Franz Lemmermeyer

Algebraic Number Theory

December 14, 2006

Contents

1. Fermat, Euler, and Nonunique Factorization

In this chapter we will motivate the introduction of algebraic integers by showing how Euler used these to solve diophantine equations in the ordinary integers. We will also recall the basic definitions like divisibility, units, primes, and irreducibles, and discuss unique factorization domains.

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

[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 $\mathbb{Z}[\sqrt{-2}\,]$.

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]

Thus Euler's "proof" sometimes produces wrong results; we will see below that this is due to the fact that unique factorization holds in the ring $\mathbb{Z}[\sqrt{-2}\,]$, but not in $\mathbb{Z}[\sqrt{-5}\,]$.

1.2 Unique Factorization Domains

Let R be a ring; we say that $a \neq 0$ is a zero divisor if there is a nonzero $b \in R$ such that $ab = 0$. A commutative ring with unit element 1 and without zero divisors is called a domain.

From now on, let R be a domain. We say that $b \mid a$ for elements $a, b \in R$ if there is some $c \in R$ such that $a = bc$. Elements dividing 1 are called units. The units of $R = \mathbb{Z}$ are ± 1.

Proposition 1.1. *The units in R form a group $R^\times$.*

Proof. We first show that $R^\times$ is closed under multiplication. To this end, let $a, b \in R^\times$; then there exist $c, d \in R$ such that $ac = 1$ and $bd = 1$. But then $(ac)(bd) = 1$ (we have used commutativity), hence $ac \in R^\times$.

Next, the unit element $1 \in R$ is a unit and serves as the neutral element of $R^\times$. If $a \in R^\times$, then $ac = 1$ for some $c \in R$; clearly c is a unit, hence every element of $R^\times$ has an inverse. Finally, $R^\times$ inherits associativits from R. □

It is quite easy to determine all units in the rings $R = \mathbb{Z}[\sqrt{m}\,]$ for negative integers m:

Proposition 1.2. *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 $R^\times = \{-1, +1\}$.*

For the proof of this result we introduce an important function, the norm. This is a multiplicative map $N : \mathbb{Q}(\sqrt{m}\,) \longrightarrow \mathbb{Q}$ defined as follows: for $\alpha = a + b\sqrt{m}$, put $\alpha' = a - b\sqrt{m}$. A simple calculation shows that $(\alpha\beta)' = \alpha'\beta'$. Now define the norm of α by $N(\alpha) = \alpha\alpha' = a^2 - mb^2$. The norm is multiplicative since $N(\alpha\beta) = (\alpha\beta)(\alpha\beta)' = \alpha\alpha'\beta\beta' = N(\alpha)N(\beta)$. Note that the norm, when restricted to $\mathbb{Z}[\sqrt{m}\,]$, gives a map $N : \mathbb{Z}[\sqrt{m}\,] \longrightarrow \mathbb{Z}$.

Lemma 1.3. *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 $N\varepsilon = \pm 1$.*

[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 $5 = x^2 - y^2 = (x - y)(x + y)$. 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.

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. 1.2. From Lemma 1.3 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 $a = \pm 1$. □

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 $p \mid b$.

Observe that primes (irreducibles) in $\mathbb{Z}$ need not be prime (irreducible) in number rings; for example, the equations $2 = (1 + i)(1 - i) = -\sqrt{-2} \cdot \sqrt{-2}$ show that 2 is not prime (not even irreducible) in $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$.

The norm is also useful for showing that certain elements are irreducible:

Lemma 1.4. *Let $\alpha \in \mathbb{Z}[\sqrt{m}]$ be an element whose norm is $N\alpha = \pm p$ for a prime p. Then α is irreducible in $\mathbb{Z}[\sqrt{m}]$.*

Proof. Assume that $\alpha = \beta\gamma$. Taking the norm gives $\pm p = N\alpha = N\beta N\gamma$. But since the left hand side is irreducible, we must have $N\beta = \pm 1$ or $N\gamma = \pm 1$. Thus β or γ is a unit, and this means that α is irreducible. □

We know the following result from elementary number theory:

Proposition 1.5. *Primes are irreducible.*

Proof. Let $\in R$ be prime, and assume that $p = ab$. We have to show that this factorization is trivial, i.e., that a or b is a unit in R. Since $p \mid ab$, the fact that p is prime implies that $p \mid a$ or $p \mid b$; assume without loss of generality that $p \mid a$, and write $a = pc$. Then $p = ab = pbc$, and since R is a domain, we must have $bc = 1$ (here is the simple argument: $0 = p - pbc = p(1 - bc)$; since $p \neq 0$ and since R does not have zero divisors, the second factor must be 0). But this shows that $b \in R^\times$. □

A domain R is called a unique factorization domain (UFD) or simply factorial if the following conditions are satisfied:

- every nonzero nonunit element of R has a factorization into irreducibles;
- these factorizations are essentially unique, that is, if $a = p_1 \cdots p_r = q_1 \cdots q_s$ for irreducible elements p_i and q_j, then $r = s$, and we can rearrange the factors in such a way that $p_i = u_i q_i$ for $i = 1, \ldots, r$ and for units $u_i \in R^\times$.

For example, the factorizations $6 = 2 \cdot 3 = (-3) \cdot (-2)$ in $\mathbb{Z}$ are esssentially the same since $2 = (-1)(-2)$ and $3 = (-1)(-3)$. In fact, we know from elementary number theory that $\mathbb{Z}$ is a UFD.

Let us now see why the domain $\mathbb{Z}[\sqrt{-5}\,]$ is not factorial. Consider the factorization $6 = (1 + \sqrt{-5}\,)(1 - \sqrt{-5}\,) = 2 \cdot 3$. Clearly the factors do not differ by units, since the only units in $\mathbb{Z}[\sqrt{-5}\,]$ are ± 1. We now claim that all the factors of 6 in the factorizations above are irreducible; this will then imply that $\mathbb{Z}[\sqrt{-5}\,]$ is not factorial.

In order to show that 2 is irreducible in $\mathbb{Z}[\sqrt{-5}\,]$, write $2 = \alpha\beta$ for $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}\,]$. We have to show that α or β is a unit. To this end, take the norm; we find $4 = N(2) = N(\alpha)N(\beta)$. This is an equation in the integers; in fact, since $N\alpha > 0$ it is an equation in natural numbers. Thus the only possibilities, are $N\alpha = 1$, $N\beta = 4$; $N\alpha = 4$, $N\beta = 1$; or $N\alpha = N\beta = 2$. Assume that $N\alpha = 2$ for $\alpha = a + b\sqrt{-5}$ and integers a, b. Then $a^2 + 5b^2 = 2$: but this equation does not have a solution: contradiction. Thus we must have $N\alpha = 1$ or $N\beta = 1$; but this implies that α or β is a unit.

The same method allows you to show that 3 and $1 \pm \sqrt{-5}$ are also irreducible.

Now 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 norms are ± 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.

1.3 Euclidean Rings

In the following, we will present techniques that allow us to prove that $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}\,]$ are indeed factorial. The key is the Euclidean algorithm. Recall that, in $\mathbb{Z}$, given any pair of nonzero integers a and b, we can find integers q and r such that $a = bq + r$ with $|r| < |b|$. The proofs that $\mathbb{Z}$ is factorial are all based more or less explicitly on this fact.

Not let R be an arbitrary domain, and assume that there is a map $f : R \longrightarrow \mathbb{N}$ with the following properties:

- $f(a) = 0$ if and only if $a = 0$;
- for all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ with $a = bq + r$ and $f(r) < f(b)$.

Then R is called a Euclidean ring, and f a Euclidean function on R. Clearly $\mathbb{Z}$ is a Euclidean ring with respect to the absolute value $f = |\cdot|$. We now

show that $\mathbb{Z}[i]$ is also Euclidean, and then prove that all Euclidean rings are actually factorial.

$\mathbb{Z}[i]$ is Euclidean

We claim that the norm is a Euclidean function on R. Clearly N maps Gaussian integers to the natural numbers, and $N(\alpha) = 0$ if and only if $\alpha = 0$. It remains to show that for nonzero $\alpha, \beta \in \mathbb{Z}[i]$ we can find $\gamma, \rho \in \mathbb{Z}[i]$ with $\alpha = \beta\gamma + \rho$ and $N(\rho) < N(\beta)$. Since the norm is multiplicative, this is equivalent to $\frac{\alpha}{\beta} = \gamma + \frac{\rho}{\beta}$ with $N(\frac{\rho}{\beta}) < 1$. Thus for any $\xi = \frac{\alpha}{\beta} \in \mathbb{Q}(i) = \{r + si : r, s \in \mathbb{Q}\}$ we have to find a Gaussian integer $\gamma \in \mathbb{Z}[i]$ such that $N(\xi - \gamma) < 1$.

This is done as follows. Write $\xi = r + si$ for rational numbers $r, s \in \mathbb{Q}$. Find integers $a, b \in \mathbb{Z}$ such that $|r-a| \leq \frac{1}{2}$ and $|s-b| \leq \frac{1}{2}$, and put $\gamma = a+bi$. Then $\xi - \gamma = t+ui$ with $|t|, |u| \leq \frac{1}{2}$, hence $N(\xi-\gamma) = t^2+u^2 \leq \frac{1}{4}+\frac{1}{4} = \frac{1}{2} < 1$.

The Euclidean Algorithm

For computing $\gcd(a, b)$ in a Euclidean ring (or for showing its existence in the first place), write

$$\begin{aligned}
a &= bq_1 + r_1, & f(r_1) &< f(b)\\
b &= r_1q_2 + r_2, & f(r_2) &< f(r_1)\\
r_1 &= r_2q_3 + r_3, & f(r_3) &< f(r_2)\\
&\dots\\
r_{n-2} &= r_{n-1}q_n + r_n, & f(r_n) &< f(r_{n-1})\\
r_{n-1} &= r_nq_{n+1}
\end{aligned}$$

Then the last nonzero remainder r_n is a gcd of a and b. In fact, the last equation tells us that $r_n \mid r_{n-1}$; the next to last equation then gives $r_n \mid r_{n-2}$, and working our way up we eventually find $r_n \mid a$ and $r_n \mid b$. Thus r_n is a common divisor. In order to show that it is the gcd, assume that $e \mid a$ and $e \mid b$. Starting from the top we now find $e \mid r_1, \dots, e \mid r_n$.

1.4 Principal Ideal Domains

Let R be a domain; a subring I of R is called an ideal if $RI = I$, that is, if $ri \in I$ for all $r \in R$ and all $i \in I$. In $\mathbb{Z}$, every subring is an ideal; in $\mathbb{Z}[i]$, the subring $\mathbb{Z}$ is not an ideal since $i \cdot 1 = i \notin \mathbb{Z}$.

For $a \in R$, the set $(a) = \{ar : r \in R\}$ is an ideal; it is called the principal ideal generated by R. More generally, for $a_1, \dots, a_m \in R$ the set $(a_1, \dots, a_m) = \{r_1a_1 + \dots + r_ma_m : r_i \in R\}$ is easily seen to be an ideal. If R

is a domain in which every ideal is principal, we call R a principal ideal ring (PID).

Our next goal is to show that every Euclidean ring is a PID, and that every PID is a UFD.

In order to become familiar with ideals, let us prove

Lemma 1.6. *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 1.7. *Let R be a ring. Then $(a) = (b)$ if and only of $a = bu$ for some unit $u \in R^\times$.*

Proof. From $(a) = (b)$ we get $(a) \subseteq (b) \subseteq (a)$, hence $b \mid a$ and $a \mid b$, or $a = bu$, $b = av$. Thus $a = bu = auv$; cancellation gives $uv = 1$, so $u, v \in R^\times$. □

Note that two factorizations $p_1 \cdots p_m = q_1 \cdots q_m$ into irreducible elements are essentially the same if, after some permutation of the indices, we have $(p_1) = (q_1), \ldots, (p_m) = (q_m)$. Thus the use of ideals in questions of unique factorization suppresses exactly the information we do not care about (factors differing by units).

Lemma 1.8. *Let R be a ring. Then $(a) \subseteq (a, b)$ for any $b \in R$.*

This is trivial.

Lemma 1.9. *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 1.10. *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 $e \mid d$.

□

Now we claim

Theorem 1.11. *Every Euclidean domain is a PID.*

Proof. Let I be an ideal in the Euclidean ring R; we have to show that I is principal. 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 1.12. *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, hence $p \mid b$. □

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 1.13. *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 $a = a_1 b_1$ with a_1 irreducible. If b_1 is irreducible, we are done; if not, write $b_1 = a_2 b_2$ 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 1.14. *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 $q_1 = p_1 u_1$ for some unit u_1; replacing q_2 by $q_2 u_1$ 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 $\mathbb{Z}$. □

Note that not every UFD is a PID; a well known counterexample is the factorial ring $\mathbb{Z}[X]$: here $(2, X)$ is not a principal ideal, i.e., cannot be generated by a single element.

1.5 The Ring of Gaussian Integers

Units and Primes

Finding all units in $R = \mathbb{Z}[i]$ is easy: a Gaussian integer is a unit if and only if its norm is 1, which immediately gives

Proposition 1.15. *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 1.16. *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!

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 1.17. *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.

1.6 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]

[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.

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

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

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

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 1.18. *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. □

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.

1.7 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" $3 = (21, 33)$ and $7 = (21, 49)$ 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 $\mathfrak{ab} = \{\sum a_i b_i : a_i \in \mathfrak{a}, b_i \in \mathfrak{b}\}$ 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

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

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

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

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

$$(21) = \mathfrak{pp}'\mathfrak{qq}'$$

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 a zero divisor in the ring $M_2(\mathbb{Z})$ of 2×2-matrices with integral coefficients.

1.2 Using the Euclidean algorithm, find the gcd of the Gaussian integers $10 + 11i$ and $11 + 16i$.

1.3 Show that the ring $\mathbb{Z}[\sqrt{-2}\,]$ is Euclidean.

1.4 Find units $\neq \pm1$ in the rings $\mathbb{Z}[\sqrt{m}\,]$ for $m = -1, 2, 3, 5, 6, 7$.

1.5 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.

1.6 Show that $21 = (1+2\sqrt{-5}\,)(1-2\sqrt{-5}\,) = 3\cdot7$ is another example of nonunique factorization in $\mathbb{Z}[\sqrt{-5}\,]$, and find a third factorization of 21 into irreducibles.

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

1.8 Discuss the factorization $6 = 2 \cdot 3 = \sqrt{6} \cdot \sqrt{6}$ in $\mathbb{Z}[\sqrt{6}\,]$.

1.9 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. $\mathfrak{q}^2 = (2 + \sqrt{-5}\,)$.

1.10 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 $\mathbb{Z}[\sqrt{10}\,]$.

1.11 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 $m + 1$.

1.12 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 $n \equiv 3 \bmod 4$?

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

1.14 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.

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

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

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

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

1.19 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 $a + bi \equiv 1 \bmod (2 + 2i)$.

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

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

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

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

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

1.25 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?

2. Rings of Integers, Modules, and Ideals

In this chapter we introduce the rings of integers in quadratic number fields. Then we shall deal with modules and ideals in these rings.

2.1 Algebraic Integers

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

Norm and Trace

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

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

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*

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

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 $m \equiv 3 \bmod 4$, 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*

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

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 easy.

2.2 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.

Given any $\alpha, \beta \in K = \mathbb{Q}(\sqrt{d}\,)$, the set $M = [\alpha, \beta] = \mathbb{Z}\alpha + \mathbb{Z}\beta$ of all $\mathbb{Z}$-linear combinations of α and β is a $\mathbb{Z}$-module. In the following, we will classify all full $\mathbb{Z}$-modules in $\mathcal{O}_K$.

Proposition 2.5. *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 2.5 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. 2.5. 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 $r + s\omega = r - uc + u(c + m\omega) = vn + u(c + m\omega) \in n\mathbb{Z} \oplus (c + m\omega)\mathbb{Z}$. □

If $n > 0$ and $0 \leq c < n$, then the integers c, n, m are uniquely determined, and we can think of $[n, c + m\omega]$ as the canonical representation of a module M.

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 2.6. *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 $0 \le r < n$; 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 $0 \le r, r' < n$ and $0 \le s, s' < m$; 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 $s = s'$. □

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 2.7. *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 $\gamma \in \mathcal{O}_K$. □

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.

In fact, consider the ring $\mathcal{O} = \mathbb{Z}[\sqrt{-3}\,]$ and $\alpha = 2$, $\beta = 1 + \sqrt{-3}$. Then $N\alpha = N\beta = 4$, and $\operatorname{Tr}\alpha\beta' = \operatorname{Tr}(2 - 2\sqrt{-3}\,) = 4$. Thus $N\alpha$, $N\beta$ and $\operatorname{Tr}\alpha\beta'$ are divisible by $m = 4$, yet $4 \nmid \alpha\beta' = 2 - 2\sqrt{-3}$.

If $M = [\alpha, \beta]$ and $N = [\gamma, \delta]$ are $\mathbb{Z}$-modules in $\mathcal{O}_K$, then we can define the product MN as the $\mathbb{Z}$-module containing all $\mathbb{Z}$-linear combinations of $\alpha\gamma$, $\alpha\delta$, $\beta\gamma$ and $\beta\delta$; in such a case we will write

$$MN = \langle \alpha\gamma, \alpha\delta, \beta\gamma, \beta\delta \rangle.$$

Proposition 2.8. *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 $N(M) = fg$.*

Full $\mathbb{Z}$-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 2.5 we can write $M = [\alpha, \beta]$ for $\alpha, \beta \in \mathcal{O}_K$ (actually Prop. 2.5 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 $\mathbb{Z}$); thus we get $MM' = f\langle \frac{N\alpha}{f}, \frac{N\beta}{f}, \frac{\alpha\beta'}{f}, \frac{\alpha'\beta}{f} \rangle$ (the generators of this $\mathbb{Z}$-module are all integral by Hurwitz's Lemma).

Now we claim that $1 \in \mathcal{O} = \langle \frac{N\alpha}{f}, \frac{N\beta}{f}, \frac{\alpha\beta'}{f}, \frac{\alpha'\beta}{f} \rangle$. In fact, 1 is a $\mathbb{Z}$-linear combination of $\frac{N\alpha}{f}$, $\frac{N\beta}{f}$ and $\frac{\mathrm{Tr}\,\alpha\beta'}{f}$ (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 $\frac{\alpha'\beta}{f}$. Now $\frac{N\alpha}{f}$ and $\frac{N\beta}{f}$ are multiples of 1, hence we have $\mathcal{O} = \langle 1, \frac{\alpha\beta'}{f}, \frac{\alpha'\beta}{f} \rangle$. Next $\frac{\alpha\beta'}{f} + \frac{\alpha'\beta}{f}$ is an integer, hence a multiple of 1, and this implies $\mathcal{O} = \langle 1, \frac{\alpha'\beta}{f} \rangle$.

Now we choose a canonical basis for M, namely $\alpha = n$ and $\beta = c + m\omega$. Then $\mathcal{O} = \langle 1, \frac{nc+nm\omega}{f} \rangle$; since $\frac{nc}{f}$ is an integer, we finally get $\mathcal{O} = \langle 1, \frac{nm}{f}\omega \rangle$. If $nm \neq 0$, then 1 and $\frac{nm}{f}\omega$ are independent, so in this case we get $\mathcal{O} = [1, \frac{nm}{f}\omega]$. In particular, $g = \frac{nm}{f}$ and $fg = nm = N(M)$ as claimed. □

This Proposition does not hold in arbitrary orders. Consider as before the ring $\mathbb{Z}[\sqrt{-3}\,]$ and the module $M = [2, 1 + \sqrt{-3}\,]$. Then it is easy to show that $N(M) = 2$, where $N(M) = (\mathbb{Z}[\sqrt{-3}\,] : M)$; in fact $\{0, 1\}$ is a complete system of residues mod M in $\mathcal{O}$. Now $M' = [2, 1 - \sqrt{-3}\,] = M$, and $MM' = [4, 2 + 2\sqrt{-3}\,] = 2M$. Since you cannot factor out any $f > 1$ from M (not inside the ring $\mathcal{O}$, anyway), this means that MM' cannot be written in the form $f\mathcal{O}$ for some order $\mathcal{O}$.

2.3 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

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

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 2.9. *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 $1 \in M$. □

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 $a + b\sqrt{-5} \bmod (2)$ 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.

[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.

An important result is

Theorem 2.10 (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. □

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

Proposition 2.11. *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. 2.5) 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. □

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 2.12. *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 2.13. *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. 2.5 we can write $\mathfrak{a} = m[a, b+\omega]$, and we have $\mathrm{N}(\mathfrak{a}) = m^2a$. It remains to show that $\mathfrak{a}\mathfrak{a}' = (m^2a)$. To this end, we compute

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

The last module is integral because of Proposition 2.11. 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). □

Lemma 2.14. *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 $g \mid b$. □

Proposition 2.13 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, $\mathfrak{a}\mathfrak{a}' = (\alpha\alpha') = (\mathrm{N}\alpha)$.

2.4 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 whole proof uses only two facts: for the existence of the prime ideal factorization, we need to know that norms of ideals are finite, and for proving uniqueness we need that for every ideal $\mathfrak{a}$ there is a nonzero ideal $\mathfrak{b}$ such that $\mathfrak{ab} = (\alpha)$ is principal.

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 $(2)(2, 1+\sqrt{-3}\,) = (1+\sqrt{-3}\,)(2, 1+\sqrt{-3}\,)$, 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 2.15. *If $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ are nonzero ideals in $\mathcal{O}_K$ with $\mathfrak{ab} = \mathfrak{ac}$, 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 $\alpha\mathfrak{b} = \mathfrak{ab} = \mathfrak{ac} = \alpha\mathfrak{c}$. 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 $\mathfrak{ab} = \mathfrak{ac}$ implies that $(\mathfrak{aa}')\mathfrak{b} = (\mathfrak{aa}')\mathfrak{c}$. Since $\mathfrak{aa}' = (N\mathfrak{a})$ is principal, the claim follows from the first part of the proof. □

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 $\mathfrak{b} = \mathfrak{ac}$. Since $\mathfrak{c} \subseteq \mathcal{O}_K$ we see $\mathfrak{b} = \mathfrak{ac} \subseteq \mathfrak{a}(1) = \mathfrak{a}$; this fact is often expressed

by saying "to divide is to contain". As a matter of fact, the converse is also true:

Proposition 2.16. *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 2.16 in the proof.

Proposition 2.17. *In rings of integers of qadratic number fields, prime ideals $\neq (0)$ 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 2.18. *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. □

2.5 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 2.19. *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 2.20. *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

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

Theorem 2.21. *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 2.11 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 squarefree.

- 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 $(d/2) = 0$.

2.6 Examples

Let us now investigate the example of $R = \mathbb{Z}[\sqrt{-5}]$, and start by listing all prime ideals of norm ≤ 7. First we observe that $\Delta = -4 \cdot 5 = -20$.

- $2 \mid \Delta$, so $(2) = \mathfrak{p}_2^2$ for $\mathfrak{p}_2 = (2, 1+\sqrt{-5})$. The ideal $\mathfrak{p}_2$ has norm 2.
- $(\frac{-5}{3}) = +1$, and $x^2 \equiv -5 \bmod 3$ has the solution $x = 1$. Thus $(3) = \mathfrak{p}_3\mathfrak{p}_3'$ for $\mathfrak{p}_3 = (3, 1+\sqrt{-5})$ and $\mathfrak{p}_3' = (3, 1-\sqrt{-5})$. Both ideals have norm 3.
- There are no prime ideals of norm 4; such a prime ideal $\mathfrak{p}$ would have to divide 2 (this follows from $\mathfrak{p} \mid N\mathfrak{p} = 4 = 2 \cdot 2$ and the assumption that $\mathfrak{p}$ is prime), hence must be equal to $\mathfrak{p}_2$.

- $5 \mid \Delta$, hence $(5) = \mathfrak{p}_5^2$ with $\mathfrak{p}_5 = (5, \sqrt{-5}\,) = (\sqrt{-5}\,)$. This is an ideal with norm 5.
- Prime ideals have prime power norm, so there are no prime ideals of norm 6.
- $(\frac{-5}{7}) = +1$, and $x^2 \equiv -5 \bmod 7$ has the solutio n $x = 3$; thus $(7) = \mathfrak{p}_7\mathfrak{p}_7'$ with $\mathfrak{p}_7 = (7, 3+\sqrt{-7}\,)$ and $\mathfrak{p}_7' = (7, 3-\sqrt{-7}\,)$. Both ideals have norm 7.

Now let us compute a few prime ideal factorizations. Let us start with the principal ideal $\mathfrak{a} = (5+\sqrt{-5}\,)$. We see that $N\mathfrak{a} = 30$, so it must be the product of prime ideals of norms 2, 3, and 5. Since there is only one prime ideal with norm 2 and 5, respectively, we know immediately that either $\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5$ or $\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3'\mathfrak{p}_5$. To find out which, we have to check whether $\alpha = 5 + \sqrt{-5}$ is an element of $\mathfrak{p}_3$ or $\mathfrak{p}_3'$: in fact, $(\alpha) = \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5$ implies $\alpha \in \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5 \subseteq \mathfrak{p}_3$.

This is most easily decided by computing $\alpha \bmod \mathfrak{p}_3$. Clearly $\alpha = 5 + \sqrt{-5} \equiv 5 + \sqrt{-5} - (1 + \sqrt{-5}\,) = 4 \equiv 4 - 3 \equiv 1 \bmod \mathfrak{p}_3$, and this shows that $\alpha \notin \mathfrak{p}_3$ (otherwise $\mathfrak{p}_3$ would contain 1, i.e., would be the unit ideal). As a check, we compute $\alpha = \alpha + 1 - \sqrt{-5} = 6 \equiv 0 \bmod \mathfrak{p}_3'$. Thus $\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3'\mathfrak{p}_5$.

Exercises

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

2.2 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 $A = a\mathbb{Z}$.

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

2.4 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 $a, b \in \mathbb{Z}$.

2.5 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.

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

2.7 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.

2.8 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).

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

2.10 Show that $[n, c + m\omega]$ is a full $\mathbb{Z}$-module if and only if $mn \neq 0$.

2.11 Let $R = \mathbb{Z}[X]$, and consider $\mathfrak{a} = (2, X)$. Show that there does not exist an ideal $\mathfrak{b} \neq (0)$ in R such that $\mathfrak{ab}$ is principal.

2.12 Let $\mathfrak{p} = (p, x + \sqrt{m}\,)$ be an ideal in $R = \mathcal{O}_K$, where p is an odd prime and $p \mid (x^2 - m)$. Find a $\mathbb{Z}$-basis of $\mathfrak{p}$ as a $\mathbb{Z}$-module.

2.13 Show that the ideal $(2, 1 + \sqrt{-5}\,)$ equals the $\mathbb{Z}$-module $[2, 1 + \sqrt{-5}\,]$.

2.14 Show that the $\mathbb{Z}$-module $M = [2, 1 + 3\sqrt{-5}\,]$ has norm 6, and that $MM' = 2[1, 3\sqrt{-5}\,]$.

2.15 Show that the full $\mathbb{Z}$-module $M = [2, \sqrt{-3}\,]$ has stabilizer ring $\mathcal{O} = [1, 2\sqrt{-3}\,]$.

2.16 Show that every order is its own stabilizer ring.

2.17 Let $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$. Prove Fermat's little theorem: $\alpha^{N\mathfrak{p}} \equiv \alpha \bmod \mathfrak{p}$ for all $\alpha \in \mathcal{O}_K$. (Hint: transfer the proof from elementary number theory to $\mathcal{O}_K$.)

2.18 Let m be a squarefree integer and p a prime number with $(\frac{m}{p}) = -1$. Derive the congruence $(a + b\sqrt{m}\,)^p \equiv a - b\sqrt{m}$ for $a, b \in \mathbb{Z}$. What happens if $(\frac{m}{p}) = +1$?

2.19 Let $K = \mathbb{Q}(\sqrt{m}\,)$ be a quadratic number field, where m is squarefree. Prove the following:

- 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.

2.20 Show that $(7, 1 + \sqrt{-5},) = (1)$.

2.21 Show more generally that $(a, \alpha) = (1)$ for $a \in \mathbb{Z}$ and $\alpha \in \mathcal{O}_K$ with $\gcd(a, N\alpha) = 1$.

2.22 Compute the prime ideal factorization of $(4 + \sqrt{-5}\,)$.

2.23 Find all prime ideals of norm ≤ 7 in $\mathbb{Z}[\sqrt{-6}\,]$, and compute the prime ideal factorizations of $(3 + \sqrt{-6}\,)$ and $(6 + \sqrt{-6}\,)$.

2.24 Find all prime ideals of norm ≤ 5 in $\mathcal{O}_K$ for $K = \mathbb{Q}(\sqrt{-23}\,)$, and compute the prime ideal factorizations of (ω) and $(1 + \omega)$, where $\omega = \frac{1+\sqrt{-23}}{2}$.

3. 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.

3.1 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 3.1. *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. □

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 3.2. *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). □

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

$$3^2 - 11 = -2,$$
$$4^2 - 11 = +5.$$

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

$$6^2 - 11 \cdot 2^2 = -8,$$
$$7^2 - 11 \cdot 2^2 = +5.$$

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.

3.2 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 3.3. *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 3.4. *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 $\pm\eta^n$. □

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 3.5. *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 3.6. *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:

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

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 3.7. *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 3.8. *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 3.9. *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 $\mathbb{Q}(\sqrt{m})$.*

3.3 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 $\mathfrak{2} = (2, \omega)$ for $\omega = \frac{1+\sqrt{-m}}{2}$ and some $m > 0$ with $-m \equiv 1 \bmod 8$. Then $\mathfrak{2}$ 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 3.10. *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 3.11. *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 3.11 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 3.12. *Let $k = \mathbb{Q}(\sqrt{m}\,)$ be a real quadratic number field, $\varepsilon > 1$ a unit in k, and $0 \neq n = |N\xi\,|$ 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!

3.4 Elements of small norms

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

Proposition 3.13. *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 3.13 one can easily show more:

Proposition 3.14. *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.*

Proposition 3.13 can be used to show that the ideal class group of $k = \mathbb{Q}(\sqrt{m})$ has non-trivial elements 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 3.12.

3.5 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	$\mathfrak{2}$	$\mathfrak{5}_1$	$\mathfrak{5}_2$	$\mathfrak{7}_1$	$\mathfrak{7}_2$
$1+\sqrt{m}$	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

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

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 $\mathfrak{2}^4\mathfrak{5}_1^{10}\mathfrak{7}_1^3\mathfrak{7}_2^3$. Since $\mathfrak{2}^2 = (2)$ and $\mathfrak{7}_1\mathfrak{7}_2 = (7)$, 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 $\mathfrak{5}_1^{10}$. But now $(59-\sqrt{m}\,)^5 = \mathfrak{2}^5\mathfrak{5}^{10}$ shows that

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

is an integer with ideal factorization $\mathfrak{2}^5$. 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

$$\begin{aligned}\gcd(152009690466841, 3431) &= 1,\\ \gcd(152009690466839, 3431) &= 3431,\end{aligned}$$

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 3.15. *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 $\varepsilon \equiv 0 - 3\sqrt{m} \equiv 3 \bmod \mathfrak{5}_1$ shows again that ε is not a square since 3 is not a square modulo $\mathfrak{5}_1$ 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 $\varepsilon \equiv 3 + \sqrt{m} \equiv 2 \bmod \mathfrak{7}_1$, and since 2 is not a cube modulo 7, it is not a cube modulo $\mathfrak{7}_1$ since $\mathcal{O}_K/\mathfrak{7}_1 \simeq \mathbb{Z}/7\mathbb{Z}$. 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
```

3.6 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

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

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

3.3 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.

3.4 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.

3.5 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 $K = \mathbb{Q}(\sqrt{m})$.

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

4. 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 3, 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.

4.1 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, $\mathfrak{ab} = (1 + \sqrt{-5})$ implies $\mathfrak{ab} \sim (1)$, 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 2.11, 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 4.1. *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 4.1 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 4.1, 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 4.2. *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 4.1 is the following observation:

Corollary 4.3. *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 $p^h = |\frac{x^2-my^2}{4}|$. □

Proof of Theorem 4.1. 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 $a \le \mu_K$, 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

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

Setting $\alpha_1 = r + \frac{1}{2}\sqrt{d}$ we find $\alpha_1 \in \mathfrak{a}$, $|N\alpha_1| \le \frac{1}{4}(a^2 - d) < a^2$, 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 $\le \mu_K$; since the norm decreases with each step, the algorithm terminates.

The proof of the inequality $|N\alpha_1| \le \frac{1}{4}(a^2 - d) < a^2$ is simple: if $d < 0$, then $|N\alpha_1| = |r^2 - \frac{d}{4}| \le \frac{a^2+|d|}{4} < 1$ 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 $\frac{1}{a}\alpha_1'\mathfrak{a} \subseteq \mathcal{O}_K \iff \alpha'\mathfrak{a} \subseteq (a) = \mathfrak{a}\mathfrak{a}' \iff (\alpha') \subseteq \mathfrak{a}'$. □

4.2 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 4.4. *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

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

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

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

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

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

Also, $\mathfrak{3}_2^3\mathfrak{7}_1 \sim (1)$ implies $\mathfrak{7}_1 \sim \mathfrak{3}_1^3 \sim \mathfrak{3}_1$. Next $\mathfrak{3}_1 \cdot \mathfrak{7}_2^2 \sim (1)$ implies $\mathfrak{7}_1^2 \sim \mathfrak{3}_1$, and we conclude that $\mathfrak{7}_1 \sim \mathfrak{7}_2 \sim (1)$. 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 $\mathfrak{2}$, and then $\frac{1}{2}\alpha^2 = \varepsilon$ will be a nontrivial unit. Let's do this now.

We have $(25 + \sqrt{478}) = \mathfrak{3}_1 \cdot \mathfrak{7}_2^2$, $(17 - \sqrt{478}) = \mathfrak{3}_1^3\mathfrak{7}_2$, as well as $\mathfrak{3}_1^2 = \frac{1}{2}(22 + \sqrt{478})^2$. Thus $\mathfrak{3}_1 \cdot \mathfrak{7}_2 = \left(\frac{2(17-\sqrt{478})}{(22+\sqrt{478})^2}\right)$, 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, $(24-\sqrt{478}\,) = \mathfrak{27}_2^2$ 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.

4.3 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 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 $d \not\equiv 7 \bmod 8$, 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 $3 \nmid h$. Thus $\mathfrak{b} = (\frac{r+s\sqrt{-d}}{2})$ 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 4.5. *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 4.6. *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. 4.5 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. □

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.

4.4 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 4.7. *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 4.8 (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 4.9. *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 automorphism 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. 4.7 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 4.7 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

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

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

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

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

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

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

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

4.8 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 $d = 3t^2 \pm 8$.

4.9 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 $d = 31$.

4.10 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".

4.11 Find an analog of Theorem 4.5 for the equation $y^2 + d = x^5$.

5. Binary Quadratic Forms

We are interested in the set of integral binary quadratic forms

$$Q = (A, B, C) = Ax^2 + Bxy + Cy^2,$$

where A, B, C are coprime integers. The integer $\Delta = B^2 - 4AC$ is called the discriminant of Q. We say that a form Q represents an integer n if there are integers x, y such that $n = Q(x, y)$. We say that Q represents n properly if there are coprime integers x, y such that $n = Q(x, y)$.

5.1 The Action of the Modular Group

In this introduction, we will concentrate on positive definite forms (those with negative discriminant) and present four different ways of looking at them.

Before we do this, we discuss the general concept of groups acting on sets.

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(!) $\{g \in G : gs = s\}$.

It is an easy exercise to check that if G acts on S from the left, then it also acts on S from the right via $ga = s^{-1}g$.

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 $g \cdot h = gh$.

Quadratic Forms

The "modular group" $\mathrm{SL}_2(\mathbb{Z})$ of 2×2-matrices with integral coefficients 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 $\Delta' = \Delta$. In order to prove that $\mathrm{SL}_2(\mathbb{Z})$ "acts" on forms of discriminant Δ we now have to verify $(Q|_M)|_N) = Q|_{MN}$ for $M, N \in \mathrm{SL}_2(\mathbb{Z})$. This is a simple if somewhat technical calculation; we can avoid the technicalities by using some linear algebra.

Linear Algebra

To every binary quadratic form $Q = (A, B, C)$ we associate the matrix $M_Q = \left(\begin{smallmatrix} A & B/2 \\ B/2 & C \end{smallmatrix}\right)$ (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 $Ax^2 + Bxy + Cy^2 = (x, y)M_Q\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Moreover we find that $\operatorname{disc} Q = B^2 - 4AC = -4 \det M_Q$. Another simple calculation shows that $M_{Q|_M} = M^t M_Q M$. This allows us to give a simple proof for the fact that $\operatorname{disc} Q|_M = \operatorname{disc} Q$:

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

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

We now prove the fundamental

Proposition 5.1. *If Q represents an integer 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; then $n = (x, y)M_Q\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. Since $M_{Q|_M} = M^t M_Q M$, we have $n = Q|_M(u, v) = (u, v)M^t M_Q M\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ 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. □

Thus Q and $Q|_M$ represent exactly the same numbers. In fact, they also represent the same integers properly: if $n = Q(x, y)$ with $\gcd(x, y) = 1$, then $n = Q|_M(u, v)$ for coprime integers u, v; for if $p \mid \gcd(u, v)$, then $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = M\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ would show that $p \mid \gcd(x, y)$.

Proposition 5.2. *If Q properly represents an integer n, then so does $Q|_M$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$.*

Let us call a form $Q = (A, B, C)$ *primitive* if $\gcd(A, B, C) = 1$. A similar argument as the one above shows

Lemma 5.3. *If Q is primitive, then so is $Q|_M$ for any $M \in \mathrm{SL}_2(\mathbb{Z})$.*

The Upper Half Plane

To every binary quadratic form $Q = (A, B, C)$ with *negative discriminant* Δ we associate the point $z_Q = \frac{-B+i\sqrt{|\Delta|}}{2A}$ in the upper half plane

$$\mathcal{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}.$$

Note that we can compute the coefficient C from z_Q by first reading off A, B and Δ, and then setting $C = \frac{B^2-\Delta}{4A}$. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathcal{H}$ 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. An easier way to prove this is via reduction to the action of $\mathrm{SL}_2(\mathbb{Z})$ on quadratic forms: simply observe that $z_{Q|_M} = M^{-1}z_Q$: this means that the action of $\mathrm{SL}_2(\mathbb{Z})$ on quadratic forms from the right corresponds to the action of $\mathrm{SL}_2(\mathbb{Z})$ on the upper half plane from the left.

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.

5.2 Reduction

In practice, the action of $\mathrm{SL}_2(\mathbb{Z})$ on binary quadratic forms is used to make the coefficients of a form Q "as small as possible".

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 $4AQ(x, y) = (2AX + BY)^2 - \Delta Y^2$.

We say that a quadratic form Q with negative discriminant is *reduced* if $-A < B \le A \le C$ or $0 \le B \le A = C$. Equivalently we may demand $|B| \le A \le C$, and $B > 0$ if we have equality $|B| = A$ or $A = C$.

Lemma 5.4. *If $Q = (A, B, C)$ is a reduced binary quadratic form with negative discriminant Δ, then $|B| \le A \le \sqrt{-\Delta/3}$ and $C \le \frac{1-\Delta}{4}$.*

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

From $4AC = B^2 - \Delta$ and the fact that $A > 0$ we get

$$C = \frac{B^2}{4A} - \frac{\Delta}{4A} \le \frac{A^2}{4A} - \frac{\Delta}{4A} = \frac{A}{4} - \frac{\Delta}{4A}.$$

As a function of A (assuming Δ to be constant), the expression on the right hand side is decreasing in the interval $[1, \sqrt{-\Delta}\,]$, hence attains its maximum at the boundary $A = 1$. This implies the claim. □

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 5.4, hence only finitely many B with $|B| \le A$. Finally, for each pair (A, B) there is at most one C because $\Delta = B^2 - 4AC$ is fixed.

Since discriminants satisfy $\Delta = B^2 - 4AC \equiv B^2 \equiv 0, 1 \bmod 4$, every discriminant has the form $\Delta = -4m$ or $\Delta = 1 - 4m$. The forms $Q_0 = (1, 0, m)$ of discriminant $\Delta = -4m$ and $Q_0 = (1, 1, m)$ of discriminant $\Delta = 1 - 4m$ are reduced; they are called the principal form of discriminant Δ. We have $h(\Delta) = 1$ if and only if Q_0 is the only reduced form of discriminant Δ.

The number of reduced forms of discriminant $\Delta < 0$ is denoted by $h(\Delta)$ and is called the class number. As an example, let us compute the class number $h(-20)$. We know that $0 < A < \sqrt{20/3} < 3$, hence $A \in \{1, 2\}$. Moreover, $-20 = B^2 - 4AC$ shows that B must be even. The following table then lists all possibilities:

A	B	forms
1	0	$x^2 + 5y^2$
2	0	$- - -$
	2	$2x^2 + 2xy + 3y^2$

Thus there are only two reduced forms, and $h(-20) = 2$.

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 fundamental 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 $\Delta < 0$.

Here is a more complex example: let us explicitly compute the class number for $\Delta = -4 \cdot 65$. We know that $|A| \leq \sqrt{-\Delta/3} < 10$. Thus we have $-9 \leq A < B \leq 9 \leq C$ 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 $B = 0$ and $B = -2$ are impossible, so we must have $B = 2$. Now $65 = 2C - 1$ gives $C = 33$, and we get the form $(2, 2, 33)$.
- $A = 3$: clearly $b \neq 0$; $b = \pm 1$ 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 $65 = 9C - b^2$ for integers b with $|b| \leq 4$ is only solvable for $b = 4$, leading to the form $(9, 8, 9)$.

Thus the set of reduced forms of discriminant $-4 \cdot 65$ is

$$\{(1,0,65),(2,2,33),(3,\pm 2,22),(5,0,13),(6,\pm 2,11),(9,8,9)\}.$$

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

$$Q|_M = (A, B+2Ab, Ab^2 + Bb + C). \tag{5.1}$$

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 $Q|_S(x,y) = A(-y)^2 + B(-xy) + Cx^2$:

$$Q|_S = (C, -B, A). \tag{5.2}$$

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 (5.1)).
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 $(C'', -B'', A'')$.

This algorithm terminates: in fact, every time the algorithm runs through step 1, the absolute value of the middle coefficient is decreased by at least 1; this clearly can happen only finitely often.

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.

Theorem 5.5. *Every primitive positive definite binary quadratic form Q is equivalent to a unique reduced form.*

For the proof that there is a unique such form we need another nice property of reduced forms, which will turn out to be an important tool in various proofs:

Lemma 5.6. *If $Q = (A, B, C)$ is reduced, then the three smallest integers properly represented by Q are A, C, and $A - |B| + C$.*

Proof. Clearly these integers are represented by Q since $Q(1,0) = A$, $Q(0,1) = C$ and $Q(1, \pm 1) = A \pm B + C$.

In order to show that these are the smallest integers represented by Q we have to show that $Q(x,y) \geq A - |B| + C$ for integers x, y with $xy > 1$. We now distinguish three cases:

- $|x| = |y|$. Then $|x| = |y| = 1$ since the representation is proper, and we find $Q(x,y) \geq A - |B| + C$.
- $|x| > |y|$. Then

$$\begin{aligned} Q(x,y) &\geq Ax^2 - |B||xy| + Cy^2 > (A - |B|)|xy| + Cy^2 \\ &\geq (A - |B| + C)y^2 > A - |B| + C. \end{aligned}$$

- $|x| < |y|$. Then $Q(x,y) \geq (A - |B| + C)x^2 > A - |B| + C$.

□

Note that these three integers A, C, and $A - |B| + C$ need not be distinct: if $Q = (1,1,1)$, then actually $A = C = A - |B| + C = 1$.

Corollary 5.7. *A (positive definite) quadratic form representing 1 is equivalent to the principal form.*

Proof. Let Q be such a quadratic form. Then Q is equivalent to some reduced form Q', which also represents 1. Since 1 is the smallest natural number represented by Q', Lemma 5.6 implies that $Q' = (A, B, C)$ with $A = 1$. Since Q' is reduced, we must have $|B| \leq |A| = 1$, hence $Q' = (1, 0, C)$ or $Q' = (1, 1, C)$. But these are exactly the principal forms. □

Proof of Thm. 5.5. We have to show that if $Q = (A, B, C)$ and $Q' = (A', B', C')$ are reduced forms with $Q \sim Q'$, then $Q = Q'$.

First we observe that the smallest natural number represented by Q and Q' is A and A', respectively. Since $Q \sim Q'$, they represent the same integers, hence we must have $A = A'$. Note that $C \geq A$ since Q is reduced; we now distinguish some cases.

1. $C > A$. Since $A = Q(\pm 1, 0)$ is represented exactly twice by Q, it is also represented exactly twice by Q', hence $C' = Q'(0, \pm 1) > A' = A$. Now C is the second smallest integer represented by Q, and therefore also by Q'. Since Q and Q' represent the same integers, we must have $C = C'$. Since $\operatorname{disc} Q = \operatorname{disc} Q'$, we see that $|B| = |B'|$. If we had $B' = -B$, then $(A, B, C) = Q \sim Q' = (A, -B, C)$. Assume that $Q' = Q|_S$ for $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Then $A = A' = Ar^2 + Brt + Ct^2$, and since $C > A$, the only solutions of this equation are $r = \pm 1$, $t = 0$. Thus $S = \left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$ or $S = \left(\begin{smallmatrix} -1 & s \\ 0 & -1 \end{smallmatrix}\right)$, hence $-B = B' = 2As + B$, or $As = -B$. Since $|B| \leq A$,

we must have $s = 0$ (and then $B = 0 = -B = B'$) or $s = 1$ (and then $B = -A$, which contradicts the assumption that Q is reduced). Thus we have $Q = Q'$ in all cases considered here.

2. $C = A$. Then $A = Q(\pm 1, 0) = Q(0, \pm 1)$, hence A is represented at least four times by Q, hence also by Q'. But this implies $C' = A$ and therefore $C = C'$. As above this implies $B' = \pm B$. But since $(A, B, A) \sim (A, B', A)$ are reduced, B and B' must be positive, and we get $Q = Q'$.

The proof is now complete. □

The Upper Half Plane

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.

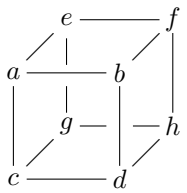
Linear Algebra

I do not know if the fact that some form Q is reduced can be read off from basic properties of the matrix M_Q in a simple form. Is there a connection between reduced forms and the eigenvalues or the characteristic polynomial of M_Q?

5.3 The Class Group

Recall that two quadratic forms Q, Q' are called equivalent if there is a matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $Q' = Q|_M$. We know that equivalent forms have the same discriminant. Now consider the equivalence classes $[Q]$ of primitive positive definite forms. We have seen that each such class contains a unique reduced form, and that the number $h(\Delta)$ of classes is finite. In this section we will show how to make the set $\mathrm{Cl}(\Delta)$ of quadratic forms with discriminant Δ (positive definite if $\Delta < 0$) into a group; in the next section we will show that $\mathrm{Cl}(\Delta) \simeq \mathrm{Cl}(K)$ if $\Delta < 0$ is a fundamental discriminant, i.e., if $\Delta = \operatorname{disc} K$ for some quadratic number field.

Giving $\mathrm{Cl}(\Delta)$ the structure of a group is not a trivial thing to do. The method that is probably the easiest to memorize is via Bhargava's cubes. This method uses cubes of integers such as



to construct a triple of quadratic forms in the following way. Each such cube can be sliced in three different ways, producing three pairs of 2×2-matrices (front-back, left-right, up-down):

$$\begin{aligned} FB \qquad & M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad & N_1 = \begin{pmatrix} e & f \\ g & h \end{pmatrix}, \\ LR \qquad & M_2 = \begin{pmatrix} a & c \\ e & g \end{pmatrix}, \qquad & N_2 = \begin{pmatrix} b & d \\ f & h \end{pmatrix}, \\ UD \qquad & M_3 = \begin{pmatrix} a & e \\ b & f \end{pmatrix}, \qquad & N_3 = \begin{pmatrix} c & g \\ d & h \end{pmatrix}. \end{aligned}$$

To each cube $\mathcal{A}$ we can associate three binary quadratic forms $Q_i = Q_i^{\mathcal{A}}$ by putting[1]

$$Q_i(x, y) = -\det(M_i x + N_i y).$$

This way we find

$$\begin{aligned} Q_1(x, y) &= (bc - ad)x^2 + (bg + cf - ah - de)xy + (fg - eh)y^2, \\ Q_2(x, y) &= (ce - ag)x^2 + (cf + de - ah - bg)xy + (df - bh)y^2, \\ Q_3(x, y) &= (be - af)x^2 + (bg + de - ah - cf)xy + (dg - ch)y^2. \end{aligned}$$

Setting $Q_i = (A_i, B_i, C_i)$ we find that in the FB-slicing we have $A_1 = -\det F$ and $C_1 = -\det B$, where F and B denote the matrices forming the front and the back face of the cube. Similarly we have $A_2 = -\det L$ and $C_2 = -\det R$ in the LR-slicing. The matrices F and L have the edge ac in common; the diagonal matrix D_{FB} satisfies $\frac{1}{2}(B_1 + B_2) = -\det D_{FB}$.

A simple calculation shows that $\operatorname{disc} Q_1 = \operatorname{disc} Q_2 = \operatorname{disc} Q_3$; thus we can define[2]

$$\begin{aligned} \operatorname{disc}(A) = {} & a^2h^2 + b^2g^2 + c^2f^2 + d^2e^2 \\ & - 2(abgh + cdef + acfh + bdeg + aedh + bfcg) + 4(adfg + bceh). \end{aligned}$$

We now introduce an action of $\mathrm{SL}_2(\mathbb{Z})$ on a cube $\mathcal{A}$ (or, rather, on its FB-slicing) as follows: for $S = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ we replace the cube $\mathcal{A}$ defined by the pair of matrices (M_1, N_1) by the cube $\mathcal{A}|_S$ defined by $(rM_1 + tN_1, sM_1 + uN_1)$. Adding the back to the front face, for example, means applying $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ to the cube.

Lemma 5.8. *Let $\mathcal{A}$ be a cube, $M \in \mathrm{SL}_2(\mathbb{Z})$, and let $\mathcal{A}' = \mathcal{A}|_M$ be the cube we get by letting M act on $\mathcal{A}$; then $\operatorname{disc} \mathcal{A}' = \operatorname{disc} \mathcal{A}$. If the associated quadratic forms are denoted by Q_i and Q_i', then $Q_1' = Q_1|_M$, $Q_2' = Q_2$, and $Q_3' = Q_3$.*

[1] The following formulas differ from those found in Bhargava's work.

[2] This discriminant is what Cayley had called the hyperdeterminant of the $2 \times 2 \times 2$-hypermatrix defined by the cube $\mathcal{A}$.

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.

Since S acts via elementary row and column operations on the matrices defining Q_2 and Q_3, we immediately see that $Q_2' = Q_2$ and $Q_3' = Q_3$. □

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 $\det(M_ix + N_iy)$.

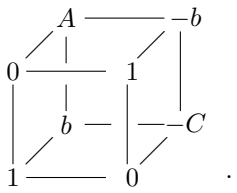
What have we achieved now? We know that if Q_1, Q_2, Q_3 are quadratic forms attached to some cube $\mathcal{A}$ and if $M \in \mathrm{SL}_2(\mathbb{Z})$, then there is a cube with associated quadratic forms $Q_1|_M$, Q_2, Q_3. This shows that we cannot hope for a composition law on quadratic forms; what we should be looking for is a composition law on equivalence classes of quadratic forms.

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 $\mathcal{A}$ 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.

Lemma 5.9. *The inverse of the class $[Q]$, where $Q = (A,B,C)$, is the class $[-Q]$, where $-Q = (A,-B,C)$.*

Proof. If $\Delta = 4m$, put $B = 2b$ and consider the cube



.

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. □

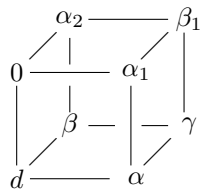
Since composition is clearly commutative (if $\mathcal{A}$ is a cube giving Q_1, Q_2, Q_3, a suitable reflection of $\mathcal{A}$ produces a cube that gives Q_2, Q_1, Q_3), all that remains is to check associativity. This is rather technical if you stick to forms; we will instead use the map sending forms to modules.

We also have to show that two primitive forms of discriminant Δ can be composed, i.e., that there is a cube giving rise to these two forms and their composed forms. This is done as follows.

We have seen how to attach three binary quadratic forms of the same discriminant to a cube A. Now we show that, conversely, to each pair of primitive binary quadratic forms of the same discriminant Δ we can find a cube A giving rise to these forms.

Proposition 5.10. *Given two primitive forms Q_1, Q_2 of discriminant Δ, there exists a primitive cube $\mathcal{A}$ such that $Q_1^{\mathcal{A}} = Q_1$ and $Q_2^{\mathcal{A}} = Q_2$.*

Proof. The proof we will give goes back to Shanks. The basic idea is the following: write $Q_i = (A_i, B_i, C_i)$ and set $B = \frac{1}{2}(B_1 + B_2)$; note that B is an integer since $B_1 \equiv B_2 \equiv \Delta \bmod 2$. Let $d = \gcd(A_1, A_2, B)$, write $A_1 = d\alpha_1$, $A_2 = d\alpha_2$ and $B = d\beta_1$, and form the cube



Then the determinants of the front face, the left face, and the diagonal already have the right values, namely $-A_1$, $-A_2$, and $-B$. Also observe that we have $A_3 = \alpha_1\alpha_2 = A_1A_2/d^2$. The values of α, β and γ can now be determined from the three equations

$$\left.\begin{array}{l}\beta_1\alpha - \alpha_1\gamma = C_2,\\ \alpha_2\alpha - \alpha_1\beta = (B_2 - B_1)/2,\\ \beta_1\beta - \alpha_2\gamma = C_1.\end{array}\right\} \tag{5.3}$$

We first have to show that these equations are consistent; to this end, multiply the first and the third equation by α_2 and α_1, respectively, and subtract them from each other; this gives

$$\beta_1(\alpha_1\beta - \alpha_2\alpha) = \alpha_2 C_2 - \alpha_1 C_1. \tag{5.4}$$

But Q_1 and Q_2 have the same discriminant $\Delta = B_1^2 - 4A_1C_1 = B_2^2 - 4A_2C_2$; this equality implies

$$\frac{B_2 + B_1}{2}\frac{B_2 - B_1}{2} = A_2C_2 - A_1C_1. \tag{5.5}$$

Dividing through by d shows that the right hand side of (5.4) equals $\beta_1(B_2 - B_1)/2$; if $\beta_1 \neq 0$, then dividing through by β_1 gives the second equation. If $\beta_1 = 0$, on the other hand, it is immediately clear that the first and the third equation in (5.3) are equivalent since $\alpha_2 C_2 = \alpha_1 C_1$ in this case.

We have seen that if we can manage to solve two out of these three equations, then we have won. In the case where $\gcd(\alpha_1, \alpha_2) = 1$, this can be easily done: solve the equation $\alpha_1 x - \alpha_2 y = 1$ using the Euclidean algorithm, then then put $\beta = \frac{B_2 - B_1}{2}x$ and $\alpha = \frac{B_2 - B_1}{2}y$. Then determine γ from either the first or the last equation. Now $\alpha_1\gamma = C_2 + \beta_1\alpha$ and $\alpha_2\gamma = C_1 + \beta_1\alpha$ are both integers; thus the denominator of γ must divide both α_1 and α_2, hence divides $\gcd(\alpha_1, \alpha_2) = 1$. Thus γ is an integer.

The same method works if $\gcd(\alpha_1, \beta_1) = 1$ or $\gcd(\alpha_2, \beta_1) = 1$: in these cases use Euclid to solve the first and the third equation, respectively, and then continue as above.

This leaves us with the case where $\gcd(\alpha_1, \alpha_2)$, $\gcd(\alpha_1, \beta_1)$ and $\gcd(\alpha_2, \beta_1)$ are all nontrivial. This can indeed happen, say if $\alpha_1 = pq$, $\alpha_2 = pr$ and $\beta_1 = qr$ for distinct primes p, q, r.

In this case, recall the equation (5.5), which in our case reads

$$\beta_1\frac{B_2 - B_1}{2} = \alpha_2 C_2 - \alpha_1 C_1.$$

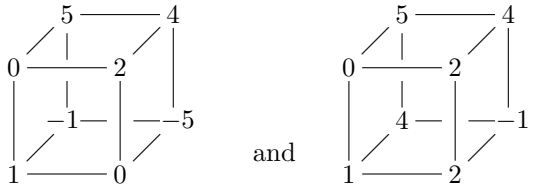
This equation immediately shows that $\gcd(\alpha_1, \beta_1) \mid \alpha_2 C_2$; but since $\gcd(\alpha_1, \beta_1)$ and α_2 are coprime, this implies $\gcd(\alpha_1, \beta_1) \mid C_2$. Thus we see that the first equation in (5.3), namely $\beta_1\alpha - \alpha_1\gamma = C_2$, has an integral solution. The other two equations then show that $\alpha_1\beta$ and $\beta_1\beta$ are integers, which implies that the denominator of β divides $H = \gcd(\alpha_1, \beta_1)$. Write $\beta = \frac{h}{H}$ and determine an integer r such that $r\alpha_2 \equiv h \bmod H$ (this can be done since $\gcd(H, \alpha_2) = 1$). If we add $\frac{r}{H}$ of the top matrix $\left(\begin{smallmatrix}0 & \alpha_1\\ \alpha_2 & \beta\end{smallmatrix}\right)$ of our cube to the matrix $\left(\begin{smallmatrix}G & \alpha\\ \beta & \gamma\end{smallmatrix}\right)$ at the bottom, then G will remain invariant, α and γ will be changed by integers (since α_1 and β_1 are multiples of H), and β will be replaced by $\beta - \frac{r}{H}\alpha_2 = \frac{h - r\alpha_2}{H}$, which is an integer. Since this transformation

does not change Q_1 and Q_2, we have finally found a cube $\mathcal{A}$ with the desired properties. □

Example. Consider the two forms $Q_1 = (2, 2, 21)$ and $Q_2 = (5, 6, 10)$ of discriminant $-4 \cdot 41$. The system (5.3) of equations now becomes

$$\begin{aligned} 4\alpha - 2\gamma &= 10, \\ 5\alpha - 2\beta &= 2, \\ 4\beta - 5\gamma &= 21. \end{aligned}$$

The solutions $(\alpha, \beta, \gamma) = (0, -1, -5)$ and $(2, 4, -1)$ give the cubes



Computing the associated forms gives Q_1 and Q_2 (of course), as well as

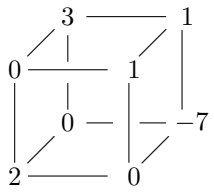
$$Q_3 = (10, -6, 5) \quad \text{and} \quad Q_3' = (10, 14, 9).$$

In particular we have $[Q_1] + [Q_2] + [Q_3] = 0$, or $[Q_1] + [Q_2] = [(10, 6, 5)] = [(5, -6, 10)]$.

Note that you can get the second cube from the first by adding the top face to the bottom. This proves in particular that $Q_3 \sim Q_3'$.

Of course we know that there must be infinitely many solutions since we can make $\mathrm{SL}_2(\mathbb{Z})$ act on the cube in such a way that the two forms Q_1 and Q_2 are not changed.

Example. Consider the forms $Q_1 = (2, 2, 21)$ and $Q_2 = (6, 2, 7)$ of discriminant $-4 \cdot 41$. Here $G = \gcd(A_1, A_2, B) = 2$ and $\gcd(\alpha_1, \beta_1) = 1$. We get the system of equations $\alpha - \gamma = 7$ and $\beta - 3\gamma = 21$, which has the solution $\gamma = -7$ and $\alpha = \beta = 0$. This gives us the cube



with the associated forms Q_1, Q_2 and $Q_3 = (3, -2, 14)$, and we find $[Q_1] + [Q_2] = [(3, 2, 14)]$.

Let us now derive a few explicit formulas for composition. Assume that $\gcd(A_1, A_2, B) = 1$. Then we have seen above that $A_3 = A_1 A_2$. Using $A_2\alpha = A_1\beta + \frac{B_2 - B_1}{2}$, we now find

$$\begin{aligned} B_3 &= A_1\beta + A_2\alpha - B \\ &= 2A_1\beta - \frac{B_2+B_1}{2} + \frac{B_2-B_1}{2} \\ &= 2A_1\beta - B_1 \\ &\equiv -B_1 \bmod 2A_1, \end{aligned}$$

and similarly we can prove

$$B_3 \equiv -B_2 \bmod 2A_2.$$

Finally we find

$$\begin{aligned} B_3B &= (A_1\beta + A_2\alpha - B)B \\ &= -B^2 + A_1B\beta + A_2B\alpha \\ &= -B^2 + A_1(C_1 + A_2\gamma) + A_2(C_2 + A_1\gamma) \\ &= -B^2 + A_1C_1 + A_2C_2 + 2A_1A_2\gamma \\ &\equiv -B^2 + A_1C_1 + A_2C_2 \\ &= -B^2 + 2A_1C_1 + B\frac{B_2-B_1}{2} \\ &= -\frac{\Delta + B_1B_2}{2} \bmod 2A_1A_2 \end{aligned}$$

Thus B_3 satisfies the congruences

$$\begin{aligned} A_2B_3 &\equiv -B_1 \bmod 2A_1A_2, \\ A_1B_3 &\equiv -B_2 \bmod 2A_1A_2, \\ BB_3 &\equiv -\frac{\Delta + B_1B_2}{2} \bmod 2A_1A_2. \end{aligned}$$

Since the coefficients of B_3 on the left hand side have greatest common divisor 1, these congruences determine B_3 uniquely modulo $2A_1A_2 = 2A_3$. In fact, if we write $\lambda A_1 + \mu A_2 + \nu B = 1$, then clearly

$$\begin{aligned} B_3 &= (\lambda A_1 + \mu A_2 + \nu B)B_3 \\ &\equiv -\lambda B_2 - \mu B_1 - \nu\frac{\Delta + B_1B_2}{2} \bmod 2A_1A_2. \end{aligned}$$

But now we observe that the residue class $B_3 \bmod 2A_3$ determines the equivalence class of Q_3; in fact we find

Corollary 5.11. *Let $Q_i = (A_i, B_i, C_i)$ be primitive binary quadratic forms with discriminant Δ and $\gcd(A_1, A_2, B) = 1$, where $B = \frac{B_1+B_2}{2}$. Then $[Q_1]+[Q_2]+[Q_3] = [Q_0]$ for $Q_3 = (A_3, B_3, C_3)$, where $A_3 = A_1A_2$, B_3 is determined by the congruences above, and $C_3 = \frac{B_3^2-\Delta}{4A_3}$.*

This is what some authors call "Dirichlet composition".

In order to show that we do indeed get a composition on $\mathrm{Cl}(\Delta)$, we have to show that different choices of the cube $\mathcal{A}$ in Prop. 5.10 give equivalent forms:

Lemma 5.12. *Let $\mathcal{A}$ and $\mathcal{A}'$ be primitive cubes with $Q_1^{\mathcal{A}} = Q_1^{\mathcal{A}'}$ and $Q_2^{\mathcal{A}} = Q_2^{\mathcal{A}'}$. Then $Q_3^{\mathcal{A}} \sim Q_3^{\mathcal{A}'}$.*

Proof. This is best done via "Gauss's Lemma", given in his Disquisitiones. Details later. □

I would like to conclude this section with the following problems. Let us call a primitive cube $\mathcal{A}$ reduced if the three associated quadratic forms $Q_i^{\mathcal{A}}$ are all reduced. Is there a simple way of characterizing reduced cubes in terms of the integers a, b, ..., h? Can we at least find good bounds on these integers? If a cube $\mathcal{A}$ is not reduced, is there a simple algorithm for reducing $\mathcal{A}$ directly? Finally: the definition of composition via primitive cubes works over any UFD; is there a sufficiently simple proof that composition is associative over general UFDs? In special cases like $\mathbb{Z}$ or $K[T]$ (see the next chapter) this can be done via "Dirichlet composition".

5.4 Orders and Modules

Modules

Finally consider the order $\mathcal{O} = [1, \omega_\Delta]$ for

$$\omega_\Delta = \begin{cases} \frac{1}{2}\sqrt{\Delta} & \text{if } \Delta \equiv 0 \bmod 4, \\ \frac{1+\sqrt{\Delta}}{2} & \text{if } \Delta \equiv 1 \bmod 4. \end{cases}$$

Δ is called a fundamental discriminant if $\mathcal{O}$ is the full ring of integers in $\mathbb{Q}(\sqrt{\Delta})$, that is, if $\Delta/4$ or Δ is squarefree.

To every quadratic form $Q = (A, B, C)$ of discriminant $\Delta < 0$ we now can attach a $\mathbb{Z}$-module $i(Q)$ in the order $\mathcal{O} = [1, \omega_\Delta]$ by putting

$$i(Q) = [A, \tfrac{-B+\sqrt{\Delta}}{2}].$$

How can we make $\mathrm{SL}_2(\mathbb{Z})$ act on full $\mathbb{Z}$-modules $[\alpha, \beta]$? The naive idea of setting $[\alpha, \beta]|_M = [r\alpha + s\beta, t\alpha + u\beta]$ for $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ does not really work, since it is an easy matter to show that M just changes the basis of the module, in other words: $[\alpha, \beta]|_M = [\alpha, \beta]$, i.e., $\mathrm{SL}_2(\mathbb{Z})$ would act trivially.

Now consider a quadratic form Q, as well as $Q' = Q|_M = (A', B', C')$; then $\iota(Q') = [A', \frac{-B'+\sqrt{\Delta}}{2}] = A'[1, z_{Q'}]$, where $z_{Q'}$ is the point in the upper half plane attached to Q'. We have already seen that $z_{Q'} = M^{-1} z_Q$, hence we find $\iota(Q') = A'[1, M^{-1} z_Q]$.

Let $Q = (A, B, C)$ be a positive definite quadratic form with discriminant Δ, and consider the $\mathbb{Z}$-module $\iota(Q) = I_Q$. If Δ is a fundamental discriminant, i.e., $\Delta = \operatorname{disc} K$ for some quadratic field $K = \mathbb{Q}(\sqrt{m}\,)$, then I_Q is an ideal in $\mathcal{O}_K$. In other words, I_Q is a $\mathbb{Z}$-module, but in general not a $\mathcal{O}_K$-module. It turns out that we can always find an order $\mathcal{O}$ (a subring of $\mathcal{O}_K$ containing $\mathbb{Z}$ and with quotient field K) such that I_Q is an $\mathcal{O}$-module, that is, an ideal in $\mathcal{O}$.

If $\mathcal{O}$ is not the maximal order $\mathcal{O}_K$, then $\mathcal{O}$ is not a Dedekind ring, and there won't be unique factorization into prime ideals. Nevertheless it turns out that we can at least imitate the definition of the class group; in fact, the class group $\mathrm{Cl}(\mathcal{O})$ of this order will be isomorphic to the class group $\mathrm{Cl}(\Delta)$ of binary quadratic forms.

Stabilizer Rings

Let M be a $\mathbb{Z}$-module in $\mathcal{O}_K$. An element $\gamma \in \mathcal{O}_K$ is called a stabilizer of M if $\gamma M \subseteq M$. Let $\mathcal{O}_M$ denote the set of all stabilizers of M.

Lemma 5.13. *Let M be a module; then $\mathcal{O}_M$ is a ring with $\mathbb{Z} \subseteq \mathcal{O}_M \subseteq \mathcal{O}_K$.*

Proof. It is easy to see that all integers in $\mathbb{Z}$ are stabilizers, and that $\gamma, \gamma' \in \mathcal{O}_M$ implies $\gamma + \gamma' \in \mathcal{O}_M$ and $\gamma\gamma' \in \mathcal{O}_M$. □

An order of $\mathcal{O}_K$ is a subring of $\mathcal{O}_K$ containing $\mathbb{Z}$, which has rank 2 as a $\mathbb{Z}$-module. For example, the $\mathbb{Z}$-module $[1, g\omega]$ is an order for any nonzero $g \in \mathbb{Z}$. The maximal order is $[1, \omega] = \mathcal{O}_K$.

Lemma 5.14. *If M is a full $\mathbb{Z}$-module in $\mathcal{O}_K$, then $\mathcal{O}_M$ is an order.*

Proof. Assume that $M = [\alpha, \beta]$ is full; then every element in K can be written as a $\mathbb{Q}$-linear combination of α and β. In particular we have

$$\omega\alpha = r\alpha + s\beta,$$
$$\omega\beta = t\alpha + u\beta$$

for $r, s, t, u \in \mathbb{Q}$. Let d be the lowest common multiple of the denominators of these rational numbers; then $dr, ds, dt, du \in \mathbb{Z}$, and this implies that $d\omega \in \mathcal{O}_M$. Since we already know that $\mathbb{Z} \subseteq \mathcal{O}_M$, this proves that $\mathcal{O}_M$ is an order. □

Modules in Orders

Let K be a quadratic number field and $\mathcal{O} \subseteq \mathcal{O}_K$ an order (a subring containing $\mathbb{Z}$ with quotient field K). With $\mathcal{O} = \mathbb{Z}\alpha + \mathbb{Z}\beta$, the discriminant of $\mathcal{O}$ is defined to be $\operatorname{disc} \mathcal{O} = \Delta = (\alpha\beta' - \alpha'\beta)^2$; this does not depend on the choice of the basis. Every order has the form $[1, f\omega]$ for some integer $f \in \mathbb{N}$ called

the conductor of $\mathcal{O}$ (here $Z \oplus \mathbb{Z}\omega = \mathcal{O}_K$ is the maximal order); comparing discriminants we see that $\operatorname{disc}\mathcal{O} = f^2 \operatorname{disc} K$, where $\operatorname{disc} K = \operatorname{disc}\mathcal{O}_K$.

A full $\mathbb{Z}$-module M is a subset $M = \mathbb{Z}\alpha + \mathbb{Z}\beta$ for $\alpha, \beta \in K$ such that $\frac{\beta}{\alpha} \in K \setminus \mathbb{Q}$. The product MN of two full $\mathbb{Z}$-modules is the full $\mathbb{Z}$-module $MN = \{mn : m \in M, n \in N\}$.

A $\mathbb{Z}$-module M is called an $\mathcal{O}$-ideal if $\mathcal{O}M = M$. An $\mathcal{O}$-ideal M is called invertible if there is an $\mathcal{O}$-ideal N with $MN = \mathcal{O}$.

Lemma 5.15. *Inverses are unique: if $MN = MN_1 = \mathcal{O}$, then $N = N_1$.*

Proof. We have $N = N\mathcal{O} = N(MN_1) = (NM)N_1 = \mathcal{O}N_1 = N_1$. □

Thus we can write $N = M^{-1}$ if $MN = \mathcal{O}$. It is now easily checked that the set of invertible $\mathcal{O}$-ideal forms a group $I_{\mathcal{O}}$ with respect to multiplication.

Lemma 5.16. *Let M be an invertible $\mathcal{O}$-ideal. Then $\mathcal{O}_M = \mathcal{O}$.*

Proof. Since M is an $\mathcal{O}$-ideal, we have $\mathcal{O}M \subseteq M$ and therefore $\mathcal{O} \subseteq \mathcal{O}_M$. Conversely, assume that $\alpha \in \mathcal{O}_M$, that is, $\alpha M \subseteq M$, and that $MN = \mathcal{O}$. Then $\alpha = \alpha \cdot 1 \in \alpha\mathcal{O} = \alpha MN \subseteq MN = \mathcal{O}$. □

Now let $\alpha, \beta \in K$ and $M = \mathbb{Z}\alpha + \mathbb{Z}\beta$ be an $\mathcal{O}$-module. Since α and β are linearly independent over $\mathbb{Q}$, the element $\gamma = \frac{\beta}{\alpha}$ is a quadratic irrationality, and there exist integers $a, b, c \in \mathbb{Z}$ with $a\gamma^2 + b\gamma + c = 0$, where we may assume that $\gcd(a, b, c) = 1$. Since $M = \alpha[1, \gamma]$, we see that $(a\gamma)\gamma = -b\gamma - c \in [1, \gamma]$, which in turn implies that $\mathbb{Z}[a\gamma] \subseteq \mathcal{O}$. Now observe that $\gamma' = -\gamma - \frac{b}{a}$, $\gamma\gamma' = \frac{c}{a}$; using these relations we find

$$\begin{aligned} MM' &= \alpha[1, \gamma] \cdot \alpha'[1, \gamma'] = N(\alpha)[1, \gamma, \gamma', \gamma\gamma'] \\ &= N(\alpha)[1, \gamma, \tfrac{b}{a}, \tfrac{c}{a}] = \tfrac{N\alpha}{a}[a, a\gamma, b, c] = \tfrac{N\alpha}{a}\ [1, a\gamma], \end{aligned}$$

where, in the last step, we have used $\gcd(a, b, c) = 1$. Thus we find that M is an invertible $\mathbb{Z}[a\gamma]$-ideal with inverse $M^{-1} = \frac{a}{N\alpha}M'$. Lemma 5.16 now implies that $\mathcal{O} = \mathbb{Z}[a\gamma]$, i.e., M is an invertible $\mathcal{O}$-ideal if and only if $\mathcal{O} = \mathbb{Z}[a\gamma]$. This is equivalent to $\operatorname{disc}\mathcal{O} = \operatorname{disc}\mathbb{Z}[a\gamma] = b^2 - 4ac$, as can be seen easily from $\gamma = \frac{-b+\sqrt{b^2-4ac}}{2a}$. We have shown:

Lemma 5.17. *Let M be an $\mathcal{O}$-module. Then M is an invertible $\mathcal{O}$-ideal if and only if one of the following equivalent conditions is satisfied:*

1. *There is a $\mathcal{O}$-module N with $MN = \mathcal{O}$;*
2. *$\mathcal{O} = \mathcal{O}_M$;*
3. *$\operatorname{disc}\mathcal{O} = b^2 - 4ac$ with a, b, c as above.*

In the following, we will often write $M = (\alpha, a, b)$ if M is given as above.

A nonzero $\mathcal{O}$-ideal M is called principal if $M = \alpha\mathcal{O}$ for some $\alpha \in K^\times$. Principal ideals are invertible: we easily find that the inverse of $M = (\alpha, a, b)$ is given by $M^{-1} = (\frac{a}{\alpha}, a, -b)$. Thus principal $\mathcal{O}$-ideals form a subgroup $P_{\mathcal{O}}$

of the group $I_{\mathcal{O}}$ of invertible ideals in $\mathcal{O}$; the quotient group $\mathrm{Cl}(\mathcal{O}) = I_{\mathcal{O}}/P_{\mathcal{O}}$ is called the class group of $\mathcal{O}$, and its cardinality the class number $h(\mathcal{O})$ of $\mathcal{O}$. If $\mathcal{O} = \mathcal{O}_K$, every $\mathcal{O}_K$-ideal is invertible, and $\mathrm{Cl}(\mathcal{O}_K)$ coincides with the ideal class group of K we have studied before.

For positive discriminants $\Delta > 0$, we do in general not get a bijection between $\mathrm{Cl}(\mathcal{O})$ and $\mathrm{Cl}(\Delta)$. This can be remedied by slightly modifying the definition of $\mathrm{Cl}(\mathcal{O})$. Call an invertible $\mathcal{O}$-ideal M principal in the strict sense if $M = \alpha\mathcal{O}$ with $N\alpha > 0$, and let $P_{\mathcal{O}}^+$ denote the group of principal $\mathcal{O}$-ideals in the strict sense. Then $\mathrm{Cl}^+(\mathcal{O}) = I_{\mathcal{O}}/P_{\mathcal{O}}^+$ is called the class group in the strict sense. If $\Delta < 0$, the condition $N\alpha > 0$ is automatically satisfied, and we always have $\mathrm{Cl}^+(\mathcal{O}) = \mathrm{Cl}(\mathcal{O})$ in this case. If $\Delta > 0$, however, we have $h^+(\mathcal{O}) = h(\mathcal{O})$ or $h^+(\mathcal{O}) = 2h(\mathcal{O})$, and both cases occur.

For orders $\mathcal{O}$ of discriminant $\Delta > 0$ it can be shown that $\mathrm{Cl}^+(\mathcal{O}) \simeq \mathrm{Cl}(\Delta)$.

5.5 The Bijection

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 $\Delta = B^2 - 4AC = f^2 \operatorname{disc} K$ to integral ideals $\mathfrak{a}_Q$ in $\mathcal{O}$ is easily defined: we just put

$$i(Q) = [A, \tfrac{-B+\sqrt{\Delta}}{2}]. \tag{5.6}$$

This is an invertible $\mathcal{O}$-ideal, where $\mathcal{O} = [1, f\omega]$, 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 $i(Q) = A[1, \gamma]$, where $\gamma = \frac{-B+\sqrt{\Delta}}{2A} = z_Q$ is the point in the upper half plane associated to Q. But with $Q' = (A', B', C') = Q|_M$ we now find $i(Q|_M) = A'[1, \gamma'] = A'[1, M^{-1}\gamma]$. It remains to show that the two ideals $A[1, \gamma]$ and $A'[1, M^{-1}\gamma]$ are equivalent.

Now let $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$; then $\{\gamma, 1\}$ is the basis of a $\mathbb{Z}$-module $[1, \gamma]$ in K if and only if $\{r\gamma + s, t\gamma + u\}$ is. But

$$[t\gamma+u, r\gamma+s] = (t\gamma+u)[1, \tfrac{r\gamma+s}{t\gamma+u}] \sim [1, M\gamma],$$

hence $[1,\gamma] = (t\gamma+u)[1, M\gamma]$ 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 $[1,\gamma] = (-t\gamma+r)[1, M^{-1}\gamma]$ and therefore

$$A'i(Q) = AA'[1,\gamma] = AA'(-t\gamma+r)[1, M^{-1}\gamma] = A(-t\gamma+r)i(Q|_M).$$

Thus equivalent quadratic forms correspond to equivalent ideals, and we have proved:

Proposition 5.18. *The map i defined in (5.6) 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 some order $\mathcal{O}$ of the complex quadratic number field K with discriminant $\Delta = f^2 \operatorname{disc} K$, we write $\mathfrak{a} = [\alpha, \beta]$ and put

$$f(\mathfrak{a}) = Q_{\mathfrak{a}}(x,y) = \frac{N(\alpha x - \beta y)}{N(\mathfrak{a})}. \tag{5.7}$$

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 $\mathcal{O}$. 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 5.19. *The map f defined in (5.7) maps ideal classes to equivalence classes of quadratic forms.*

Now we can state

Theorem 5.20. *Consider the ideal class group $\mathrm{Cl}(\mathcal{O})$ of some order $\mathcal{O}$ of discriminant $\Delta < 0$, and the class group $\mathrm{Cl}(\Delta)$ of primitive quadratic forms of discriminant Δ. Then the maps $i : \mathrm{Cl}(\Delta) \longrightarrow \mathrm{Cl}(\mathcal{O})$ and $f : \mathrm{Cl}(\mathcal{O}) \longrightarrow \mathrm{Cl}(\Delta)$ are group homomorphisms and inverse to each other. In particular, $\mathrm{Cl}(\mathcal{O}) \simeq \mathrm{Cl}(\Delta)$.*

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.

It remains to show that ι is a homomorphism. Assume for simplicity that the forms Q_1 and Q_2 satisfy $\gcd(A_1, A_2, B) = 1$. Then $[Q_1] + [Q_2] = [Q_3]$, where Q_3 is determined up to equivalence by $A_3 = A_1A_2$ and the unique integer $B_3 \bmod 2A_3$ determined by the congruences

$$\begin{aligned} B_3 &\equiv -B_1 \bmod 2A_1, \\ B_3 &\equiv -B_2 \bmod 2A_2, \\ BB_3 &\equiv \frac{\Delta - B_1B_2}{2} \bmod 2A_3. \end{aligned}$$

Now

$$\begin{aligned} \iota(Q_1) &= [A_1, \tfrac{-B_1+\sqrt{\Delta}}{2}] = [A_1, \tfrac{-B_3+\sqrt{\Delta}}{2}], \\ \iota(Q_2) &= [A_2, \tfrac{-B_2+\sqrt{\Delta}}{2}] = [A_2, \tfrac{-B_3+\sqrt{\Delta}}{2}], \\ \iota(Q_3) &= [A_3, \tfrac{-B_3+\sqrt{\Delta}}{2}]. \end{aligned}$$

Setting $\gamma = \frac{-B_3+\sqrt{\Delta}}{2}$ we find that we have to show that $[A_1, \gamma][A_2, \gamma] = [A_1A_2, \gamma]$. But

$$[A_1, \gamma][A_2, \gamma] = [A_1A_2, A_1\gamma, A_2\gamma, \gamma^2] = [A_1A_2, A_1\gamma, A_2\gamma, -B_3\gamma] = [A_1A_2, \gamma]$$

since $\gamma^2 \equiv -B_3\gamma \bmod A_1A_2$ and $\gcd(A_1, A_2, B_3) = 1$.

In fact, we have

$$\gamma^2 = \Big(\frac{B_3 - \sqrt{\Delta}}{2}\Big)^2 = \frac{B_3^2 - 2B_3\sqrt{\Delta} + B_3^2 - 4A_3C_3}{4}$$
$$= -B_3\frac{-B_3 + \sqrt{\Delta}}{2} - A_3C_3 \equiv -B_3\gamma \bmod A_3,$$

and this implies the claimed congruence since $A_3 = A_1A_2$. □

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 relation $I_2I_3 = (2) \sim (1)$.

Exercises

5.1 Show that $\mathrm{SL}_2(\mathbb{Z})$ acts on the upper half plane, i.e., that $(MN)z = M(Nz)$ for $z \in \mathcal{H}$.

5.2 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.

5.3 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 $(MN)^{-1}z = N^{-1}M^{-1}z$.

5.4 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 $(g, x) \longmapsto xg^{-1}$.

5.5 Show that the two binary quadratic forms $(1, 0, 3)$ and $(1, 1, 1)$ represent the same integers, but that they are not equivalent.

5.6 Show that every form with discriminant $\Delta < 0$ represents some integer $n \neq 0$ with $n \leq \sqrt{-\Delta/3}$.

5.7 Show that if (A, B, C) and (A', B', C') have the same discriminant, then $B \equiv B' \bmod 2$.

5.8 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 $\left(\begin{smallmatrix}\gamma\\ \delta\end{smallmatrix}\right) = M\left(\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right)$.

5.9 Show that

$$(1, -B, AC) \sim \begin{cases} (1, 0, -m) & \text{if } B^2 - 4AC = 4m, \\ (1, 1, -m) & \text{if } B^2 - 4AC = 4m + 1. \end{cases}$$

5.10 Show that any quadratic form of type $(a, 0, c)$ or (a, a, c) and $a > 1$ has order 2 in the class group.

5.11 Consider a discriminant $\Delta = 4m < 0$ with squarefree $m \equiv 3 \bmod 4$. Assume that m has exactly t distinct (odd) prime factors. Show that the following forms are all reduced, and that they form a subgroup of $\mathrm{Cl}(\Delta)$ isomorphic to $(\mathbb{Z}/2\mathbb{Z})^t$:

1. $Q = (a, 0, c)$ for all factorizations $m = ac$ with $a < c$.
2. $Q = (2a, 2a, c)$ for all factorizations $m = a(a - 2c)$ with $2a < c$.

5.12 Prove analogs for the statements in the preceding exercises for forms of discriminants $\Delta = 8m$ and $\Delta = m \equiv 1 \bmod 4$.

6. Elliptic and Hyperelliptic Curves

Let K be a field of characteristic $\neq 2$. Then an equation $\mathcal{C} : Y^2 = f(X)$, where $f \in K[X]$ is a squarefree polynomial, A defines a hyperelliptic curve. We will usually assume that f has odd degree; in this case, $\deg f = 2g + 1$, and the integer g is called the genus of $\mathcal{C}$.

In this chapter we will construct a group $\mathrm{Jac}(\mathcal{C})$ called the Jacobian of $\mathcal{C}$; if $\mathcal{C}$ is elliptic, i.e., if it has genus 1, then $\mathrm{Jac}(\mathcal{C})$ can be identified with the group of points on E. For the construction we consider quadratic forms with coefficients in the polynomial ring $\mathcal{O} = K[T]$, which is a UFD with unit group $\mathcal{O}^\times = K^\times$. The sign of some nonzero $f \in \mathcal{O}$ is defined by $\mathrm{sgn}(f) = a_n$, where $f(T) = a_n T^n + \ldots + a_1 T + a_0$. Note that a necessary condition for $a \in \mathbb{Z}$ to be a square is that $a > 0$, and a necessary condition for $f \in \mathcal{O}$ to be a square is that $\deg f$ is even and $\mathrm{sgn}(f)$ is a square.

6.1 Quadratic Forms

A quadratic form is a triple $Q = (A, B, C)$ of elements $A, B, C \in \mathcal{O}$. It is called primitive if $\gcd(A, B, C) = 1$. The discriminant of Q is the class of $B^2 - 4AC$ modulo squares, that is, $\Delta = (B^2 - 4AC)K^{\times 2}$.

In order to get a reduction theory for quadratic forms, we need a measure for the size of the coefficients; such a measure is provided by the degree function. In order to make it multiplicative, we choose a real number $\rho > 1$ and put $|A| = \rho^{\deg A}$ for $A \in \mathcal{O}$.

Two forms Q, Q' are called equivalent if there is a matrix $M \in \mathrm{SL}_2(\mathcal{O})$ such that $Q' = Q|_M$. We say that a primitive form (A, B, C) of discriminant Δ is reduced if

$$|B| < |A| \leq |C| \quad \text{and} \quad \mathrm{sgn}(A) = 1.$$

Since $\Delta = B^2 - 4AC$ and $|B^2| < |\Delta|$, we must have $|AC| = |\Delta|$.

Lemma 6.1. *If the form (A, B, C) is reduced, then $|A| < \sqrt{|\Delta|}$.*

Proof. If Q is reduced, then $B^2 - \Delta = 4AC$ shows that $\deg \Delta = \deg A + \deg C \geq 2 \deg A$. □

Note that if K is a finite field, then there are only finitely many reduced forms since there are only finitely many polynomials A with degree bounded by $\frac{1}{2}\deg \Delta$. If K has infinitely many elements (for example if $K = \mathbb{Q}$), then it is often difficult to decide whether there are infinitely many reduced forms or not.

Example. Let us determine all reduced forms of discriminant $\Delta = 1 - T^2$ over $\mathbb{F}_3$. If $\deg A = 0$, then $A = 1$. If $\deg A = 1$, then $A = T - a$ for some $a \in \mathbb{F}_3$, and going through all possibilities shows that the following list of reduced forms is complete:

A	B	Q
1	0	$(1, 0, T^2)$
T	± 1	$(T, \pm 1, T)$
$T-1$	0	$(T-1, 0, T+1)$
$T+1$	0	$(T+1, 0, T-1)$

The last two forms are clearly equivalent (switch A and C). The second pair of forms similarly are equivalent to each other since $(T, 1, T) \sim (T, -1, T)$. In particular, there are reduced forms that are equivalent to each other.

In order to keep things as simple as possible we restrict our attention from now on to discriminants Δ satisfying the following properties:

1. $\Delta(T)$ is squarefree;
2. $\deg \Delta$ is odd;
3. $\operatorname{sgn}(\Delta) = -1$.

Remark. If $\deg \Delta$ is odd, we always have $\deg A < \deg C$ for reduced forms. In fact, from $B^2 - 4AC = \Delta$ and $\deg B < \deg A, \deg C$ it follows that $\deg A + \deg C = \deg \Delta$. This implies that $\deg A$ and $\deg C$ have distinct parity, hence we cannot have $|A| = |C|$ in this case.

A form Q is called positive definite if $\operatorname{sgn}(Q(x, y))$ is a square for all nonzero $x, y \in \mathcal{O}$. Since equivalent forms represent the same elements, $Q|_M$ will be positive definite if and only if Q is. Note that over an algebraically closed field, every form is positive definite.

Example. The form $Q = (1, 0, T^2)$ over $\mathbb{F}_3$ is not positive definite since it represents $2T^2 = Q(T, 1)$, and $\operatorname{sgn}(2T^2) = 2$ is not a square in $\mathbb{F}_3$.

Lemma 6.2. *If* $\deg \Delta$ *is odd, then the form* $Q = (A, B, C)$ *is positive definite if and only if* $\operatorname{sgn}(A)$ *is a square.*

Proof. Using the action of $\mathrm{SL}_2(\mathcal{O})$ we may assume that Q is reduced, i.e., that $|B| < |A| \leq |C|$. Now consider the expression $Q(x, y) = Ax^2 + Bxy + Cy^2$. Since $|B| < |A|, |C|$, the middle term never influences the leading coefficient $\operatorname{sgn}(Q(x, y))$: in fact,

- if $|x| \geq |y|$ then $|Bxy| \leq |Bx^2| < |Ax^2|$;
- if $|x| \leq |y|$ then $|Bxy| \leq |By^2| < |Cy^2|$;

thus the dominating term is either Ax^2 or Cy^2. Since $4AC = B^2 - \Delta$ and $-\Delta$ is monic, we have $A > 0$ if and only if $C > 0$. □

Theorem 6.3. *If* $\deg \Delta$ *is odd, then every positive definite form of discriminant* Δ *is equivalent to a unique reduced form.*

Proof. Let us first recall that, for $M = \left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$ and a quadratic form $Q = (A, B, C)$, we have

$$Q|_M = (A, B + 2Ab, Ab^2 + Bb + C). \tag{6.1}$$

Similarly,

$$Q|_S = (C, -B, A) \tag{6.2}$$

for $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

Now take a positive definite primitive quadratic form (A, B, C) with discriminant Δ, and perform the following operations:

1. If $\deg B > \deg A$, find $b \in \mathcal{O}$ with $\deg(B + 2Ab) < \deg A$, and replace (A, B, C) by $(A', B', C') = (A, B, C)|M = Q|_M$ for $M = \left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$. Then $\deg B' = \deg(B + 2Ab) < \deg A'$ since $A' = A$.
2. If $\deg A' \leq \deg C'$ goto step 3. If $\deg A' > \deg C'$, use S to replace the form (A', B', C') by $(C', -B', A')$. If $\deg B' > \deg C'$, goto step 1.
3. Now we have a quadratic form (A'', B'', C'') with $\deg B'' < \deg A'' \leq \deg C''$. Since Q was assumed to be positive definite, $\operatorname{sgn}(A) = a^2$ will be a square. Now replace Q'' by $Q''|_R$, where $R = \left(\begin{smallmatrix} 1/a & 0 \\ 0 & a \end{smallmatrix}\right)$.

For showing that two reduced forms are never equivalent we use Lemma 6.4 below. Assume that $(A, B, C) \sim (A', B', C')$ are reduced forms of the same discriminant. Then A is the minimal element represented by the first, and since equivalent forms represent the same elements, also by the second form. This implies $A = A'$.

Now write $(A', B', C') = (A, B, C)|_M$ for $M = \left(\begin{smallmatrix} r & s \\ t & u \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathcal{O})$; then $A = A' = Ar^2 + Brt + Ct^2$. Since (A, B, C) is reduced, the dominating term on the right hand side is either Ar^2 or Ct^2 (these cannot cancel since $\deg A$ and $\deg C$ have different parity). But this term must have the same degree as A, hence $r = \pm 1$ and $t = 0$. Since $M \in \mathrm{SL}_2(\mathcal{O})$, we must have $r = u$, hence $M = \left(\begin{smallmatrix} \pm 1 & s \\ 0 & \pm 1 \end{smallmatrix}\right)$.

But now $B' = B \pm 2sA$; since $\deg B, \deg B' < \deg A$, this is only possible if $s = 0$. Since $M = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ acts trivially on forms, we now see that $B' = B$ and $C' = C$ as claimed. □

Lemma 6.4. *The minimal monic element represented properly by a reduced form* Q *is* $Q(1, 0) = A$.

Proof. Assume that $R = Q(x,y) = Ax^2 + Bxy + Cy^2$ satisfies $\deg R \leq \deg A$. We have already observed that the middle term never dominates, hence $\deg R = \deg A + 2\deg x \geq \deg A$ or $\deg R = \deg C + 2\deg y \geq \deg C \geq \deg A$. Thus if R is minimal, we must have $\deg R = \deg A$ and $x \in K^\times$. Since R and A are monic, this implies $R = A$. □

Let us now determine all reduced forms of discriminant $\Delta = 1 - T^3$ over $\mathbb{F}_5$ (note that $1 - T^3 = (1-T)^3$ over $\mathbb{F}_3$). Except for the principal form $Q_0 = (1, 0, -T^3)$, a reduced form can be written as (A, B, C) with $1 \leq \deg A < \deg \Delta$. Since $\deg \Delta = 3$, we find $\deg A = 1$, hence $A = T + a$ for some $a \in K = \mathbb{F}_5$. From $1 - T^3 = \Delta = B^2 - 4AC = B^2 - 4(T+a)C$ we deduce that we have to choose $B = b \in K$ for a given $A = T + a$ in such a way that $(T + a) \mid b^2 - \Delta$. This is equivalent to $\Delta(-a) = b^2$. This equation in K will not always have a solution; we find

A	B	Q	(a,b)
T	± 1	$(T, \pm 1, -T^2)$	$(0, \pm 1)$
$T-1$	0	$(T-1, 0, -T^2 - T - 1)$	$(1, 0)$
$T-2$	$-$	$-$	$-$
$T-3$	± 2	$(T-3, \pm 2, -T^2 + 2T - 1)$	$(3, \pm 2)$
$T-4$	$-$	$-$	

Thus there are exactly 6 reduced forms of discriminant $1 - T^3$.

Elliptic Curves

Now assume that $\Delta(T) = -T^3 + -aT + b$. Every form of discriminant Δ is equivalent to a reduced form (A, B, C); such a form is reduced if $\deg B < \deg A < \deg \Delta = 3$, that is, if $\deg A \leq 1$ and $\deg B \leq 0$.

If $\deg A = 0$ and $\operatorname{sgn}(A) = 1$, then $A = 1$, and the unique reduced form is $Q_0 = (1, 0, -\frac{\Delta}{4})$. This form is called the principal form of discriminant Δ.

If $\deg A = 1$ and $\operatorname{sgn}(A) = 1$, then $A = T - a$ and $B = b$ for $a, b \in K$. From $\Delta(T) = B^2 - 4AC = b^2 - 4(T-a)C$ we get $\Delta(a) = b^2$ by plugging in $T = a$. Thus the reduced quadratic form (A, B, C) corresponds to a point (a, b) on the elliptic curve $Y^2 = \Delta(T)$. Conversely, any such point (a, b) gives rise to a reduced form (A, B, C) with discriminant Δ by putting $A = T - a$, $B = b$, and $C = \frac{B^2 - \Delta}{4A}$ (note that $b^2 - \Delta(T)$ is divisible by $T - a$, so we do in fact have $C \in \mathcal{O}$). If we agree to send the principal quadratic form $Q_0 = (1, 0, -\frac{\Delta}{4})$ to the point at infinity on $E : Y^2 = \Delta(T)$, we get

Theorem 6.5. *There is a bijection between the points on the projective elliptic curve $E : Y^2 = \Delta(T)$ with $\Delta(T) = -T^3 + aT + b \in K[T]$ and the reduced quadratic forms of discriminant Δ.*

We observe again that if $K = \mathbb{F}_q$ is a finite field, then there are only finitely many points on E (and thus only finitely many reduced forms); in

this case, the class number of forms of discriminant Δ coincides with the number of points on $E(\mathbb{F}_p)$.

If Δ is squarefree of odd degree ≥ 5, the curves $\mathcal{C} : Y^2 = \Delta(X)$ are called hyperelliptic curves. In this case, the curve has too few points to carry a nice group structure. Our discussion above, however, shows exactly what we have to do to get a group law. A reduced quadratic form (A, B, C) for hyperelliptic curves with $\deg \Delta = 5$ satisfies $\deg A \leq 2$ and $\deg B \leq 1$. Consider the case $\deg A = 2$; let α, α' denote the two roots of A in some extension of K. Then $B^2 - 4AC = \Delta$ implies $\Delta(\alpha) = B(\alpha)^2$, hence $(\alpha, B(\alpha))$ and $(\alpha', B(\alpha'))$ are points on $\mathcal{C}$ defined over some quadratic extension of K. Thus we get essentially a bijection between pairs of such points and reduced forms (there are a couple of details to fill in for forms with $\deg A \leq 1$ etc.). The group of such pairs of points on $\mathcal{C}$ is called the Jacobian of $\mathcal{C}$; as for elliptic curves, the group law has a geometric interpretation.

Example. Consider the hyperelliptic curve $Y^2 = 1 - T^5$ over $\mathbb{F}_3$. Its points $\mathcal{O}$, $(a, b) = (0, \pm 1), (1, 0)$ correspond to the reduced forms $(1, 0, -\Delta)$, $(T, \pm 1, T^4)$ and $(T-1, 0, T^4+T^3+T^2+T+1)$. A brute force enumeration shows that the remaining reduced forms are given by $(T^2, \pm 1, T^3)$, $(T^2 - T, \pm(T-1), T^3 + T^2 + T - 1)$ and $(T^2 - T - 1, \pm T, T^3 + T^2 - T + 1)$.

6.2 The Class Group

If F is a finite field with p elements, then we always take $\rho = p$; in this case, $|A|$ is the number of elements in the residue class ring $\mathcal{O}/A$.

Now assume that $K = \mathbb{F}_p$ is a finite field with $p > 2$ elements. Let $|f| = p^{\deg f}$ denote the norm of f, that is, the cardinality of the system of residue classes $\mathcal{O}/(f)$.

Proposition 6.6. *Assume that $K = \mathbb{F}_q$ is a finite field. Then there are only finitely many reduced forms of discriminant $\Delta < 0$.*

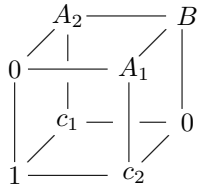
Proof. This is obvious since there are only finitely many linear polynomials over K. □

The equivalence classes of forms of discriminant Δ can be made into a group via Bhargava's cubes as before. Now consider reduced forms $Q_j = (X - x_j, y_j + \sqrt{\Delta})$ corresponding to the points (x_j, y_j) on the elliptic curve $E : Y^2 = \Delta(X)$, and assume that $[Q_1] + [Q_2] + [Q_3] = [Q_0]$. We would like to derive formulas for Q_3 in terms of x_1, y_1, x_2, y_2.

The forms are $Q_j = (T - x_j, y - j, C_j)$; assume first that $x_1 \neq x_2$; then $\gcd(A_1, A_2) = 1$, and composition is easy: recall that for composing Q_1 and Q_2 we have to solve the system of equations

$$\begin{aligned} B\alpha - A_1\gamma &= C_2, \\ A_2\alpha - A_1\beta &= (B_2 - B_1)/2, \\ B\beta - A_2\gamma &= C_1. \end{aligned}$$

In our current situation, $B = \frac{1}{2}(B_1 + B_2)$ is a unit in $\mathcal{O}$. This means that we can choose an arbitrary $\gamma \in \mathcal{O}$, and then solve the first and the third equation for α and β. If we take $\gamma = 0$ and introduce the abbreviations $c_j = C_j/B$, then we get the following cube:



This gives us the form $Q_3 = (A_1A_2, B_3, C_3)$, where

$$\begin{aligned} A_3 &= A_1A_2, \\ B_3 &= c_1A_1 + c_2A_2 - B, \\ C_3 &= c_1c_2. \end{aligned}$$

Note that, according to our calculations in Chapter 5,

$$BB_3 = A_1C_1 + A_2C_2 - B^2 = -\frac{\Delta + B_1B_2}{2},$$

hence $B_3 = -\frac{\Delta + B_1B_2}{B_1 + B_2}$.

As you can see, the new form Q_3 will in general not be reduced. Our first task will be to reduce B_3 mod A_3 in such a way that the degree of B_3 decreases. This is done as follows:

Lemma 6.7. *If $A_1 - A_2, B_1 + B_2 \in K^\times$ then we have*

$$\frac{\Delta + B_1B_2}{B_1 + B_2} \equiv \frac{A_1B_2 - A_2B_1}{A_1 - A_2} \bmod A_1A_2.$$

In the special case $\deg \Delta = 3$, we know that B_1 and B_2 have degree 0, whereas A_1 and A_2 will be monic of degree 1 (unless one of the forms involved is the principal form); in particular, $A_1 - A_2$ and $B_1 + B_2$ *are* constants (possibly 0). Thus the expression on the right hand side (if it is defined) has degree 1, whereas the one on the left has degree 3.

Proof. We have to verify that

$$(A_1 - A_2)(\Delta + B_1B_2) \equiv (A_1B_2 - A_2B_1)(B_1 + B_2) \bmod A_1A_2.$$

Plugging in $\Delta = B_1^2 - 4A_1C_1$, multiplying out, cancelling and omitting the term $4A_1A_2C_1 \equiv 0 \bmod A_1A_2$ shows that this congruence is equivalent to

$$A_1(B_1^2 - B_2^2 - 4A_1C_1) \equiv 0 \bmod A_1A_2.$$

But now $B_1^2 - 4A_1C_1 = B_2^2 - 4A_2C_2$, hence the last congruence becomes $4A_1A_2C_2 \equiv 0 \bmod A_1A_2$, which is obviously correct. □

Example. Consider the forms

$$Q_1 = (T, 1, -T^2) \quad \text{and} \quad Q_2 = (T-1, 0, -T^2 - T - 1)$$

of discriminant $\Delta = 1 - T^3$ over $\mathbb{F}_5$. Then $A_3 = T^2 - T$ and $B_3 = -\frac{\Delta + B_1B_2}{B_1 + B_2} = T^3 - 1$; using Lemma 6.7 we get $B_3 \equiv -\frac{A_1B_2 - A_2B_1}{A_1 - A_2} = T - 1 \bmod A_1A_2$: in fact, we have $T^3 - 1 - (T+1)(T^2 - T) = T - 1$. Thus $Q_3 \sim (T^2 - T, T - 1, C_3')$ with $C_3' = -T - 2$. The reduction process now yields $Q_3 \sim (-T - 2, 1 - T, T^2 - T) \sim (-T - 2, 3, T^2 - 2T - 1)$. Getting rid of the coefficient $-1 = 2^2$ via $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right)$ finally gives $Q_3 \sim (T + 2, -2, -T^2 + 2T + 1)$.

Let us go back to the problem of composing two forms $Q_j = (A_j, B_j, C_j)$, $j = 1, 2$, in the case where $\deg \Delta = 3$, and where $A_j = T - a_j$ for distinct elements $a_1, a_2 \in K$. We have seen so far that $[Q_1] + [Q_2] + [Q_3] = 0$ for the form

$$Q_3 = \Big(A_1A_2, -\frac{\Delta + B_1B_2}{B_1 + B_2}, *\Big) \sim \Big(A_1A_2, -\frac{A_1B_2 - A_2B_1}{A_1 - A_2}, *\Big).$$

Let us now compute the third coefficient C of the last form. It is given by $C = \frac{\Delta - B^2}{4A}$, where $A = A_1A_2$ and $B = \frac{A_1B_2 - A_2B_1}{A_1 - A_2}$. Since $\Delta = -T^3 - aT + b$ and $\deg B \leq 1$, we find that $\Delta - B^2$ is a cubic polynomial; moreover, it is divisible by $A_1 = T - a_1$ and $A_2 = T - a_2$, hence we can write $B^2 - \Delta = (T - a_1)(T - a_2)(T - a_3)$ for some constant a_3. The easiest way of computing a_3 is by comparing the coefficients of T^2 on both sides. For doing this we need

Lemma 6.8. *If $A_j = T - a_j$ and $B_j = b_j$ for constants $a_1, a_2, b_1, b_2 \in K$ with $a_1 \neq a_2$, then*

$$\frac{A_1B_2 - A_2B_1}{A_1 - A_2} = \frac{b_2 - b_1}{a_2 - a_1}(T - a_1) + b_1.$$

Proof. This can be checked by a straightforward coputation:

$$\frac{A_1B_2 - A_2B_1}{A_1 - A_2} = \frac{(T - a_1)b_2 - (T - a_2)b_1}{a_2 - a_1} = \frac{b_2 - b_1}{a_2 - a_1}(T - a_1) + b_1.$$

As we will see below, it is no accident that the expression on the right hand side is part of the equation of a line. □

Thus if we put $m = \frac{b_2 - b_1}{a_2 - a_1}$, then the coefficient of T^2 in $B^2 - \Delta$ is m^2, and we find

Theorem 6.9. *Assume that $Q_1 = (A_1, B_1, C_1)$ and $Q_2 = (A_2, B_2, C_2)$ are forms of discriminant $\Delta = -T^3 - aT + b$, where $A_1 = T - a_1$ and $A_2 = T - a_2$ for distinct $a_1, a_2 \in K$. Then $[Q_1] + [Q_2] + [Q_3] = 0$ for the form $Q_3 = (T - a_3, b_3, *)$, where a_3 and b_3 are determined by the equations $a_1 + a_2 + a_3 + m^2 = 0$, $m = \frac{b_2 - b_1}{a_2 - a_1}$, and $b_3 = m(T - a_1) + b_1$.*

The Geometric Group Law

Let us now consider quadratic forms whose discriminant is a squarefree cubic, and show that composition of forms induces the well known geometric group law for adding points on the corresponding elliptic curve. The latter is defined by the two conditions, namely that the point $\mathcal{O}$ at infinity be the neutral element, and that collinear points add up to 0.

Assume that we are given points $P_1 = (a_1, b_1)$ and $P_2 = (a_2, b_2)$ on an elliptic curve $Y^2 = -X^3 - aX + b$. Assume first that $a_1 \neq a_2$. Then the slope of the line through P_1P_2 is given by $m = \frac{b_2 - b_1}{a_2 - a_1}$. The line through P_1 and P_2 is thus given by $Y = m(X - a_1) + b_1$; intersecting this line with the cubic gives us the equation

$$(m(X - a_1) + b_1)^2 + X^3 + aX - b = 0.$$

This is a cubic equation with solutions $X = a_1, a_2, a_3$, hence we must have

$$(m(X - a_1) + b_1)^2 + X^3 + aX - b = (X - a_1)(X - a_2)(X - a_3).$$

Comparing the coefficients of X^2 on both sides shows

$$a_1 + a_2 + a_3 = -m^2,$$

that is,

$$a_3 = -a_1 - a_2 - m^2.$$

Plugging this into the line equation gives us the value of b_3 as

$$b_3 = m(a_3 - a_1) + b_1.$$

The point (a_3, b_3) corresponds to the class of the form $(T - a_3, b_3, *)$; since we have seen above that $(T - a_3, b_3, *) \sim Q_3 = (A_3, B_3, C_3)$ for the form Q_3 composed from $Q_1 = (T - a_1, b_1, *)$ and $Q_2 = (T - a_2, b_2, *)$, we are done.

Exercises

6.1 Construct a matrix in $\mathrm{SL}_2(\mathcal{O})$ with nonconstant entries.

6.2 Show that if Q is a form with discriminant Δ over some finite field K, and if $\deg \Delta$ is even, then Q is not definite.

6.3 Consider the cubic $\mathcal{C} : Y^2 = -X^3$. Show that the points on $\mathcal{C} \setminus \{(0,0)\}$ carry a natural group structure by setting up a bijection between all points on $\mathcal{C}$ and the classes of forms of discriminant $\Delta(T) = -T^3$. Observe that the point $(0,0)$ corresponds to the class of the non-primitive form $(T, 0, -T^2/4)$. Show that this group is isomorphic to the additive group of K.

6.4 Show similarly that the cubic $Y^2 = -X^3 + X^2$ can be given a group structure.