and Q_i', then $Q_1' = Q_1|_M$, $Q_2' = Q_2$, and $Q_3' = Q_3$.*

Proof. We know that $Q_1 = -\det(M_1x + N_1y)$; applying $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right)$ we see that

$$\begin{aligned} Q_1'(x,y) &= -\det\big((rM_1 + tN_1)x + (sM_1 + uN_1)y)\big) \\ &= -\det\big(M_1(rx+sy) + N_1(tx+uy)\big). \end{aligned}$$

Since $Q_1 = (A, B, C) = -\det(M_1x + N_1y)$, we find

$$\begin{aligned} Q_1'(x,y) &= A(rx+sy)^2 + B(rx+sy)(tx+uy) + C(tx+uy)^2 \\ &= (A', B', C') \end{aligned}$$

for

$$\begin{aligned} A' &= Ar^2 + Brt + Ct^2, \\ B' &= 2(Ars + Ctu) + B(ru + st), \\ C' &= As^2 + Bsu + Cu^2. \end{aligned}$$

Thus we see that $Q_1'(x,y) = Q|_M(x,y)$ as claimed.

The proof of $Q_2' = Q_2$ is a lengthy calculation best done by `pari`. If you prefer working by hand, recall that every $M \in \mathrm{SL}_2(\mathbb{Z})$ can be written as a product of matrices $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$; thus it suffices to prove the claim for these two matrices. We already did this for the first one, and the calculations for the second one can be done manually. □

Now instead of letting $\mathrm{SL}_2(\mathbb{Z})$ act on the pair (M_1, N_1) as above we can also let it act on (M_2, N_2) and (M_3, N_3). In this way we get an action of the group $\Gamma = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ on the set of cubes; note that the action

of the three factors in Γ commutes: if you let an element (T_1, T_2, T_3) act on a cube then it does not matter whether you first let T_1 act on (M_1, N_1) and then T_2 on (M_2, N_2) or the other way round (check this!).

Observe also that the action of the subgroup $I \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ of Γ is trivial on the quadratic form Q_1, since this subgroup acts by row and column operations on M_1 and N_1, hence does not change the determinant $\det(M_i x + N_i y)$.

What have we achieved now? We know that if Q_1, Q_2, Q_3 are quadratic forms attached to some cube A and if $M \in \mathrm{SL}_2(\mathbb{Z})$, then there is a cube with associated quadratic forms $Q_1|_M$, Q_2, Q_3. This shows that we cannot hope for a composition law on quadratic forms; what we should be looking for is a composition law on equivalence classes of quadratic forms, where two forms Q and Q' are called equivalent if there exists a matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $Q' = Q|_M$.

Lemma 7.20. *The relation $Q \sim Q|_M$ for $M \in \mathrm{SL}_2(\mathbb{Z})$ is an equivalence relation.*

This is easily proved. Now let $[Q]$ denote the equivalence class of the quadratic form Q. Then we define a group law on the set of equivalence classes of quadratic forms with the same discriminant d by $[Q_1] \oplus [Q_2] \oplus [Q_3] = 0$ whenever there is a cube A with associated quadratic forms Q_1, Q_2, Q_3.

Actually, to get a group law from this formula we need to specify the neutral element or the inverse of a form. Let us fix the group law by demanding that the class I of the principal form Q_0 is the neutral element.

Bhargava Composition is Gauss Composition

Now we will show that Bhargava's composition law coincides with Gauss's. We start with a triviality:

Lemma 7.21. *The inverse of the class $[Q]$, where $Q = (A, B, C)$, is the class $[-Q]$, where $-Q = (A, -B, C)$.*

Proof. If $\Delta = 4m$, we consider the cube

$$\begin{tikzcd}
 & A \arrow[rr, dash] \arrow[dd, dash] & & -b \arrow[dd, dash] \\
0 \arrow[ur, dash] \arrow[rr, dash] \arrow[dd, dash] & & 1 \arrow[ur, dash] \arrow[dd, dash] & \\
 & b \arrow[rr, dash] & & -C \\
1 \arrow[ur, dash] \arrow[rr, dash] & & 0 \arrow[ur, dash] &
\end{tikzcd}.$$

Its quadratic forms are

$$\begin{aligned}
Q_1(x, y) &= x^2 - my^2, \\
Q_2(x, y) &= Ax^2 + Bxy + Cy^2, \\
Q_3(x, y) &= Ax^2 - Bxy + Cy^2.
\end{aligned}$$

This implies that $[I] + [Q] + [-Q] = 0 = [I]$, and now the claim follows. $\square$

Note that the cube A_d only shows that $I + I + I = 0$, which does not allow us directly to conclude that I is the neutral element with respect to composition.

Now we have shown that the classes of quadratic forms have inverses; for verifying that they form a commutative group we still have to show that composition is commutative and associative. Commutativity is easy: if there is a cube A with quadratic forms Q_1, Q_2, Q_3, then there is a cube A' with quadratic forms Q_2, Q_1, and Q_3. Associativity, on the other hand, is not easy to verify using the definition.

Our first proof that Bhargava's composition coincides with Gauss's is by showing that both give the same result for concordant forms. To this end we consider the cube A given by

$$\begin{tikzcd}
 & 0 \arrow[rr, dash] \arrow[dl, dash] \arrow[dd, dash] & & -C \arrow[dl, dash] \arrow[dd, dash] \\
1 \arrow[rr, dash] \arrow[dd, dash] & & 0 \arrow[dd, dash] & \\
 & -A' \arrow[rr, dash] \arrow[dl, dash] & & -B \arrow[dl, dash] \\
0 \arrow[rr, dash] & & -A &
\end{tikzcd}$$

Its associated quadratic forms are

$$\begin{aligned}
Q_1 &= Ax^2 + Bxy + A'Cy^2, \\
Q_2 &= A'x^2 + Bxy + ACy^2, \\
Q_3 &= Cx^2 + Bxy + AA'y^2.
\end{aligned}$$

Thus Bhargava's law implies

$$[Q_1] \oplus [Q_2] = -[Q_3] = [(C, -B, AA')] = [(AA', B, C)],$$

and this coincides with Dirichlet composition.

The following idea going back to Gauss, Dedekind and Speiser shows directly that Bhargava's and Gauss's composition laws coincide:

The Isomorphism. We now prove that Bhargava's construction defines a group law on equivalence classes of binary quadratic forms; we will accomplish this by writing down a bijection between such classes and ideal classes in quadratic number fields. In order to simplify things (and for lack of time),

we will only do this for negative discriminants Δ that are also discriminants of quadratic fields (note that we have worked with quadratic forms $x^2 + 4y^2$ of discriminant -16, which is not a field discriminant).

We now claim that the map $i : [Q] \longrightarrow [i(Q)]$ is a homomorphism from the set of equivalence classes of quadratic forms with discriminant $\Delta < 0$ to the ideal class group of the complex quadratic number field K with discriminant Δ.

The map will be a homomorphism if we can show that $[Q_1] \oplus [Q_2] \oplus [Q_3] = 0$ implies that $I_1 I_2 I_3 \sim (1)$, where $i(Q_j) = I_j$.

Assume that $I_1 I_2 I_3 = (\alpha)$; replacing I_3 by $\frac{1}{\alpha} I_3$ we may assume that $I_1 I_2 I_3 = (1)$. write $I_1 = [\alpha_1, \alpha_2]$, $I_2 = [\beta_1, \beta_2]$, and $I_3 = [\gamma_1, \gamma_2]$. Then for $i, j, k \in \{1, 2\}$ there exist a_{ijk}, c_{ijk} such that

$$\alpha_i \beta_j \gamma_k = c_{ijk} + a_{ijk}\sqrt{m}.$$

Since $I_1 I_2 I_3 = (1)$, the elements on the right hand side are algebraic integers, hence $a_{ijk}, c_{ijk} \in \mathbb{Z}$.

The eight integers a_{ijk} define a cube A in a natural way, whose associated quadratic forms Q_i correspond, up to equivalence, to the ideals I_i.

Theorem 7.22. *If $\Delta < 0$, the correspondence between ideal classes and classes of quadratic forms induces an isomorphism between the ideal class group $\mathrm{Cl}(K)$ of the field $K = \mathbb{Q}(\sqrt{\Delta}\,)$ and the class group $\mathrm{Cl}(\Delta)$ of quadratic forms.*

7.7 Cubic Forms

In his first article published in 1844, Eisenstein sets out to generalize Gauss's theory of binary quadratic forms to cubic forms

$$ax^3 + 3bx^2y + 3cxy^2 + dy^3.$$

To each such cubic he associates the binary quadratic form

$$(A, B, C) = Ax^2 + 2Bxy + Cy^2,$$

where

$$A = b^2 - ac, \quad 2B = bc - ad, \quad C = c^2 - bd.$$

He then observes that a transformation $x = \alpha x' + \beta y'$, $y = \gamma x' + \delta y'$ with $\alpha\delta - \beta\gamma = 1$ applied to the cubic will have the corresponding effects on the associated binary quadratic form.

This is actually a special case of a later result due to Hesse. Given a plane algebraic curve $f(x, y) = 0$, he defined what today is called the Hessian as the curve with equation

$$\begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = 0.$$

For the special case of the cubic $f(x, y) = ax^3 + 3bx^2y + 3cxy^2 + dy^3$ we get, up to a constant, the Hessian turns out to be the conic

$$0 = -\begin{vmatrix} ax + by & bx + cy \\ bx + cy & cx + dy \end{vmatrix} = Ax^2 + 2Bxy + Cy^2.$$

Exercises

7.1 Show that $\mathrm{SL}_2(\mathbb{Z})$ acts on the upper half plane, i.e., that $(MN)z = M(Nz)$ for $z \in \mathcal{H}$.

7.2 Show that $Q \sim Q|_M$ for $M \in \mathrm{SL}_2(\mathbb{Z})$ defines an equivalence relation on the set of binary quadratic forms of fixed discriminant Δ.

7.3 Show that if Q corresponds to $z \in \mathcal{H}$, then for $M \in \mathrm{SL}_2(\mathbb{Z})$, the form $Q|_M$ corresponds to $M^{-1}z$. In particular, $Q|_{MN}$ corresponds to $(MN)^{-1}z = N^{-1}M^{-1}z$.

7.4 Show that if a group G acts on X from the left via $(g,x) \longmapsto gx$, then G acts on X from the right via $(g,x) \longmapsto xg^{-1}$.

7.5 Show that the two binary quadratic forms $(1,0,3)$ and $(1,1,1)$ represent the same integers, but that they are not equivalent.

7.6 Show that every form with discriminant $\Delta < 0$ represents some integer $n \neq 0$ with $n \leq \sqrt{-\Delta/3}$.

7.7 Show that if (A,B,C) and (A',B',C') have the same discriminant, then $B \equiv B' \bmod 2$.

7.8 Show that $[\alpha,\beta] = [\gamma,\delta]$ for elements $\alpha,\beta,\gamma,\delta \in \mathcal{O}_K$ if and only if there is some $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $\left(\begin{smallmatrix}\gamma\\ \delta\end{smallmatrix}\right) = M\left(\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right)$.

7.9 Show that

$$(1,-B,AC) \sim \begin{cases} (1,0,-m) & \text{if } B^2-4AC = 4m, \\ (1,1,-m) & \text{if } B^2-4AC = 4m+1. \end{cases}$$

Bibliography

1. N. C. Ankeny, S. Chowla, H. Hasse, *On the class number of the real subfield of a cyclotomic field*, J. Reine Angew. Math. **217** (1965), 217–220; cf. p. 48
2. M. Bhargava, *Higher composition laws. I: A new view on Gauss composition, and quadratic generalizations*, Ann. Math. **159** (2004), 217–250
3. H. Cohn, *Advanced Number Theory*, Dover
4. R. Dedekind, *Über trilineare Formen und die Komposition der binären quadratischen Formen*, J. Reine Angew. Math. **129** (1905), 1–34
5. D.E. Flath, *Introduction to Number Theory*, Wiley 1989
6. H. Hasse, *Über mehrklassige, aber eingeschlechtige reell-quadratische Zahlkörper*, El. Math. **20** (1965), 49–59
7. J. Riss, *La composition des formes quadratiques binaires (d'après Gauss)*, Sém. Théor. Nombr. Bordeaux (1978), exp. 18, 16pp
8. D. Shanks, *Class number, a theory of factorization, and genera*, Proc. Symp. Pure Math. **20** (1970), 415–440
9. D. Shanks, *The infrastructure of a real quadratic field and its applications*, Proc. Number Theory Conf. Boulder 1972, 217–224
10. D, Shanks, *A matrix underlying the composition of quadratic forms and its implications for cubic extensions*, Notices Amer. Math. Soc. **25** (1978), p. A305
11. D. Shanks, *On Gauss and Composition I*, in *Number Theory and Applications* (R. Mollin, ed.), 1989, 163–178
12. D. Shanks, *On Gauss and Composition II*, in *Number Theory and Applications* (R. Mollin, ed.), 1989, 179–204
13. Speiser, , H. Weber-Festschrift (1912), 375–395
14. D. Zagier, *The Birch-Swinnerton-Dyer Conjecture from a naive point of view*, Arithmetic Algebraic Geometry (G. van der Geer et al., eds.), 377–389, Birkhäuser 1991 109