The Arithmetic of Binary Quadratic Forms

Bachelorarbeit in Mathematik
eingereicht an der Fakultät für Mathematik und Informatik
der Georg-August-Universität Göttingen
am 22. Juli, 2019

von
Radu Toma

Erstgutachter:
Prof. Dr. Valentin Blomer

Zweitgutachter:
Prof. Dr. Preda Mihăilescu

Contents

1 Introduction

One of the most famous results in elementary number theory is Fermat's theorem on sums of two squares. It states that an odd prime is the sum of two squares if and only if it is congruent to 1 modulo 4. This is one of the earliest examples of the study of integral binary quadratic forms, i.e. quadratic polynomials in two variables of the shape

$$f(x, y) = ax^2 + bxy + cy^2$$

with integer coefficients. Some of the brightest mathematicians have contributed to this theory since then, including Euler, Lagrange, Gauß, and many others.

The systematic approach began with Lagrange, who introduced many of the concepts presented in the second chapter of this thesis. The elementary theory of binary quadratic forms then culminated with Gauß' great insights published in his monumental book, *Disquisitiones Arithmeticae.* Subsequently, it was discovered that this theory is intimately connected to the arithmetic of ideals in quadratic number fields. This abundance of structure is what allowed mathematicians to prove very satisfying and, indeed, beautiful results about binary quadratic forms.

Although this theory is already very rich and mature, there is still progress being made currently. Many applications have been found, one of the more recent and important ones being cryptography. Some of these new results and applications concern a simple question: How many integers can be written as $ax^2 + bxy + cy^2$ for some integers x and y? This is the main topic of this thesis.

To answer the question, we will first need an overview of the theory of integral binary quadratic forms and the arithmetic of ideals in quadratic orders, which is given in chapters 2 and 3, respectively. The tools that we need for our purposes can be understood with very few prerequisites. The lure of a self-contained and complete thesis ultimately led to a lengthier treatment of the theory due to giving almost all proofs fully. A reader inexperienced in algebraic number theory will thus be provided with a thorough but relatively concise (in comparison to most books on the topic) introduction necessary for the rest of this thesis. The already knowledgeable reader might enjoy the simplified exposition of more general concepts and theorems applied to the imaginary quadratic case.

The fourth chapter is dedicated to the main theorem of this thesis. It is a generalization of Theorem 2 from Valentin Blomer and Andrew Granville's paper [BG06] on representation numbers of quadratic forms. Let $R(g, n)$ be the number of integer solutions to the equation

$$n = g(x, y),$$

where $g(x, y) = ax^2 + bxy + cy^2$ is an integral binary quadratic form. Suppose the discriminant of g, defined as

$$D := b^2 - 4ac,$$

is negative and that a is positive. For a non-negative exponent β, the main theorem asserts the following asymptotic:

$$\sum_{\substack{1 \leq n \leq x \\ n \in \mathbb{N}}} R(g, n)^\beta \sim C_{g,\beta} \frac{x}{\sqrt{|D|}},$$

where $C_{g,\beta}$ is a positive constant depending on g and β.

This has been proved by Blomer and Granville when D is essentially square-free, i.e. a so-called fundamental discriminant, and generalized to arbitrary negative discriminants in this thesis. The result should be appreciated with care, since it only holds if x is not too big in relation to $|D|$. Indeed, the behaviour changes in different ranges of x, as it is described in [BG06].

Chapter 4 analyses in its first section some of these limitations and shortcomings of the main theorem. Although further study is needed, the result is strong enough to be applied successfully in a counting problem involving integral Apollonian circle packings. These are fractal-like configurations of circles, all of which have integer curvatures. The question is whether the set of all curvatures found in a given circle packing has positive density inside the natural numbers. Jean Bourgain and Elena Fuchs showed in [BF11] that the answer is affirmative.

Indeed, one can show that the set of curvatures in an integral circle packing contains the set of numbers properly represented by some shifted binary quadratic forms. A quadratic form g represents an integer n properly if there exist integers x and y, such that $g(x, y) = n$, with the extra condition that x and y are coprime. This suggests that we would need a analogue of the main theorem for proper representations. The difficulties of producing such a result are discussed in the second section of chapter 4. Nevertheless, we will gather enough estimates to prove the positive density theorem of Bourgain and Fuchs in the last section of this thesis.

2 Binary quadratic forms

This section presents the basic theory of binary quadratic forms. The exposition is mainly distilled from the books [Cox13] and [Zag81].

2.1 Equivalence of forms and reduction

Definition 2.1. An *(integral binary quadratic) form* is a homogeneous quadratic polynomial in two variables

$$f(x,y) = ax^2 + bxy + cy^2$$

with integer coefficients. We call $D = D_f = b^2 - 4ac \in \mathbb{Z}$ the *discriminant* of f and we say that f is a *primitive* form if $\gcd(a,b,c) = 1$.

From this definition we infer directly that the discriminant of a form can only be congruent to 0 or 1 modulo 4. Moreover, an arbitrary form is obviously equal to a multiple of a primitive form.

Definition 2.2. A number $D \in \mathbb{Z}$ is called a *fundamental discriminant* if $D \equiv 1 \pmod 4$ and D is squarefree or if $D \equiv 0 \pmod 4$, $D/4$ is squarefree and $D/4 \equiv 2$ or $3 \pmod 4$. In general, an integer $D \neq 0$ is called a (quadratic) *discriminant* if $D \equiv 0, 1 \pmod 4$. Any discriminant is the unique product of a fundamental discriminant and a square: if we write $D = D_0 f_D^2$ with D_0 a fundamental discriminant and $f_D \in \mathbb{Z}$, then we call f_D the *conductor* of D.

The implicit assertion in this definition is seen directly by simple manipulations for which we refer to [HK13, Theorem 1.1.6].

Remark 2.3. Calling each integer $D \equiv 0, 1 \pmod 4$ a discriminant is reasonable, since there are always forms such that D is their discriminant. Indeed, in each case we have the form

$$f(x,y) = \begin{cases} x^2 - \dfrac{D}{4}y^2, & D \equiv 0 \pmod 4, \\ x^2 - xy + \dfrac{1-D}{4}y^2, & D \equiv 1 \pmod 4, \end{cases}$$

which is called the *principal form* of discriminant D.

In his famous book, *Disquisitiones arithmeticae* [Gau01], Gauß introduces an equivalence relation on the set of quadratic forms[1], which lies at the foundation of this theory.

[1] To be precise, Gauß only considered forms with an even middle coefficient. Nevertheless, he was the first to understand the importance of *proper* equivalence. Before him, Lagrange had used a weaker notion of equivalence, where $ps - qr$ in Definition 2.4 was allowed to be -1 as well. Gauß makes this distinction in [Gau01, §157, §158] and assures us that its usefulness will soon reveal itself.

Definition 2.4. We say that two forms f and g are (properly) *equivalent* if there exist integers p, q, r and s such that

$$f(x, y) = g(px + qy, rx + sy) \quad \text{and} \quad ps - qr = 1.$$

We may rephrase this last definition by noting that two forms are equivalent if and only if they lie in the same orbit with respect to the action of the group $\mathrm{SL}_2(\mathbb{Z})$ defined by

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} f(x, y) = f(px + qy, rx + sy). \tag{2.1}$$

One can easily check that this is indeed a group action and, thus, equivalence of forms is an equivalence relation. Simple computations show that the discriminant is invariant under this action and that forms equivalent to a primitive form are primitive as well.

Lemma 2.5. *Let $f(x, y) = ax^2 + bxy + cy^2$ be a form with discriminant $D < 0$ and first coefficient $a > 0$. Then $f(x, y) > 0$ for all pairs of integers $(x, y) \neq (0, 0)$. Moreover, if $g(x, y) = a'x^2 + b'xy + c'y^2$ is equivalent to f, then $a' > 0$.*

Proof. The first claim follows from the identity

$$4af(x, y) = (2ax + by)^2 - Dy^2. \tag{2.2}$$

For the second assertion notice that if

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} f(x, y) = a'x^2 + b'xy + c'y^2, \quad \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),$$

then $a' = f(p, r)$. Since $ps - qr = 1$, it follows that $(p, r) \neq (0, 0)$ and therefore $a' > 0$. □

Thus, the action of $\mathrm{SL}_2(\mathbb{Z})$ is well-defined if we restrict to primitive forms with negative discriminant and positive first coefficient.

Definition 2.6. A form $f = ax^2 + bxy + cy^2$ with discriminant $D < 0$ is called *positive definite* if $a > 0$ and *negative definite* if $a < 0$. We denote the set of equivalence classes of primitive positive definite forms with fixed discriminant D by $\mathfrak{F}_D$ and denote the class of f by $[f] = [a, b, c]$. We call $\mathfrak{F}_D$ the *form class group.*

Remark 2.7. We have not yet shown that the form class group has indeed a suitable group structure. One can introduce the so-called *Dirichlet composition* of forms explicitly (see

[Cox13, Chapter 3]), but we shall take a more abstract route and transfer the group structure from the ideal class group to $\mathfrak{F}_D$ in section 3.4.

Clearly, a form f is positive definite if and only if the form $-f$ is negative definite. Thus, without loss of generality, we may from now on restrict our considerations to primitive positive definite forms, which have more robust properties. The theory of forms with positive discriminant, called indefinite forms, can also be developed in an analogous way, but there are fundamental dissimilarities, the treatment of which would exceed the scope of this thesis.

The following definition and theorem give a set of representatives for $\mathfrak{F}_D$.

Definition 2.8. A positive definite form $f = ax^2 + bxy + c^2$ of discriminant $D < 0$ is called *reduced* if either $-a < b \leq a < c$ or $0 \leq b \leq a = c$.

Theorem 2.9. *Let $D < 0$ be a quadratic discriminant. Every class $F \in \mathfrak{F}_D$ contains exactly one reduced form.*

Proof. We only show here that every given primitive positive definite form is equivalent to one satisfying $|b| \leq a \leq c$. From the equivalence class of the given form, choose $f(x, y) = ax^2 + bxy + cy^2$ such that $|b|$ is minimal. For any integer m, the form

$$g(x, y) = f(x + my, y) = ax^2 + (2am + b)xy + c'y^2$$

is equivalent to f. If $a < |b|$, then we can choose m such that $|2am + b| < |b|$, which contradicts the choice of b. Thus $|b| \leq a$ and $|b| \leq c$ follows analogously. If $a > c$, we interchange the outer coefficients by the transformation $(x, y) \mapsto (-y, x)$.

This first step is enough to prove Corollary 2.11. For the rest of the proof, including uniqueness, we refer to Theorem 2.8 in [Cox13]. □

Remark 2.10. If $f(x, y) = ax^2 + bxy + cy^2$ is a reduced positive definite form, then $a = \min_{x,y\in\mathbb{Z}} f(x, y)$. Indeed, this follows from the inequality

$$ax^2 + bxy + cy^2 \geq a(x^2 + y^2) + bxy \geq a \cdot 2|xy| - |bxy| \geq a|xy|.$$

Corollary 2.11. *Let $D < 0$ be a quadratic discriminant. The number of classes of positive definite primitive forms, i.e. the cardinality of $\mathfrak{F}_D$, is finite.*

Proof. The claim follows from Theorem 2.9 by proving there are only finitely many reduced forms. Indeed, for $f = ax^2 + bxy + cy^2$ reduced we have

$$|D| = -D = 4ac - b^2 \geq 4a^2 - a^2 = 3a^2. \tag{2.3}$$

Thus, there are only finitely many possible values for a. Since $|b| \leq a$ and c is determined by a and b, the claim follows. □

The equivalence notion thus reduces the infinite number of forms having a given discriminant to a finite set, which, although still difficult to understand, is much more approachable. The usefulness of the equivalence of forms will become even more apparent as the theory develops.

2.2 Representation of integers

Definition 2.12. If f is a form and n an integer, we say that f *represents* n if there exist $x, y \in \mathbb{Z}$ such that $f(x, y) = n$. If, in addition, x and y are coprime, then we say that f *properly represents* n.

Remark 2.13. A key observation is that, for an integer n and a matrix $\gamma = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \gamma \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} xp + yq \\ xr + ys \end{pmatrix}$$

restricts to a bijection from the set of representations $\{(x, y) \in \mathbb{Z}^2 \mid (\gamma f)(x, y) = n\}$ onto the set of representations $\{(x', y') \in \mathbb{Z}^2 \mid f(x', y') = n\}$. This follows from (2.1) and the fact that $\mathrm{SL}_2(\mathbb{Z})$ matrices are invertible. We can see easily that this bijection restricts to proper representations. Thus, all forms in a particular class (properly) represent the same integers.

In general, an arbitrary integer could be represented by a form in more than one way, that is, with different pairs (x, y). How many such pairs is a central and very old question in the theory of quadratic forms.

Definition 2.14. For $n \in \mathbb{N}$ and f a binary quadratic form we define the *number of representations* as

$$R(f, n) = \#\{(x, y) \in \mathbb{Z}^2 : f(x, y) = n\}.$$

For a class $C \in \mathfrak{F}_D$, we set $R(C, n) = R(f, n)$ for some $f \in C$. By Remark 2.13, this is well-defined. Similarly, we define

$$R^*(f, n) = \#\{(x, y) \in \mathbb{Z}^2 : f(x, y) = n, \gcd(x, y) = 1\}$$

to be the *number of proper representations*.

The following results about the representation of integers will be used later on.

Lemma 2.15. *A form f properly represents an integer m if and only if f is equivalent to the form $mx^2 + b'xy + c'y^2$ for some $b', c' \in \mathbb{Z}$.*

Proof. Suppose that $f(p, r) = m$ for coprime integers p and r. By Bézout's lemma we find $q, s \in \mathbb{Z}$ such that $pq - rs = 1$. If $f(x, y) = ax^2 + bxy + cy^2$, then

$$f(px + qy, rx + sy) = f(p, r)x^2 + Bxy + Cxy,$$

for some $B, C \in \mathbb{Z}$. Conversely, the form $mx^2 + Bxy + Cy^2$ properly represents m by $(x, y) = (1, 0)$ and the claim follows by Remark 2.13. □

Lemma 2.16. *For an integer m, a primitive form f properly represents at least one integer coprime to m.*

Proof. If $f(x, y) = ax^2+bxy+cy^2$ with $\gcd(a, b, c) = 1$ and p is a prime, then at least one of the values $f(1, 0)$, $f(0, 1)$ and $f(1, 1)$ is coprime to p. Thus, for each prime dividing m we have congruence conditions on x and y such that p does not divide $f(x, y)$. The claim now follows by combining these conditions using Chinese Remainder Theorem. □

3 Orders in quadratic fields

Binary quadratic forms are intimately connected to quadratic number fields, i.e. quadratic extensions of the field of rational numbers. We will first look at particular subrings called orders and study the arithmetic of their ideals. Section 3.4 is the highlight of this chapter, showing the correspondence between forms and ideals of orders. We will later make use of this connection and apply the tools and notions presented below to obtain more insight into forms and the representation of integers.

The theory developed here is mostly self-contained, in the sense that knowledge of algebraic number theory is not necessary. Nevertheless, I do assume some acquaintance with linear algebra over modules, for which some references from Serge Lang's standard textbook [Lan02] are given. The exposition is inspired by the books [Cox13], [BS66], [HK13], and the online notes of Keith Conrad. Most of the proofs in this section are adapted versions or a combination of the ones found in the references.

3.1 Basic definitions

Remark 3.1. A quadratic number field can be written uniquely as $\mathbb{Q}(\sqrt{D_0})$ with D_0 a fundamental discriminant (see [HK13, Theorem 1.1.9]). For any $D = D_0 f^2$ we have $\mathbb{Q}(\sqrt{D}) = \mathbb{Q}(\sqrt{D_0})$.

Definition 3.2. A *lattice* in a quadratic number field K is a subset $L \subset K$ that is a free $\mathbb{Z}$-module of rank 2. Equivalently, a lattice $L \subset K$ is a finitely generated $\mathbb{Z}$-module that contains a $\mathbb{Q}$-basis of K. An *order* $\mathcal{O}$ of K is a lattice that is also a subring of K containing 1.

Remark 3.3. The equivalence of the definitions is shown in [HK13, p. 115]. For a lattice L we shall write

$$L = [\omega_1, \omega_2] = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2,$$

where (ω_1, ω_2) is a basis of L.

Definition 3.4. Let D be a quadratic discriminant, $K = \mathbb{Q}(\sqrt{D})$ and D_0 the fundamental discriminant associated to D. Define

$$\omega_D = \frac{\sigma_D + \sqrt{D}}{2}, \quad \text{where} \quad \sigma_D = \begin{cases} 0, & D \equiv 0 \pmod{4}, \\ 1, & D \equiv 1 \pmod{4}. \end{cases}$$

Then we call $\mathcal{O}_D = [1, \omega_D]$ the *quadratic order of discriminant* D and we call $\mathcal{O}_K = \mathcal{O}_{D_0}$ the *maximal order* or the *ring of integers* of K.

One sees quickly that $\mathcal{O}_D$ is indeed an order in K and the following theorems justify the terminology introduced above and the usage of definite articles.

Theorem 3.5. *Let D be a quadratic discriminant and $K = \mathbb{Q}(\sqrt{D})$. Then the following assertions hold.*

1. *If $f \in \mathbb{N}$, then $\mathcal{O}_{Df^2} = [1, f\omega_D] = \mathbb{Z} + f\mathcal{O}_D \subset \mathcal{O}_D$ and $[\mathcal{O}_D : \mathcal{O}_{Df^2}] = f$. In particular, $\mathcal{O}_D \subset \mathcal{O}_K$ and $[\mathcal{O}_K : \mathcal{O}_D] = f_D$ is the conductor of D.*

2. *Let D' be a quadratic discriminant. Then $\mathcal{O}_{D'}$ is a subset of $\mathcal{O}_D$ if and only if $D' = Df^2$ for some $f \in \mathbb{N}$. In particular, $\mathcal{O}_D = \mathcal{O}_{D'}$ if and only if $D = D'$.*

Proof. 1. Let $f \in \mathbb{N}$. The identities

$$\omega_{Df^2} = \begin{cases} \dfrac{\sqrt{Df^2}}{2} = f\omega_D, & D \equiv 0 \pmod 4, \\ \dfrac{\sqrt{Df^2}}{2} = -\dfrac{f}{2} + f\omega_D, & D \equiv 1 \pmod 4, \quad f \equiv 0 \pmod 2, \\ \dfrac{1+\sqrt{Df^2}}{2} = \dfrac{1-f}{2} + f\omega_D, & D \equiv 1 \pmod 4, \quad f \equiv 1 \pmod 2, \end{cases}$$

show that $\mathcal{O}_{Df^2} = [1, \omega_{Df^2}] = [1, f\omega_D] \subset \mathcal{O}_D$. The rest of the assertions now follows easily.

2. If $D' = Df^2$, then $\mathcal{O}_{D'} \subset \mathcal{O}_D$ by 1. Conversely, if $\mathcal{O}_{D'} \subset \mathcal{O}_D$ and $f = [\mathcal{O}_D : \mathcal{O}_{D'}]$, then $\mathcal{O}_{D'}$ is an order in K and, by 1, we see that $f_{D'} = [\mathcal{O}_K : \mathcal{O}_{D'}] = [\mathcal{O}_K : \mathcal{O}_D]f = f_D f$. Hence, we obtain $D' = D_0 f_{D'}^2 = D_0 f_D^2 f^2 = Df^2$. □

Definition 3.6. Let $K = \mathbb{Q}(\sqrt{D})$ be a quadratic number field.

1. An element $\xi \in K$ is called an *algebraic integer* if the minimal polynomial of ξ over $\mathbb{Q}$, i.e. the monic polynomial of least degree with rational coefficients for which ξ is a root, has integer coefficients. In this context, we shall refer to numbers in $\mathbb{Z}$ as *rational integers*.

2. Let $\alpha \mapsto \alpha'$ be the nontrivial automorphism of K, that is, $a + b\sqrt{D} \mapsto a - b\sqrt{D}$. We define the *norm* of $\alpha \in K$ as $N(\alpha) = \alpha\alpha'$ and its *trace* as $T(\alpha) = \alpha + \alpha'$.

Theorem 3.7. *Let K be a quadratic field. Then $\xi \in K$ is an algebraic integer if and only if $\xi \in \mathcal{O}_K$.*

Proof. Let $K = \mathbb{Q}(D)$ with D a fundamental discriminant. Let $\xi = a + b\sqrt{D} \in K$ with $a, b \in \mathbb{Q}$. If $b = 0$, then ξ is the root of the irreducible polynomial $X - a$ and ξ is an algebraic integer if and only if $\xi \in \mathbb{Z} \subset \mathcal{O}_K$. If $b \neq 0$, then ξ is the root of the irreducible polynomial $X^2 - 2aX + a^2 - Db^2 = X^2 - T(\xi)X + N(\xi)$. Thus, ξ is an algebraic integer if and only if $2a \in \mathbb{Z}$ and $a^2 - Db^2 \in \mathbb{Z}$. Using that D is of the form $2^r N$ with $r = 2$ or 3 and N squarefree, we see that this last condition is equivalent to $\xi \in \mathcal{O}_K$. □

Theorem 3.8. *If $\mathcal{O}$ is an order of the quadratic number field K, then there exists a quadratic discriminant D, such that $\mathcal{O} = \mathcal{O}_D$.*

Proof. The quotient of free $\mathbb{Z}$-modules of equal rank is finite (see [BS66, Lemma 1, Chap. 2, Sec. 6]; this also follows, for instance, from the elementary divisor theorem in [Lan02, Chap. III, Theorem 7.8]). Thus, $[\mathcal{O}_K : \mathcal{O}] = f \in \mathbb{N}$ and then $f\mathcal{O}_K \subset \mathcal{O}$. It follows that $\mathbb{Z} + f\mathcal{O}_K \subset \mathcal{O}$ and by Theorem 3.5 we obtain that $\mathcal{O} = \mathcal{O}_{D_0 f^2}$, where $\mathcal{O}_{D_0} = \mathcal{O}_K$, since $\mathcal{O}_{D_0 f^2} \subset \mathcal{O}$ and $[\mathcal{O}_K : \mathcal{O}] = f = [\mathcal{O}_K : \mathcal{O}_{D_0 f^2}]$. □

Definition 3.9. If the quadratic number field K is fixed and D_0 is the fundamental discriminant associated to K, then the *conductor of an order* $\mathcal{O}$ is the conductor of D when we write $\mathcal{O} = \mathcal{O}_D = \mathcal{O}_{D_0 f^2}$ using the previous theorem.

From now on we shall only develop the theory for *imaginary* quadratic orders, i.e. orders $\mathcal{O}_D$ with $D < 0$, analogously to our treatment of binary quadratic forms. In this case, the nontrivial automorphism is given by complex conjugation and the norm of an element is the square of its absolute value. Thus, the norm is multiplicative and non-negative. This gives us strong restrictions on the group of units.

Theorem 3.10. *Let $D < 0$ be a quadratic discriminant. Then the group of units $\mathcal{O}_D^*$ of $\mathcal{O}_D$ is a finite group given by*

$$\mathcal{O}_D^* = \{\xi \in \mathcal{O}_D \mid N(\xi) = 1\} = \begin{cases} \langle e^{2\pi i/6} \rangle, & D = -3, \\ \langle i \rangle, & D = -4, \\ \langle -1 \rangle, & D < -4, \end{cases} \quad \text{and } |\mathcal{O}_D^*| = \begin{cases} 6, & D = -3, \\ 4, & D = -4, \\ 2, & D < -4, \end{cases}$$

Proof. If $\xi \in \mathcal{O}_D$ is a unit, then ξ is an algebraic integer and its norm $N(\xi)$ is an integer (as in the proof of Theorem 3.7). By multiplicativity of the norm we must have $N(\xi) \mid 1$ and it follows that $N(\xi) = 1$. We can find $a, b \in \mathbb{Z}$ such that $\xi = \frac{a+b\sqrt{D}}{2}$, so ξ is a unit if and only if

$$a^2 - Db^2 = 4 \tag{3.1}$$

The rest follows by simple computations, using that D is negative. □

We shall later denote the number of units in the order $\mathcal{O}_D$ by w_D.

3.2 Ideals of quadratic orders

Next, we study the arithmetic of ideals in quadratic orders.

Remark 3.11. Since orders are free $\mathbb{Z}$-modules of rank 2 and ideals are submodules, one can show that ideals must also be lattices (see [Lan02, Theorem 7.1]).

Lemma 3.12. *Let $\mathcal{O}$ be a quadratic order and $\mathfrak{a} \subset \mathcal{O}$ a nonzero ideal. Then $\mathcal{O}/\mathfrak{a}$ is finite.*

Proof. We first show that $\mathfrak{a}$ contains a nonzero integer. Indeed, notice that if $\mathcal{O} = \mathcal{O}_D$, then any $\alpha \in \mathcal{O}$ is of the form $\alpha = (a + b\sqrt{D})/2$ with $a, b \in \mathbb{Z}$, implying that $\overline{\alpha} \in \mathcal{O}$. If $0 \neq \alpha \in \mathfrak{a}$, then $N(\alpha) = \alpha\overline{\alpha} \in \mathfrak{a} \cap \mathbb{Z}$, since α is an algebraic integer.

Now let $0 \neq m \in \mathfrak{a} \cap \mathbb{Z}$. Since $\mathcal{O}$ is a free $\mathbb{Z}$-module of rank 2, it follows that $\mathcal{O}/m\mathcal{O} \cong \mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z}$ is finite. The canonical map $\mathcal{O}/m\mathcal{O}_D \twoheadrightarrow \mathcal{O}_D/\mathfrak{a}$ is surjective and therefore $\mathcal{O}_D/\mathfrak{a}$ is finite as well. □

Definition 3.13. The *norm* of a nonzero ideal $\mathfrak{a}$ in a quadratic order $\mathcal{O}_D$ is defined as $N(\mathfrak{a}) = |\mathcal{O}_D/\mathfrak{a}|$.

We would like the ideals of a given order to behave like numbers. For instance, they should be invertible (the definition will be given below) and their norms should be multiplicative. In general, some of these ideals are in fact ideals of bigger orders as well. Colloquially speaking, they do not really belong to the given order and this ruins the afore mentioned properties we wish to have. Therefore we need to restrict the definition of ideals to the ones which properly belong to an order.

Definition 3.14. Let $\mathcal{O}$ be an order in the quadratic number field K. A *fractional $\mathcal{O}$-ideal* is a set of the form $\alpha\mathfrak{a}$ for $\alpha \in K^*$ and $\mathfrak{a}$ a nonzero ideal of $\mathcal{O}$. A fractional $\mathcal{O}$-ideal $\mathfrak{b}$ is called *proper* if $\{\beta \in K | \beta\mathfrak{b} \subset \mathfrak{b}\} = \mathcal{O}$ and we call $\mathfrak{b}$ *invertible* if there exists another fractional $\mathcal{O}$-ideal $\mathfrak{c}$ such that $\mathfrak{b}\mathfrak{c} = \mathcal{O}$.

For example, all principal ideals $\xi\mathcal{O}$, $\xi \in \mathcal{O}$, are proper and invertible as well. More generally, we have the following equivalence.

Theorem 3.15. *Let $\mathcal{O}$ be an order in a quadratic field K and let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal. Then $\mathfrak{a}$ is proper if and only if $\mathfrak{a}$ is invertible.*

Proof. If $\mathfrak{a}$ is invertible, then there exists a fractional $\mathcal{O}$-ideal $\mathfrak{b}$ such that $\mathfrak{a}\mathfrak{b} = \mathcal{O}$. By our definition of ideals, we have the inclusion $\mathcal{O} \subset \{\beta \in K | \beta\mathfrak{a} \subset \mathfrak{a}\}$. Conversely, if $\beta \in K$ satisfies $\beta\mathfrak{a} \subset \mathfrak{a}$, then $\beta\mathcal{O} = \beta(\mathfrak{a}\mathfrak{b}) = (\beta\mathfrak{a})\mathfrak{b} \subset \mathfrak{a}\mathfrak{b} = \mathcal{O}$ and thus $\beta \in \mathcal{O}$ since $1 \in \mathcal{O}$. Therefore $\mathfrak{a}$ is proper. For the other direction we first prove a lemma, following [Cox13].

Lemma 3.16. *Let $K = \mathbb{Q}(\tau)$ be a quadratic number field and let $ax^2 + bx + c$ be the minimal polynomial of τ, where a, b and c are relatively prime integers. Then $[1, \tau]$ is a proper fractional ideal for the order $[1, a\tau]$ of K.*

Proof. Computing the minimal polynomial of $a\tau$ shows that $a\tau$ is an algebraic integer, which implies that $[1, a\tau]$ is an order. Then $[1, \tau] = \frac{1}{a}[a, a\tau]$ is a fractional ideal of $[1, a\tau]$. For $\beta \in K$ note that $\beta[1, \tau] \subset [1, \tau]$ is equivalent to $\beta \in [1, \tau]$ and $\beta\tau \in [1, \tau]$. The first condition says that $\beta = m + n\tau$ for $m, n \in \mathbb{Z}$. It now follows that

$$\beta\tau = m\tau + n\tau^2 = m\tau + \frac{n}{a}(-b\tau - c) = \frac{-cn}{a} + \left(\frac{-bn}{a} + m\right)\tau.$$

Since $\gcd(a, b, c) = 1$, we see that $\beta\tau \in [1, \tau]$ if and only if a divides n. It follows that $\{\beta \in K \mid \beta[1, \tau] \subset [1, \tau]\} = [1, a\tau]$. □

Now let $\mathfrak{a}$ be a proper ideal of $\mathcal{O}$. Write $\mathfrak{a} = [\alpha, \beta]$ for $\alpha, \beta \in K$ (see Remark 3.11). Then $\mathfrak{a} = \alpha[1, \tau]$ with $\tau = \beta/\alpha$. Let $ax^2 + bx + c$ be the minimal polynomial of τ with a, b, c coprime integers (we find this since τ lies in the quadratic number field K). Lemma 3.16 implies that $\mathcal{O} = [1, a\tau]$. Since $\overline{\tau}$ is the other root of this polynomial, it follows again by Lemma 3.16 that $\overline{\mathfrak{a}} = \overline{\alpha}[1, \overline{\tau}]$ is a fractional ideal of $\mathcal{O} = [1, a\overline{\tau}] = [1, a\tau]$, since $\overline{\mathcal{O}} = \mathcal{O}$ by Theorem 3.8. We now prove that

$$\mathfrak{a}\overline{\mathfrak{a}} = \frac{N(\alpha)}{a}\mathcal{O}, \tag{3.2}$$

from which it follows that $\mathfrak{a}$ is invertible. Note that

$$a\mathfrak{a}\overline{\mathfrak{a}} = a\alpha\overline{\alpha}[1, \tau][1, \overline{\tau}] = N(\alpha)[a, a\tau, a\overline{\tau}, a\tau\overline{\tau}].$$

Since by Vieta $\tau + \overline{\tau} = -b/a$ and $\tau\overline{\tau} = c/a$, this becomes

$$a\mathfrak{a}\overline{\mathfrak{a}} = N(\alpha)[a, a\tau, -b, c] = N(\alpha)[1, a\tau] = N(\alpha)\mathcal{O},$$

since a, b, c are coprime. □

The following lemma reaffirms that the notion of properness is well-chosen, since norms of proper ideals behave just like norms of numbers.

Lemma 3.17. *Let $\mathcal{O}$ be an order in an imaginary quadratic field and let $\mathfrak{a}, \mathfrak{b} \subset \mathcal{O}$ be proper ideals. Then:*

1. *$N(\alpha\mathcal{O}) = N(\alpha)$ for $\alpha \in \mathcal{O}, \alpha \neq 0$;*

2. $N(\mathfrak{a}\mathfrak{b}) = N(\mathfrak{a})N(\mathfrak{b})$*;*

3. $\mathfrak{a}\overline{\mathfrak{a}} = N(\mathfrak{a})\mathcal{O}$*.*

Proof. Let the given order have the form $\mathcal{O} = [1, u]$. For $\alpha \in \mathcal{O}$ we find $a, b, c, d \in \mathbb{Z}$ such that $\alpha = a + bu$ and $u\alpha = c + du$. Then $u\overline{\alpha} = au + b|u|^2$ and $\overline{\alpha} = uc/|u|^2 + d$ and thus

$$N(\alpha) = \alpha\overline{\alpha} = (a + bu)\overline{\alpha} = (ac/|u|^2 + ba)u + ad + b^2|u|^2.$$

Since $N(\alpha) \in \mathbb{Z}$, this implies that $ac/|u|^2 = -ba$ and then $N(\alpha) = ad - bc$ easily follows. Now from linear algebra (see [BS66, Lemma 1, Chap. 2, Sec. 6]) we know that the index $[\mathcal{O} : \alpha\mathcal{O}]$ is equal to the absolute value of the determinant of any transition matrix from a basis of $\mathcal{O}$ to a basis of $\alpha\mathcal{O}$. Since a basis of $\alpha\mathcal{O}$ is $[\alpha, \alpha u]$, it follows that $N(\alpha\mathcal{O}) = [\mathcal{O} : \alpha\mathcal{O}] = ad - bc = N(\alpha)$.

Next, we generalize the result above and prove that $N(\alpha\mathfrak{a}) = N(\alpha)N(\mathfrak{a})$ for an $\mathcal{O}$-ideal $\mathfrak{a}$. The inclusions $\alpha\mathfrak{a} \subset \alpha\mathcal{O} \subset \mathcal{O}$ imply that $[\mathcal{O} : \alpha\mathfrak{a}] = [\mathcal{O} : \alpha\mathcal{O}][\alpha\mathcal{O} : \alpha\mathfrak{a}]$. Since multiplication by α induces an isomorphism $\mathcal{O}/\mathfrak{a} \cong \alpha\mathcal{O}/\alpha\mathfrak{a}$, we get $N(\alpha\mathfrak{a}) = N(\alpha\mathcal{O})N(\mathfrak{a})$, and our claim now follows from *1.*

We may write $\mathfrak{a} = \alpha[1, \tau]$ as in the proof of Theorem 3.15 and then Lemma 3.16 tells us that $\mathcal{O} = [1, a\tau]$, where we can choose a to be positive without loss of generality. Since obviously $[a, a\tau]$ has index a in $[1, a\tau]$, we obtain that $N(a[1, \tau]) = a$. Applying norms to the equality $a\mathfrak{a} = \alpha a[1, \tau]$, we infer by the above that

$$N(\mathfrak{a}) = N(\alpha)/a. \tag{3.3}$$

Assertion *3.* now follows by recalling the identity (3.2), that is, $\mathfrak{a}\overline{\mathfrak{a}} = (N(\alpha)/a)\mathcal{O}$.

Since by *3.* we have $N(\mathfrak{a}\mathfrak{b})\mathcal{O} = \mathfrak{a}\mathfrak{b} \cdot \overline{\mathfrak{a}\mathfrak{b}} = \mathfrak{a}\overline{\mathfrak{a}} \cdot \mathfrak{b}\overline{\mathfrak{b}} = N(\mathfrak{a})N(\mathfrak{b})\mathcal{O}$, claim *2.* now follows. □

3.3 Prime ideals and unique factorization

In this section we will study the factorization of ideals into prime ideals. For non-maximal orders, not all proper ideals factorize uniquely, which is one of the difficulties of this theory (for counterexamples see [Conb, Section 8]). Nevertheless, a big enough collection of ideals does have unique factorization. These are the *ideals coprime to the conductor* f of the order $\mathcal{O}$, i.e. ideals $\mathfrak{a} \subset \mathcal{O}$ such that $\mathfrak{a} + f\mathcal{O} = \mathcal{O}$.

Lemma 3.18. *Let* $\mathcal{O}$ *be an order of conductor* f*.*

1. A nonzero $\mathcal{O}$*-ideal* $\mathfrak{a}$ *is coprime to* f *if and only if its norm* $N(\mathfrak{a})$ *is coprime to* f*.*

2. Every $\mathcal{O}$-ideal coprime to f is proper.

Proof. To prove *1*, let $m_f : \mathcal{O}/\mathfrak{a} \longrightarrow \mathcal{O}/\mathfrak{a}$ be the multiplication by f homomorphism. Then we easily see the equivalences

$$\mathfrak{a} + f\mathcal{O} = \mathcal{O} \iff m_f \text{ is surjective } \iff m_f \text{ is an isomorphism,}$$

where the last equivalence follows from the finiteness of $\mathcal{O}/\mathfrak{a}$. By the structure theorem for finite Abelian groups (see [Lan02, Theorems 8.1 and 8.2]), m_f is an isomorphism if and only if f is coprime to the order $N(\mathfrak{a})$ of $\mathcal{O}/\mathfrak{a}$.

To show *2*, let $\beta \in K$ satisfy $\beta\mathfrak{a} \subset \mathfrak{a}$. If $\mathfrak{a} = [\omega_1, \omega_2]$, then we see that multiplication by β corresponds to an integral matrix, i.e.

$$\beta \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = A \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}, A \in \mathcal{M}_2(\mathbb{Z}).$$

Then β is a root of the characteristic polynomial of A, which is a monic polynomial with integral coefficients, and thus $\beta \in \mathcal{O}_K$. We now have

$$\beta\mathcal{O} = \beta(\mathfrak{a} + f\mathcal{O}) = \beta\mathfrak{a} + \beta f\mathcal{O} \subset \mathfrak{a} + f\mathcal{O}_K.$$

Since $f\mathcal{O}_K \subset \mathcal{O}$ by Theorem 3.5, this implies that $\beta\mathcal{O} \subset \mathcal{O}$ and thus $\beta \in \mathcal{O}$, since $1 \in \mathcal{O}$. We have now proven that $\{\beta \in K \mid \beta\mathfrak{a} \subset \mathfrak{a}\} \subset \mathcal{O}$ and the other inclusion is obvious, because $\mathfrak{a}$ is an ideal of $\mathcal{O}$. □

Remark 3.19. Using the same proof, we can generalize the first statement of Lemma 3.18: for any $m \in \mathbb{Z} \setminus \{0\}$, a nonzero $\mathcal{O}$-ideal $\mathfrak{a}$ is coprime to m if and only if its norm $N(\mathfrak{a})$ is coprime to m. (This is stated in Lemma 5.9.1 in [HK13].)

The former lemma shows in particular that all ideals of the maximal order are proper. The arithmetic of ideals in the ring of integers is indeed easier to describe and handle. We will now show how to relate smaller orders to the maximal one, so that we may translate back some of the properties that we later compute for the ring of integers.

Theorem 3.20. *Let $\mathcal{O}$ be the order of conductor f in an imaginary field K.*

1. *If $\mathfrak{a}$ is an $\mathcal{O}_K$-ideal coprime to f, then $\mathfrak{a} \cap \mathcal{O}$ is an $\mathcal{O}$-ideal coprime to f of the same norm. If $\mathfrak{a}$ is prime, then so is $\mathfrak{a} \cap \mathcal{O}$.*

2. *If $\mathfrak{a}$ is an $\mathcal{O}$-ideal coprime to f, then $\mathfrak{a}\mathcal{O}_K$ is an $\mathcal{O}_K$-ideal coprime to f of the same norm.*

3. The map $\mathfrak{a} \mapsto \mathfrak{a} \cap \mathcal{O}$ *from the monoid of* $\mathcal{O}_K$*-ideals coprime to* f *to the monoid of* $\mathcal{O}$*-ideals coprime to* f *is an isomorphism and its inverse is given by* $\mathfrak{a} \mapsto \mathfrak{a}\mathcal{O}_K$*.*

Proof. For the first claim, let $\mathfrak{a}$ be an $\mathcal{O}_K$ ideal coprime to f. Consider the injection $\mathcal{O}/(\mathfrak{a} \cap \mathcal{O}) \hookrightarrow \mathcal{O}_K/a$ given by the inclusion $\mathcal{O} \subset \mathcal{O}_K$. Multiplication by f induces an isomorphism of $\mathcal{O}_K/\mathfrak{a}$ because $\mathfrak{a}$ is coprime to f. Since $f\mathcal{O}_K \subset \mathcal{O}$, the injection above is also a surjection and, thus, an isomorphism. This shows that the norms of $\mathfrak{a}$ and $\mathfrak{a} \cap \mathcal{O}$ are equal. By Lemma 3.18, $\mathfrak{a} \cap \mathcal{O}$ is coprime to f. The isomorphism above also shows that, if $\mathfrak{a}$ is prime, then $\mathcal{O}_K/\mathfrak{a} \cong \mathcal{O}/(\mathfrak{a} \cap \mathcal{O})$ is an integral domain. Therefore $\mathfrak{a} \cap \mathcal{O}$ is prime as well.

To prove *2*, let $\mathfrak{a}$ be an $\mathcal{O}$-ideal coprime to f. Since

$$\mathfrak{a}\mathcal{O}_K + f\mathcal{O}_K = (\mathfrak{a} + f\mathcal{O})\mathcal{O}_K = \mathcal{O}\mathcal{O}_K = \mathcal{O}_K,$$

we see that $\mathfrak{a}\mathcal{O}_K$ is also coprime to f. The statement about the norms follows from *3*, which we prove in the next paragraphs.

We first show that $\mathfrak{a}\mathcal{O}_K \cap \mathcal{O} = \mathfrak{a}$ for $\mathfrak{a}$ an $\mathcal{O}$-ideal coprime to f. We have the inclusion

$$\begin{aligned}\mathfrak{a}\mathcal{O}_K \cap \mathcal{O} &= (\mathfrak{a}\mathcal{O}_K \cap \mathcal{O})\mathcal{O} \\ &= (\mathfrak{a}\mathcal{O}_K \cap \mathcal{O})(\mathfrak{a} + f\mathcal{O}) \\ &\subset \mathfrak{a} + f(\mathfrak{a}\mathcal{O}_K \cap \mathcal{O}) \subset \mathfrak{a} + \mathfrak{a} \cdot f\mathcal{O}_K \subset \mathfrak{a},\end{aligned}$$

since $f\mathcal{O}_K \subset \mathcal{O}$. The other inclusion is obvious.

Next we show that $(\mathfrak{a} \cap \mathcal{O})\mathcal{O}_K = \mathfrak{a}$ for $\mathfrak{a}$ an $\mathcal{O}_K$-ideal coprime to f. Note that

$$\mathfrak{a} = \mathfrak{a}\mathcal{O} = \mathfrak{a}(\mathfrak{a} \cap \mathcal{O} + f\mathcal{O}) \subset (\mathfrak{a} \cap \mathcal{O})\mathcal{O}_K + f\mathfrak{a}.$$

Since $f\mathfrak{a} \subset f\mathcal{O}_K \subset \mathcal{O}$, we have the inclusion $f\mathfrak{a} \subset \mathfrak{a} \cap \mathcal{O} \subset (\mathfrak{a} \cap \mathcal{O})\mathcal{O}_K$, which implies $\mathfrak{a} \subset (\mathfrak{a} \cap \mathcal{O})\mathcal{O}_K$. The other inclusion is obvious. Thus the two maps in *1.* and *2.* are inverses of each other.

Now to show that the maps preserve the multiplicative structures, it is enough to prove that one of them is a homomorphism, since they are inverses of each other. This is obvious for $\mathfrak{a} \mapsto \mathfrak{a}\mathcal{O}_K$:

$$(\mathfrak{a}\mathfrak{b})\mathcal{O}_K = \mathfrak{a}\mathcal{O}_K \cdot \mathfrak{b}\mathcal{O}_K.$$

Finally, *1.* and *3.* imply the norm statement of *2.* □

We will now look at the multiplicative structure of ideals in more detail.

Lemma 3.21. *Let $\mathcal{O}$ be a quadratic order. If $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero proper $\mathcal{O}$-ideals, then $\mathfrak{a} \subset \mathfrak{b}$ if and only if $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ for some proper $\mathcal{O}$-ideal $\mathfrak{c}$.*

Proof. If $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$, then $\mathfrak{a} \subset \mathfrak{b}$ follows since $\mathfrak{b}$ is an $\mathcal{O}$-ideal and $\mathfrak{c} \subset \mathcal{O}$.

If $\mathfrak{a} \subset \mathfrak{b}$, then let $\mathfrak{c} = \mathfrak{b}^{-1}\mathfrak{a} \subset \mathfrak{b}^{-1}\mathfrak{b} \subset \mathcal{O}$. Thus $\mathfrak{c}$ is an $\mathcal{O}$-ideal and $\mathfrak{a} = \mathfrak{c}\mathfrak{b}$. Since $\mathfrak{a}$ is invertible, the last equality shows that $\mathfrak{c}$ is invertible and, equivalently, proper. □

Theorem 3.22. *Let $\mathcal{O}$ be an imaginary quadratic order with conductor f.*

1. *If $\mathfrak{p}$ is an nonzero prime ideal of $\mathcal{O}$, then $\mathfrak{p}$ is a maximal ideal, $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ for some (rational) prime p that is the only prime lying in $\mathfrak{p}$.*

2. *Every ideal is contained in some prime ideal.*

3. *Let $\mathfrak{a}, \mathfrak{b}$ be ideals of an order $\mathcal{O}$ and $\mathfrak{p} \subset \mathcal{O}$ be a prime ideal. If $\mathfrak{a}\mathfrak{b} \subset \mathfrak{p}$, then $\mathfrak{a} \subset \mathfrak{p}$ or $\mathfrak{b} \subset \mathfrak{p}$.*

4. *Let $\mathfrak{a}$ be a nonzero ideal of $\mathcal{O}$ such that $(N(\mathfrak{a}), f) = 1$ Then $\mathfrak{a}$ is a product of prime ideals in a unique way up to the order of factors. In particular, if $\mathcal{O}$ is the maximal order $\mathcal{O}_K$, then every nonzero ideal is a product of prime ideals in an essentially unique way.*

Proof. 1. Let $\mathfrak{p}$ be a nonzero prime ideal of $\mathcal{O}$. Then $\mathcal{O}/\mathfrak{p}$ is a finite integral domain by Lemma 3.12, hence a field, and thus $\mathfrak{p}$ is maximal. Since $\mathfrak{p} \cap \mathbb{Z}$ is a nonzero prime ideal of $\mathbb{Z}$, we have $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ for some prime p, which is therefore the only rational prime inside $\mathfrak{p}$.

2. This follows by an easy induction on the norm. Alternatively, this is traditionally shown for arbitrary rings by Zorn's Lemma.

3. Let $\mathfrak{a}\mathfrak{b} \subset \mathfrak{p}$. Suppose that $\mathfrak{a}$ is not a subset of $\mathfrak{p}$. Then $\mathfrak{p} \subsetneq \mathfrak{p} + \mathfrak{a}$ and therefore $\mathfrak{p} + \mathfrak{a} = \mathcal{O}$ follows by maximality of $\mathfrak{p}$. Thus $\mathfrak{b} = \mathfrak{b} \cdot \mathcal{O} = \mathfrak{b}(\mathfrak{p} + \mathfrak{a}) = \mathfrak{b}\mathfrak{p} + \mathfrak{b}\mathfrak{a} \subset \mathfrak{p}$.

4. By Lemma 3.18, $\mathfrak{a}$ is a proper ideal and every ideal containing $\mathfrak{a}$ is proper, since if $\mathfrak{a} \subset \mathfrak{b}$, then $N(\mathfrak{b}) \mid N(\mathfrak{a})$ (by the third isomorphism theorem). Now we show the claim by induction on the norm.

Assume that $\mathfrak{a} \subsetneq \mathcal{O}$. Then by *2*, there exists a prime ideal $\mathfrak{p}$ containing $\mathfrak{a}$. Since $\mathfrak{p}$ is proper, we find an $\mathcal{O}$-ideal $\mathfrak{b}$ such that $\mathfrak{a} = \mathfrak{p}\mathfrak{b}$. Since $\mathfrak{p} \neq \mathcal{O}$ by definition of prime ideals, we infer that $\mathfrak{a} \subsetneq \mathfrak{b}$. Hence $N(\mathfrak{b}) < N(\mathfrak{a})$ and, by induction, we find a factorization into prime ideals $\mathfrak{b} = \mathfrak{p}_1 \ldots \mathfrak{p}_n$, so that $\mathfrak{a} = \mathfrak{p}\mathfrak{p}_1 \ldots \mathfrak{p}_n$.

To prove uniqueness, assume that $\mathfrak{a} = \mathfrak{p}_1 \ldots \mathfrak{p}_n = \mathfrak{p}'_1 \ldots \mathfrak{p}'_m$ with $n, m \in \mathbb{N}$ and prime $\mathcal{O}$-ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n, \mathfrak{p}'_1, \ldots, \mathfrak{p}'_m$. We proceed by induction on n and note that $n = 1$ if and only if $m = 1$, and in this case the claim is obvious. Now suppose that $n \geq 2$. Since $\mathfrak{p}'_1 \dots \mathfrak{p}'_m \subset \mathfrak{p}_1$, by renumbering if necessary, we may deduce from *3.* that $\mathfrak{p}'_1 \subset \mathfrak{p}_1$. By *1.* we find that $\mathfrak{p}_1 = \mathfrak{p}'_1$ and, since these are invertible ideals, we obtain $\mathfrak{p}_2 \dots \mathfrak{p}_n = \mathfrak{p}'_2 \dots \mathfrak{p}'_m$. By induction we find that $n = m$ and $\mathfrak{p}_j = \mathfrak{p}'_j$ for all $j = 2, \dots, n$. □

Now we would like to study the prime ideals of the maximal order. We have a complete classification given by the relation between the discriminant and the rational primes inside the prime ideals.

Definition 3.23. Let n be a positive integer and p a prime. For p odd, we call a a *quadratic residue* modulo p if there exists $b \in \mathbb{Z}$ such that $b^2 \equiv a \pmod{p}$. We call a a quadratic *nonresidue* otherwise. The *Kronecker symbol* is defined as

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & p \mid a, \\ 1, & p \nmid a \text{ and } a \text{ is a quadratic residue modulo } p, \\ -1, & p \nmid a \text{ and } a \text{ is quadratic nonresidue modulo } p, \end{cases}$$

where a is a square modulo p if there exist $b \in \mathbb{Z}$ such that $b^2 \equiv a \pmod{p}$. For $p = 2$, we define

$$\left(\frac{a}{2}\right) = \begin{cases} 0, & 2 \mid a, \\ 1, & a \equiv \pm 1 \pmod{8}, \\ -1, & a \equiv \pm 3 \pmod{8}. \end{cases}$$

Theorem 3.24. *Let $K = \mathbb{Q}(D)$ be an imaginary quadratic field with D a fundamental discriminant. Let p be a prime in $\mathbb{Z}$.*

1. *If $\left(\frac{D}{p}\right) = 0$, i.e. $p \mid D$, then $p\mathcal{O}_K = \mathfrak{p}^2$ for a prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$ with $\mathfrak{p} = \overline{\mathfrak{p}}$. The prime p is called ramified.*

2. *If $\left(\frac{D}{p}\right) = 1$, then $p\mathcal{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$, where $\mathfrak{p} \neq \overline{\mathfrak{p}}$ are prime ideals in $\mathcal{O}_K$. The prime p is called split.*

3. *If $\left(\frac{D}{p}\right) = -1$, then $p\mathcal{O}_K$ is a prime ideal in $\mathcal{O}_K$. The prime p is called inert.*

Proof. We prove this theorem by explicitly constructing the prime ideals.

1. Let p be an odd prime dividing D and define $\mathfrak{p} = p\mathcal{O}_K + \sqrt{D}\mathcal{O}_K \subset \mathcal{O}_K$. Then $\mathfrak{p}$ is an ideal and by squaring we obtain

$$\mathfrak{p}^2 = p^2\mathcal{O}_K + p\sqrt{D}\mathcal{O}_K + D\mathcal{O}_K.$$

By definition of a fundamental discriminant, $p^2 \nmid D$ since p is odd, so that $(p^2, D) = p$. Thus it follows from the above that $\mathfrak{p}^2 = (p^2\mathcal{O}_K + D\mathcal{O}_K) + p\sqrt{D}\mathcal{O}_K = p\mathcal{O}_K$. Applying norms we obtain that $N(\mathfrak{p}) = p$. If $\mathfrak{p} = \mathfrak{a}\mathfrak{b}$, then $p = N(\mathfrak{a})N(\mathfrak{b})$, so that $\mathfrak{a} = \mathcal{O}_K$ or $\mathfrak{b} = \mathcal{O}_K$. Thus, by the unique factorization into prime ideals we find that $\mathfrak{p}$ is prime. By the norm formula in Lemma 3.17 and invertibility, we have $\mathfrak{p}^2 = \mathfrak{p}\overline{\mathfrak{p}}$, which implies $\mathfrak{p} = \overline{\mathfrak{p}}$.

If $p = 2$, write $D = 4N$ and set

$$\mathfrak{p} = \begin{cases} 2\mathcal{O}_K + (1+\sqrt{N})\mathcal{O}_K, & N \text{ odd}, \\ 2\mathcal{O}_K + \sqrt{N}\mathcal{O}_K, & N \text{ even}. \end{cases}$$

Now the computations are analogous to the ones done in the paragraph above.

2. If $\left(\frac{D}{p}\right) = 1$, then $p \nmid D$ and, since $D \equiv 0, 1 \pmod 4$, we may find an integer b such that $D \equiv b^2 \pmod{4p}$, where we use the Chinese Remainder Theorem if p is odd. Hence

$$D \equiv b \pmod 2, \quad 4p \mid b^2 - D, \quad p \nmid b \quad \text{and} \quad b \not\equiv -b \pmod{2p}.$$

By the above, straight forward computations show that

$$\mathfrak{p} = \left[p, \frac{b+\sqrt{D}}{2}\right] \text{ and } \overline{\mathfrak{p}} = \left[p, \frac{b-\sqrt{D}}{2}\right]$$

are $\mathcal{O}_K$-ideals.

Suppose now that $(b - \sqrt{D})/2 \in \mathfrak{p}$. Then there are $x, y \in \mathbb{Z}$ such that $(b - \sqrt{D})/2 = xp + y(b + \sqrt{D})/2$. It follows that $y = -1$ and then we must have $b = xp$, which is a contradiction. Thus $\mathfrak{p} \neq \overline{\mathfrak{p}}$. Moreover, again since p and b are coprime and writing $b^2 - D = 4px$, we compute

$$\begin{aligned} \mathfrak{p}\overline{\mathfrak{p}} &= \left[p, \frac{b+\sqrt{D}}{2}\right]\left[p, \frac{b-\sqrt{D}}{2}\right] = \left[p^2, p\frac{b+\sqrt{D}}{2}, p\frac{b-\sqrt{D}}{2}, \frac{b^2-D}{4}\right] \\ &= p\left[p, b, \frac{b+\sqrt{D}}{2}, x\right] = p\left[1, \frac{b+\sqrt{D}}{2}\right] = p\,[1, \omega_D] = p\mathcal{O}_K. \end{aligned}$$

Therefore $N(\mathfrak{p}) = N(\overline{\mathfrak{p}}) = p$ and using unique prime factorization and the multiplicativity of the norm we deduce that $\mathfrak{p}$ and $\overline{\mathfrak{p}}$ are prime ideals.

3. Let $\left(\frac{D}{p}\right) = -1$. For proving that $p\mathcal{O}$ is a prime ideal, let $x, y \in \mathcal{O}_K$ such that $xy \in p\ \mathcal{O}_K$, i.e. $xy = p\alpha$ for some $\alpha \in \mathcal{O}_K$. It follows that $N(x)N(y) = N(p)N(\alpha) =$

$p^2N(\alpha)$. We may assume that $N(p) \mid N(x)$. Write

$$x = \frac{u + v\sqrt{D}}{2} \text{ and thus } N(x) = \frac{u^2 - v^2D}{4},$$

where $u, v \in \mathbb{Z}$ and $u \equiv vD \pmod 2$, by our description of $\mathcal{O}_K$. Hence $4p$ divides $u^2 - v^2D$. If $p \neq 2$, then D is a quadratic non-residue modulo p, which implies that $u \equiv v \equiv 0 \pmod p$, hence $x \in p\mathcal{O}_K$. If $p = 2$, then $D \equiv 5 \pmod 8$ and $u^2 - v^2D \equiv u^2 - 5v^2 \equiv 0 \pmod 8$. Considering all cases shows that either $u \equiv v \equiv 0 \pmod 4$ or $u \equiv v \equiv 2 \pmod 4$. It follows that $x \in 2\mathcal{O}_K$. □

Remark 3.25. By Theorem 3.22, every prime $\mathcal{O}_K$-ideal contains a rational prime and thus the prime ideal appears in the unique factorization of the principal ideal generated by that rational prime. Therefore, the theorem above indeed classifies all prime $\mathcal{O}_K$-ideals by the rational primes they contain.

Corollary 3.26. *Let $\mathcal{O}$ be the order with conductor f in the quadratic field $K = \mathbb{Q}(D_0)$ with D_0 a fundamental discriminant. Suppose $\mathfrak{a} \subset \mathcal{O}$ is an ideal with $(N(\mathfrak{a}), f) = 1$ and*

$$N(\mathfrak{a}) = \prod_{i=1}^{n_p} p_i \prod_{j=1}^{n_q} q_j \prod_{k=1}^{n_r} r_k,$$

where $p_i \in \{p \text{ prime} \mid (\frac{D_0}{p}) = 0\}$, $q_j \in \{p \text{ prime} \mid (\frac{D_0}{p}) = 1\}$ and $r_k \in \{p \text{ prime} \mid (\frac{D_0}{p}) = \ -1\ \}$. Then we have the factorization

$$\mathfrak{a} = \prod_{i=1}^{n_p} \mathfrak{p}_i \prod_{j=1}^{n_q} \mathfrak{q}_j \prod_{k=1}^{n_r} (r_k),$$

where $\mathfrak{p}_i$ is a prime $\mathcal{O}$-ideal such that $\overline{\mathfrak{p}_i} = \mathfrak{p}_i, N(\mathfrak{p}_i) = p_i$ and $\mathfrak{q}_j$ is a prime $\mathcal{O}$-ideal such that $\overline{\mathfrak{q}_j} \neq \mathfrak{q}_j, N(\mathfrak{q}_j) = q_j$. The principal ideals (r_k) are prime $\mathcal{O}$-ideals as well.

Proof. By Theorem 3.20 and Theorem 3.22, the $\mathcal{O}_K$-ideal $\mathfrak{a}\mathcal{O}_K$ has norm $N(\mathfrak{a})$ and a unique factorization into prime ideals. Theorem 3.24 and Lemma 3.17 together imply that each prime $\mathcal{O}_K$-ideal has prime norm. Applying the norm on the factorization of $\mathfrak{a}\mathcal{O}_K$ shows that each prime ideal corresponds bijectively to a prime in the factorization of $N(\mathfrak{a})$. The claim now follows by Theorem 3.20, applying the isomorphism $I \mapsto I \cap \mathcal{O}$, which commutes with complex conjugation and preserves norms and principal ideals generated by rational integers, since $\mathbb{Z} \subset \mathcal{O} \subset \mathcal{O}_K$, so that $(r\mathcal{O}_K) \cap \mathcal{O} = r\mathcal{O}$. □

3.4 Orders and quadratic forms

Remark 3.27. By our definition, we easily see that the product of two fractional $\mathcal{O}$-ideals is again a fractional $\mathcal{O}$-ideal. Thus, the set of invertible ideals forms a group, inside which we find the subgroup of principal fractional ideals.

Definition 3.28. The *ideal class group* $\mathfrak{C}(\mathcal{O}_D) = \mathfrak{C}(D)$ of the imaginary quadratic order $\mathcal{O}_D$ is defined as the quotient of the group of proper fractional $\mathcal{O}_D$-ideals and the subgroup of principal fractional $\mathcal{O}_D$-ideals.

In the following let $\mathbb{H} = \{z \in \mathbb{C} \mid \Im z > 0\}$ denote the upper half plane. We now finally exhibit the previously advertised correspondence between forms and ideals.

Theorem 3.29. *Let $\mathcal{O} = \mathcal{O}_D$ be the order of discriminant $D < 0$ in the imaginary quadratic field K. Then the map sending a proper ideal $\mathfrak{a} = [\alpha, \beta]$ with $\beta/\alpha \in \mathbb{H}$ to the quadratic form $f(x, y) = N(x\alpha + y\beta)/N(\mathfrak{a})$ induces a bijection $\Psi : \mathfrak{C}(D) \longrightarrow \mathfrak{F}_D$.*

Proof. We first show that the map is well-defined. Let $\mathfrak{a}$ be a proper $\mathcal{O}$-ideal and choose a basis $\mathfrak{a} = [\alpha, \beta]$ such that $\tau = \beta/\alpha \in \mathbb{H}$, switching α and β if necessary. Let $ax^2 + bx + c$ be the minimal polynomial of τ with a, b, c coprime integers and we may assume $a > 0$, so that $\tau = (-b + \sqrt{b^2 - 4ac})/2a$. If we define

$$\begin{aligned} f(x, y) = N(x\alpha + y\beta)/N(\mathfrak{a}) &= \frac{(x\alpha + y\beta)(x\overline{\alpha} + y\overline{\beta})}{N(\mathfrak{a})} \\ &= \frac{\alpha\overline{\alpha}}{N(\mathfrak{a})}x^2 + \frac{\alpha\overline{\beta} + \overline{\alpha}\beta}{N(\mathfrak{a})}xy + \frac{\beta\overline{\beta}}{N(\mathfrak{a})}y^2, \end{aligned}$$

then $f(\tau, 1) = 0$. Note that the coefficient of x^2 in $f(x, 1)$ is $\alpha\overline{\alpha}/N(\mathfrak{a}) = N(\alpha)/N(\mathfrak{a}) = a$, by the formula (3.3). It follows that $f(x, 1) = ax^2 + bx + c$ and thus f is a primitive integral form.

To show that f has discriminant D, note first that Lemma 3.16 implies the equality $\mathcal{O} = [1, \omega_D] = [1, a\tau]$. Since $\omega_D \in [1, a\tau]$, there exist $x, y \in \mathbb{Z}$ such that

$$\omega_D = \frac{\sigma_D + \sqrt{D}}{2} = x + y\frac{-b + \sqrt{b^2 - 4ac}}{2}$$

and it follows that $\sqrt{D} = y\sqrt{b^2 - 4ac}$. Since $a\tau \in [1, \omega_D]$, we find analogously that $a\tau = y'\sqrt{D}$ for some $y' \in \mathbb{Z}$. We deduce that $D = b^2 - 4ac$, so that f is a primitive positive definite form of discriminant D.

We have established that our map has indeed a well-defined codomain. We now need to show that Ψ depends neither on the particular representative of an ideal class, nor on the basis of the representative. For the latter we need the following:

Lemma 3.30. *If a lattice has two bases (α, β) and (α', β') with $\beta/\alpha, \beta'/\alpha' \in \mathbb{H}$, then there exists a matrix $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ such that*

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix}$$

Proof. Since $[\alpha', \beta'] \subset [\alpha, \beta]$, the exist integers p, q, r, s such that $\alpha' = p\alpha + q\beta$ and $\beta' = r\alpha + s\beta$. Conversely, since $[\alpha, \beta] \subset [\alpha', \beta']$, we deduce that the matrix $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ has an inverse with integer entries, so that $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{GL}(2, \mathbb{Z})$, implying that $ps - rq = \pm 1$. We may also infer that

$$\frac{p\frac{\beta}{\alpha} + q}{r\frac{\beta}{\alpha} + s} = \frac{\beta'}{\alpha'}$$

and one can easily compute that

$$\Im\left(\frac{p\frac{\beta}{\alpha} + q}{r\frac{\beta}{\alpha} + s}\right) = \det\begin{pmatrix} p & q \\ r & s \end{pmatrix} \frac{\Im\left(\frac{\beta}{\alpha}\right)}{|r\frac{\beta}{\alpha} + s|^2}$$

Since both β/α and β'/α' lie in the upper half plane, it follows that $\det\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) = 1$. □

By the lemma above, if we choose another basis of $\mathfrak{a}$ we find $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that this basis has the form $(p\alpha + q\beta, r\alpha + s\beta)$. This maps to the form $f(px + qy, rx + sy)$, which is equivalent to f. Thus Ψ does not depend on the choice of basis for the ideal. To show that it only depends on the class of the ideal, let $\lambda \in K$. Then the function

$$(x, y) \mapsto \frac{N(x\lambda\alpha + y\lambda\beta)}{N(\lambda\mathfrak{a})} = \frac{N(\lambda)N(x\alpha + y\beta)}{N(\lambda)N(\mathfrak{a})} = \frac{N(x\alpha + y\beta)}{N(\mathfrak{a})}$$

is identical to f. This proves that Ψ is well-defined.

To show that it is surjective, let $f(x, y) = ax^2 + bxy + cy^2$ be a primitive positive definite form of discriminant $D < 0$. Let $\tau = \frac{-b+\sqrt{D}}{2a} \in \mathbb{H}$ be the root of $f(x, 1)$ that lies in the upper half plane. If $\mathfrak{a} = [1, \tau]$, then $\mathfrak{a}$ is a proper fractional ideal of the order $[1, a\tau]$, by Lemma 3.16. Since $D = b^2 - 4ac$, we have the congruence $D \equiv b \pmod 2$. Now it follows easily that $[1, a\tau] = [1, \frac{-b+\sqrt{D}}{2}] = [1, \omega_D] = \mathcal{O}_D$, so that $a\mathfrak{a} = [a, a\tau]$ is a proper $\mathcal{O}_D$-ideal. By (3.3), we have $N(a\mathfrak{a}) = a$ and thus the class of $\mathfrak{a}$ is mapped to the form

$$(x, y) \mapsto \frac{N(xa + ya\tau)}{N(a\mathfrak{a})} = \frac{a^2}{a} \cdot \left(x^2 + \frac{b}{a}xy + \frac{c}{a}y^2\right) = f(x, y),$$

using the Vieta formulas $\tau + \overline{\tau} = b/a$ and $\tau\overline{\tau} = c/a$. Hence our map is surjective.

For injectivity, let $\mathfrak{a} = [\alpha, \beta]$ be an $\mathcal{O}$-ideal with $\beta/\alpha \in \mathbb{H}$. If $\mathfrak{a}$ maps to the form $f(x, y) = ax^2 + bxy + cy^2$ and $[a, a\tau]$ is the preimage of f found in the paragraph above, then $\tau = \beta/\alpha$ since both are roots of $f(x, 1)$ in the upper half plane and such roots are unique. Thus $[a, a\tau] = [a, a\beta/\alpha] = a\alpha^{-1}\mathfrak{a}$, so that $[a, a\tau]$ is equivalent to $\mathfrak{a}$. □

Remark 3.31. Using the bijection in the theorem above, we may transfer the group structure of $\mathfrak{C}(D)$ to the form class group $\mathfrak{F}_D$, which now justifies the terminology. The identity element in $\mathfrak{F}_D$, i.e. the image $\Psi([1, \omega_D]) = \Psi(\mathcal{O}_D)$, is given by the principal form (recall Remark 2.3). Using Corollary 2.11 we deduce the following highly non-trivial result.

Corollary 3.32. *The ideal class group $\mathfrak{C}(D)$ is finite.*

For a primitive positive definite form $f(x, y) = ax^2 + bxy + cy^2$, we have seen that a representative of the preimage $\Psi^{-1}(f)$ is the proper ideal $\mathfrak{a} = [a, a\tau]$ for $\tau = (-b + \sqrt{D})/2a$. Recall from ring theory that we call two elements $\xi, \xi' \in \mathcal{O}_D$ *associated* if there is a unit $u \in \mathcal{O}_D$ such that $\xi' = u\xi$. We can map pairs of integers $(x, y) \in \mathbb{Z}^2$ to elements $xa + ya\tau \in \mathfrak{a}$ and we call two pairs *equivalent* if the corresponding elements in $\mathfrak{a}$ are associated[2]. The usefulness of this map becomes obvious in the proof of the following theorem.

Theorem 3.33. *Let $f(x, y) = ax^2 + bxy + cy^2$ be a primitive positive definite form of discriminant $D < 0$. There is a bijection between inequivalent solutions of $f(x, y) = n > 0$ and proper ideals of norm n in the class $(\Psi^{-1}(f))^{-1} =: C_f$.*

Proof. Let $\tau = (-b + \sqrt{D})/2a$ and $\mathfrak{a} = [a, a\tau] \in \Psi^{-1}(f)$. The proof of Theorem 3.29 shows that

$$f(x, y) = \frac{N(xa + ya\tau)}{N(\mathfrak{a})}.$$

The map $(x, y) \mapsto xa + ya\tau$ gives a bijection between the solutions of $f(x, y) = n$ and elements $\xi \in \mathfrak{a}$ with $N(\xi) = nN(\mathfrak{a})$. It suffices to show that the map $\xi \mapsto \xi\mathfrak{a}^{-1}$ induces a bijection between non-associated elements $\xi \in \mathfrak{a}$ with $N(\xi) = nN(\mathfrak{a})$ and proper $\mathcal{O}_D$-ideals in C_f with norm n.

[2] This is the definition given in [BS66, p. 143, Sec. 7.6]. In [Zag81, p. 58] we find another definition that is perhaps more enlightening, which can be seen with some extra work to be equivalent to ours. Namely, two solutions $(x_1, y_1), (x_2, y_2)$ are equivalent if they can be transformed into one another as in Remark 2.13 by $\mathrm{SL}_2(\mathbb{Z})$ matrices which leave the form invariant, that is, if $\gamma\binom{x_1}{y_1} = \binom{x_2}{y_2}$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma f = f$. This definition also shows that the equivalence relation can be restricted to the set of proper representations.

Let $\xi \in \mathfrak{a}$ have norm $nN(\mathfrak{a})$. Then the fractional ideal $I = \xi\mathfrak{a}^{-1}$ lies in $\mathcal{O}_D$ since $I\mathfrak{a} = \xi\mathfrak{a}^{-1}\mathfrak{a} = \xi\mathcal{O}_D \subset \mathfrak{a}$ and $\mathfrak{a}$ is proper. Furthermore, I is in the class C_f, because obviously $\mathfrak{a}^{-1} \in (\Psi^{-1}(f))^{-1}$, and the norm of I is $N(\xi)N(\mathfrak{a})^{-1} = n$.

Conversely, let $I \in C_f$ be an $\mathcal{O}_D$-ideal with norm n. Since $\mathfrak{a} \in C_f^{-1}$, there is an element $\xi \in \mathcal{O}_D$, unique up to association, such that $I\mathfrak{a} = \xi\mathcal{O}_D$. It follows that $I = \xi\mathfrak{a}^{-1}$, for $\xi \in \mathfrak{a}$ (because $\mathfrak{a}$ is an ideal), and $N(\xi) = nN(\mathfrak{a})$. □

Definition 3.34. For a primitive positive definite form f we denote the number of inequivalent solutions of the equation $f(x, y) = n$ by $r_f(n)$ and the number of inequivalent proper representations by $r_f^*(n)$.

Remark 3.35. Note that equivalence of representations stems from the action of the group of units in $\mathcal{O}_D$. If w_D denotes the number of units in $\mathcal{O}_D$, then we have $r_f(n) = R(f, n)/w_D$ and $r_f^*(n) = R^*(f, n)/w_D$.

3.5 Ambiguous classes

The last section of this chapter presents results about particular classes of ideals, called ambiguous, which are in general easier to handle. By Theorem 3.33 we may speak of ideals and forms concomitantly.

Definition 3.36. A class of ideals (or of forms) is called *ambiguous* if it has order 1 or 2 in the ideal (or form) class group. We denote the subgroup of ambiguous classes by $\mathfrak{G}(D) \subset \mathfrak{C}(D)$.

Remark 3.37. We will also view $\mathfrak{G}(D)$ as a subgroup of $\mathfrak{F}_D$ in virtue of Theorem 3.33. There will be in the rest of this thesis no cause for confusion in doing so.

We have seen that the inverse of the class of $\mathfrak{a} \subset \mathcal{O}$ is the class of $\overline{\mathfrak{a}}$, since $\mathfrak{a}\overline{\mathfrak{a}} = N(\mathfrak{a})\mathcal{O}$. Using the bijection of Theorem 3.33, we can compute that the inverse class of a form $f(x, y) = ax^2 + bxy + cy^2$ is the class of the so-called *opposite* form $\overline{f}(x, y) = ax^2 - bxy + cy^2$. Indeed, $[1, (-b + \sqrt{D})/2] \in \Psi^{-1}(f)$ and $\overline{[1, (-b + \sqrt{D})/2]} = [1, (-b - \sqrt{D})/2] = [1, (b + \sqrt{D})/2]$, which is mapped to $\overline{f}$ by Ψ. Now this allows us to describe ambiguous form classes explicitly using the special representatives we discovered in Theorem 2.9.

Lemma 3.38. *The class of a reduced form $f(x, y) = ax^2 + bxy + cy^2$ of discriminant $D < 0$ is ambiguous if and only if $b = 0, a = b$ or $a = c$.*

Proof. The class of f is ambiguous if and only if the forms f and $\overline{f}$ are equivalent. If $|b| < a < c$, then $\overline{f}$ is also reduced, so that, by Theorem 2.9, f and $\overline{f}$ are equivalent if and only if $b = 0$. In the case that $a = b$ we easily see the equivalence $\overline{f} = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right) f$ as in the proof of Theorem 2.9. Finally, if $a = c$, then $\overline{f} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) f$. □

We will later also need to know what the number of ambiguous classes is.

Theorem 3.39. *Let $D < 0$ be a quadratic discriminant and let r be the number of odd primes dividing D. Define μ as follows: if $D \equiv 1 \pmod 4$, then $\mu = r$, and if $D \equiv 0 \pmod 4$ and $D = -4n$ with $n > 0$, then*

$$\mu = \begin{cases} r, & n \equiv 3 \pmod 4, \\ r+1, & n \equiv 1, 2 \pmod 4 \text{ or } n \equiv 4 \pmod 8, \\ r+2, & n \equiv 0 \pmod 8. \end{cases}$$

The class group $\mathfrak{F}_D$ has exactly $2^{\mu-1}$ ambiguous classes, that is, $|\mathfrak{G}(D)| = 2^{\mu-1}$.

Proof. By Theorem 2.9, we only need to count the reduced forms with the special properties from Lemma 3.38. Some of the computations and case-work can be seen in [Cox13, Prop. 3.11]. □

4 The main theorem

We now turn our attention to the task of estimating the number of representations and state the central result of this thesis. For convenience, we make the following convention that in all sums of the form $\sum_{k\leq x}$ the variable k ranges over the subset of the natural numbers $\{k \in \mathbb{N} \mid k \leq x\}$. Moreover, raising to the power 0 is defined as

$$k^0 = \begin{cases} 0, & \text{if } k = 0; \\ 1, & \text{otherwise.} \end{cases}$$

Theorem 4.1 (Main theorem)**.** *Let g be a binary quadratic form of discriminant $D = D_0 f^2 < 0$ with conductor f. For each divisor d of f, let $a_{\theta_d(g)}$ be the smallest positive integer represented by $\theta_d(g)$ and let $u_{\theta_d(g)}$ be the smallest positive integer coprime to f/d that can be represented by some form in the coset $\theta_d(g)\mathfrak{G}(D/d^2)$. For any $\beta \geq 0$ we have:*

$$\sum_{n\leq x} r_g(n)^\beta = \frac{2}{w_D} \cdot \frac{\pi x}{\sqrt{D}} \left(1 + (2^{\beta-1} - 1) \sum_{d|f} \frac{\varphi(f/d)}{f u_{\theta_d(g)}}\right) + E_\beta(x, D),$$

where

$$E_\beta(x, D) \ll \begin{cases} \displaystyle\sum_{d|f} \frac{2^{\omega(f/d)}}{d} \sqrt{\frac{x}{a_{\theta_d(g)}}} + \tau(f^2) + \tau(D) \left(\frac{x \log x}{D} + \frac{x}{D^{3/4}}\right), & 0 \leq \beta \leq 2, \\ \displaystyle\sum_{d|f} \frac{2^{\omega(f/d)}}{d} \sqrt{\frac{x}{a_{\theta_d(g)}}} + \tau(f^2) + \tau(D) \frac{x(\log x)^{(2/q)(2^{(\beta-2)q+1}-1)+1}}{D^{(3/4)(1-1/q)}}, & \beta > 2, \end{cases}$$

for any real $q > 1$, where $\tau(D)$ denotes the number of divisors of D. The implied constants depend at most on β and q.

The next sections present various results about representation numbers. Although most are interesting in their own right, we will ultimately use them all for proving the main theorem.

4.1 A reduction theorem for the number of representations

The correspondence between representations and ideals, i.e. Theorem 3.33, and factorization into prime ideals, i.e. Corollary 3.26, give us a tool for handling representations of numbers coprime to the conductor. In order to gain some control in the case of other numbers as well, this section shows how to reduce the general problem to the coprime case. The following results have been stated by Sun and Williams in [SW06].

Lemma 4.2. *Let D be a discriminant with conductor f, $m \in \mathbb{N}$ and $C \in \mathfrak{C}(D)$.*

1. *There exist integers a, b, c such that $C = [a, b, c]$ with $(a, m) = 1$.*
2. *If $m \mid f$, then there exist integers a, b and c such that $C = [a, bm, cm^2]$, a is coprime to m, and $b \equiv D \pmod 2$.*

Proof. 1. This follows from Lemma 2.16 and Lemma 2.15 together.

2. By Lemma 2.15, choose integers a, b_1, c_1 such that $C = [a, b_1, c_1]$ and $(a, m) = 1$. Setting $\Delta = D/m^2 \in \mathbb{Z}$, we have $b_1^2 - 4ac_1 = \Delta m^2$ and thus $b_1 \equiv \Delta m \pmod 2$. Let u and v be integers such that $b_1 = \Delta m + 2u$ and $va \equiv -u \pmod m$, where we use that $(a, m) = 1$. Then $b_1 + 2va \equiv b_1 - 2u = \Delta m \pmod{2m}$ and, in particular, $m \mid b_1 + 2va$. Choosing $b \in \mathbb{Z}$ such that $b_1 + 2va = bm$, we have the congruence $bm \equiv \Delta m \pmod{2m}$, hence $b \equiv \Delta \pmod 2$, and

$$\begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} (ax^2 + b_1xy + c_1y^2) = ax^2 + (b_1 + 2av)xy + c_2y^2 = ax^2 + bmxy + c_2y^2,$$

with $c_2 \in \mathbb{Z}$ and $4ac_2 = (b^2 - \Delta)m^2$. Since $4 \mid b^2 - \Delta$ and $(a, m) = 1$, it follows that $m^2 \mid c_2$. Writing $c_2 = cm^2$ we obtain the claim. □

Lemma 4.3. *Let $a, b, c \in \mathbb{Z}$ and $m, n \in \mathbb{N}$ with $(a, m) = 1$ and $m^2 \mid n$. If $n = ax^2 + bmxy + cm^2y^2$ for $x, y \in \mathbb{Z}$, then $m \mid x$.*

Proof. By (2.2) we have $(2ax + bmy)^2 = 4an + (b^2 - 4ac)m^2y^2$ and, since $m^2 \mid n$, this implies that $m \mid 2ax$, so that $\frac{m}{(2,m)} \mid x$. Hence $ax^2 = n - bmxy - xm^2y^2 \equiv 0 \pmod{\frac{m^2}{(2,m)}}$ and thus $m \mid x$. □

Lemma 4.4. *Let D be a discriminant with conductor f and let $m \in \mathbb{N}$ such that $m|f$. The map $\theta_m : \mathfrak{C}(D) \longrightarrow \mathfrak{C}(D/m^2), [a, bm, cm^2] \mapsto [a, b, c]$ is well-defined.*

Proof. For an arbitrary class $C \in \mathfrak{C}(D)$, Lemma 4.2 gives us $a, b, c \in \mathbb{Z}$ with $(a, m) = 1$ such that $f(x, y) = ax^2 + bmxy + cm^2y^2$ is a form in the class C. Then it is obvious that $ax^2 + bxy + cy^2$ is a primitive positive definite form of discriminant $b^2 - 4ac = D/m^2$.

Now let $g(x', y') = a'x^2 + b'mxy + c'm^2y^2$ be in the class C as well, so that there exists $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ with $g = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) f$. As one can easily compute (see [Zag81, p. 58]), it follows that $m^2c' = f(q,s)$ and by Lemma 4.3 we have $m \mid q$. Finally, one can check that

$$a'x^2 + b'xy + c'y^2 = \begin{pmatrix} p & \frac{q}{m} \\ rm & s \end{pmatrix} (ax^2 + bxy + xy^2),$$

proving that the map is well-defined. □

We may now state and prove Theorem 3.2 from [SW06], which will be an important tool in the proof of the main theorem.

Theorem 4.5 (Reduction Theorem)**.** *Let $D = f^2D_0$ be a quadratic discriminant with conductor f. Let $n \in \mathbb{N}$ and $C \in \mathfrak{C}(D)$. Then*

$$R(C,n) = \begin{cases} 0 & \text{if } (n,f^2) \text{ is not a square;} \\ R(\theta_m(C), n/m^2) & \text{if } (n,f^2) = m^2 \text{ for } m \in \mathbb{N}. \end{cases}$$

Proof. By Lemma 4.2 we may assume that $C = [a,b,c]$ with $(a,f) = 1$. Suppose $R(C,n) > 0$, so that $n = ax^2 + bxy + cy^2$ for some $x, y \in \mathbb{Z}$. We can reformulate this to $4an = (2ax+by)^2 - Dy^2$. Since $(a,f) = 1$ and $f^2 \mid D$, it follows that $(4n, f^2) = (4an, f^2) = ((2ax+by)^2, f^2) = u^2$ for some $u \in \mathbb{Z}$. Hence (n,f^2) is a square if $\nu_2(n) \neq \nu_2(f^2) - 1$, where $\nu_2(n)$ is the exponent of 2 in the factorization of n. Now assume $\nu_2(n) = \nu_2(f^2) - 1$, which implies that $2 \mid f$ and $2 \nmid a$. Let $n = 2^{\nu_2(n)}n_0, f = 2^{\nu_2(f)}f_0$ with $2 \nmid n_0, 2 \nmid f_0$. Then dividing by $2^{\nu_2(f^2)} = 2^{2\nu_2(f)}$ yields

$$2an_0 = ((2ax+by)/2^{\nu_2(f)})^2 - f_0^2D_0y^2.$$

Since $D_0 \equiv 0, 1 \pmod 4$, we see that the right hand side cannot be congruent to 2 modulo 4. However, $2an_0 \equiv 2 \pmod 4$, so this is a contradiction. Thus, (n,f^2) is always a square if n is representable.

Now suppose $(n,f^2) = m^2$ for some $m \in \mathbb{N}$. Since $m \mid f$, we may assume that $C = [a, mb, m^2c]$ with $(a,m) = 1$. If $n = ax^2 + bmxy + cm^2y^2$ for some $(x,y) \in \mathbb{Z}$, then $m \mid x$ by Lemma 4.3. Thus $n/m^2 = aX^2 + bXy + cy^2$ for $X = x/m \in \mathbb{Z}$. Conversely, if $n/m^2 = aX^2 + bXy + cy^2$ for some $X, y \in \mathbb{Z}$, then (mX, y) is a solution to $n = ax^2 + bmxy + cm^2y^2$. □

Remark 4.6. If $(n,f^2) = m^2$, then n/m^2 is coprime to f/m, that is, to the conductor of D/m^2. Thus we reduced the general case to the coprime case.

4.2 Preliminary estimates

Finding the exact number of representations for arbitrary numbers and forms is difficult. Precise formulas have been found for special cases (see for instance [SW06]), but in this thesis we are concerned only with giving good estimates and asymptotics. This section gathers some weaker results that we will use in the proof of the main theorem. Recall the notation $r_g(n)$ for the number of inequivalent representations from Definition 3.34.

Lemma 4.7. *For a primitive positive definite form g of discriminant $D = f_D^2 D_0$ and an integer $n \in \mathbb{Z}$ coprime to the conductor f_D, we have*

$$r_g(n) \leq \tau(n),$$

where $\tau(n)$ is the number of divisors of n.

Proof. By the correspondence between representations and ideals, i.e. Theorem 3.33, the number of inequivalent representations of n is certainly bounded by the number of all $\mathcal{O}_D$-ideals having norm n. We can factorize these ideals uniquely into primes. More precisely, if $n = \prod p_i \prod q_j \prod r_k$ as in the statement of Corollary 3.22, then the ideals corresponding to representations of n have the form $\mathfrak{a} = \prod \mathfrak{p}_i \prod \mathfrak{q}_j \prod (r_k)$, where $\mathfrak{p} = \overline{\mathfrak{p}}$ and $\mathfrak{q} \neq \overline{\mathfrak{q}}$. We see that the products $\prod \mathfrak{p}$ and $\prod (r_k)$ are unique. Each different prime q dividing n, such that $(\frac{D_0}{q}) = 1$, contributes to the product $\prod \mathfrak{q}_j$ by a factor of the form $\mathfrak{q}^\alpha \cdot \overline{\mathfrak{q}}^\beta$, where $\alpha + \beta$ equals the maximal exponent of q in the factorization of n, i.e., $\alpha + \beta = \nu_q(n)$. There are thus $\nu_q(n) + 1 = \tau(q^{\nu_q(n)})$ such different possible factors. Considering all such primes q and the fact that the divisor function τ is multiplicative, we find that there are at most $\tau(\prod q^{\nu_q(n)}) \leq \tau(n)$ different ideals with norm n. □

We will later use the inequality provided by the former lemma, but embellished with a positive exponent. Since we want to sum over all positive integers up to some x, we will also need the following estimate.

Lemma 4.8. *Let $\alpha > 0$. Then*

$$\sum_{k \leq x} \tau(k)^\alpha \ll x (\log x)^{2^\alpha - 1}.$$

Proof. The statement was proved in [Del71] with elementary methods inspired by Paul Erdős. A stronger asymptotic result can be proved using powers of the Riemann zeta function as in the Selberg-Delange method, explained in [Ten95, Chapter II.5]. □

The following lemmata culminate with an asymptotic for the sum of representation numbers for the integers smaller than a given x and coprime to the conductor.

Lemma 4.9. *For $f \in \mathbb{N}$ we have the identities*

$$\sum_{d|f} \frac{\mu(d)}{d} = \frac{\varphi(f)}{f}, \quad \sum_{d|f} \frac{\varphi(d)}{f} = 1, \quad \textit{and} \quad \sum_{d|f} 2^{\omega(d)} = f^2,$$

where μ is the Möbius function, φ is Euler's totient function and $\omega(a)$ the number of prime divisors of $a \in \mathbb{Z}$.

Proof. The first two formulas follow directly or by Möbius transformation from the well-known convolution identity $\sum_{d|f} \varphi(d) = f$. For the last one, note that 2^ω is a multiplicative function, i.e., $2^{\omega(nm)} = 2^{\omega(n)+\omega(m)} = 2^{\omega(n)}2^{\omega(m)}$ whenever $(n, m) = 1$. Therefore, we may check the convolution identity $\sum_{d|f} 2^{\omega(d)} = f^2$ on prime powers. Indeed, for a prime p and $\alpha \in \mathbb{N}_0$ we have

$$\sum_{d|p^\alpha} 2^{\omega(d)} = \sum_{k=0}^{\alpha} 2^{\omega(p^k)} = 1 + 2\alpha = \tau(p^{2\alpha}).$$

□

Lemma 4.10. *For $a \in \mathbb{N}$ and $x > 0$ we have the estimate*

$$\sum_{\substack{1 \le n \le x \\ (n,a)=1}} 1 = \frac{\varphi(a)}{a} x + O\left(2^{\omega(a)}\right).$$

Proof. By Lemma 4.9 and Möbius inversion we have

$$\begin{aligned}\sum_{\substack{1 \le n \le x \\ (n,a)=1}} 1 &= \sum_{n \le x} \sum_{d|(n,a)} \mu(d) \\ &= \sum_{d|a} \mu(d) \left[\frac{x}{d}\right] = \sum_{d|a} \frac{\mu(d)x}{d} + O\left(2^{\omega(a)}\right) = \frac{\varphi(a)}{a} x + O\left(2^{\omega(a)}\right),\end{aligned}$$

since there are $2^{\omega(a)}$ square-free divisors of a. □

Lemma 4.11. *If g is a positive definite primitive quadratic form of discriminant $-D < 0$*

with conductor f and a is the smallest integer represented by g, then

$$\#\{(m,n) \in \mathbb{Z}^2 \mid g(m,n) \leq x, (g(m,n), f) = 1\} = \frac{\varphi(f)}{f} \cdot \frac{2\pi x}{\sqrt{D}} + O\left(2^{\omega(f)}\left(1 + \sqrt{\frac{x}{a}}\right)\right).$$

The implied constant is absolute.

Proof. First, we need to characterize representations of numbers coprime to the conductor. Since the number of representations depends only on the class of a form, we may assume that $g(x_1, x_2) = ax_1^2 + b'x_1x_2 + c'x_2^2$ is reduced. In particular, a is the smallest integer represented by g and it will be useful to note that $a \ll \sqrt{D}$ as in (2.3). Moreover, if $f = f_a\tilde{f}$ is a decomposition of the conductor such that $(a, \tilde{f}) = 1$ and all primes dividing f_a divide a as well, then assume by Lemma 4.2 that $g(x_1, x_2) = ax_1^2 + b\tilde{f}x_1x_2 + c\tilde{f}^2x_2^2$.

By Theorem 4.5, since a is represented by g, we have $(a, f^2) = (a, f_a^2) = k^2$ for some $k \in \mathbb{Z}$. By our construction of f_a we find that all primes p dividing f_a must satisfy $p^2 \mid a$ and, since $D = b^2 - 4ac$ and g is primitive, p also divides b and does not divide c. Now let $g(x_1, x_2) = n$ for some $n \in \mathbb{N}$ and suppose p is a prime such that $p \mid (n, f)$. If $p \mid f_a$, then $(c\tilde{f}^2, p) = 1$, $p \mid b$ and $p^2 \mid a$, so that from Lemma 4.3 it follows that $p \mid x_2$. If $p \mid \tilde{f}$, then $p \mid x_1$, again by Lemma 4.3. It follows that $(g(x_1, x_2), f) = 1$ if and only if $(x_1, \tilde{f}) = 1$ and $(x_2, f_a) = 1$.

Consequently, using (2.2), we need to count the number of integer solutions to $(2am + bn)^2 + Dn^2 \leq 4ax$, where m is coprime to $\tilde{f}$ and n is coprime to f_a. The inequality is equivalent to the two conditions $|n| \leq \sqrt{4ax/D}$ and

$$\frac{-\sqrt{4ax - Dn^2} - bn}{2a} \leq m \leq \frac{\sqrt{4ax - Dn^2} - bn}{2a}.$$

By Lemma 4.10, there are

$$\frac{\varphi(\tilde{f})}{\tilde{f}} \cdot \frac{\sqrt{4ax - Dn^2}}{a} + O\left(2^{\omega(\tilde{f})}\right) \tag{4.1}$$

integers m coprime to $\tilde{f}$ in this range. There are $2\frac{\phi(f_a)}{f_a}\sqrt{4ax/D} + O(2^{\omega(f_a)})$ possible values for n; summing the expression (4.1) over these, the $O(2^{\omega(\tilde{f})})$ errors add up to $O(2^{\omega(\tilde{f})}\sqrt{ax/D} + 2^{\omega(f_a)}2^{\omega(\tilde{f})}) = O(2^{\omega(\tilde{f})}\sqrt{x/a} + 2^{\omega(f)})$, since $a \ll \sqrt{D}$ and 2^{ω} is multiplicative. Next we compute the sum

$$\sum_{\substack{|n|\leq\sqrt{4ax/D}\\(n,f_a)=1}}\frac{\sqrt{4ax-Dn^2}}{a}=\sum_{|n|\leq\sqrt{4ax/D}}\frac{\sqrt{4ax-Dn^2}}{a}\sum_{d|(n,f_a)}\mu(d)$$
$$=\sum_{d|f_a}\mu(d)\sum_{|k|\leq\sqrt{4ax/D}/d}\frac{\sqrt{4ax-Dd^2k^2}}{a}.$$

We approximate the inner sum by the corresponding integral:

$$\int_{-\sqrt{4ax/D}/d}^{\sqrt{4ax/D}/d}\frac{\sqrt{4ax-Dd^2t^2}}{a}dt.$$

Since the integrand is decreasing from zero to either endpoint, the error made in the approximation is at most twice the value at zero, i.e. $O(\sqrt{x/a})$. To evaluate the integral, we make the change of variable $v=t\sqrt{4ax/D}/d$, obtaining

$$\frac{(4x)}{d\sqrt{D}}\int_{-1}^{1}\sqrt{1-v^2}dv=\frac{2\pi x}{d\sqrt{D}}.$$

Therefore we have

$$\frac{\varphi(\tilde{f})}{\tilde{f}}\cdot\sum_{\substack{|n|\leq\sqrt{4ax/D}\\(n,f_a)=1}}\frac{\sqrt{4ax-Dn^2}}{a}=\frac{\varphi(\tilde{f})}{\tilde{f}}\cdot\sum_{d|f_a}\mu(d)\left(\frac{2\pi x}{d\sqrt{D}}+O\left(\frac{x}{a}\right)\right),$$

where we note that $\varphi(\tilde{f})/\tilde{f}\leq 1$. We use Lemma 4.9 to compute $\sum_{d|f_a}\mu(d)/d=\phi(f_a)/f_a$. Noting that $\frac{\varphi(\tilde{f})}{\tilde{f}}\cdot\frac{\varphi(f_a)}{f_a}=\frac{\varphi(f)}{f}$, by multiplicativity of φ, yields the claimed main term. Gathering the error terms gives an error of

$$O\Big(2^{\omega(\tilde{f})}\sqrt{x/a}+2^{\omega(f)}+2^{\omega(f_a)}\sqrt{x/a}\Big)=O\Big(2^{\omega(f)}\Big(1+\sqrt{x/a}\Big)\Big),$$

using that $\sum_{d|f_a}\mu(d)\leq 2^{\omega(f_a)}$. □

Remark 4.12. In the special case when $f=1$ we can improve the error term to $O(\sqrt{x/a})$, since now the possible values of n are bounded by a multiple of $\sqrt{ax/D}$. Consequently, when we add the $O(2^{\omega(1)})=O(1)$ errors in equation (4.1) together, we obtain the claimed bound for the error. We refer to Lemma 3.1 in [BG06] for the simplified proof in this special case.

4.3 Proof of the main theorem

The strategy in the proof of the main theorem is to first prove a special result, counting the representations of numbers coprime to the conductor. Afterwards we will use the reduction theorem from section 4.1 to extend this count to all positive integers.

Theorem 4.13. *For a binary quadratic form g having discriminant $-D = D_0 f_D^2 < 0$ with conductor f_D, let a_g be the smallest positive integer that is represented by g, and let u_g be the smallest positive integer coprime to f_D that can be represented by some form in the coset $g\mathfrak{G}(D)$. For any $\beta \geq 0$ we have:*

$$\sum_{\substack{n \leq x \\ (n, f_D)=1}} r_g(n)^\beta = \frac{\varphi(f_D)}{f_D} \cdot \frac{2}{w_D} \left(1 + \frac{2^{\beta-1}-1}{u_g}\right) \frac{\pi x}{\sqrt{D}} + E_\beta(x, D),$$

where

$$E_\beta(x, D) \ll \begin{cases} 2^{\omega(f)} \left(1 + \sqrt{\dfrac{x}{a_g}}\right) + 2^{\omega(D)} \left(\dfrac{x \log x}{D} + \dfrac{x}{D^{3/4}}\right), & 0 \leq \beta \leq 2, \\ 2^{\omega(f)} \left(1 + \sqrt{\dfrac{x}{a_g}}\right) + 2^{\omega(D)} \dfrac{x (\log x)^{(2/q)(2^{(\beta-2)q+1}-1)+1}}{D^{(3/4)(1-1/q)}}, & \beta > 2, \end{cases}$$

for any real $q > 1$, where $\omega(D)$ denotes the number of prime divisors of D. The implied constants depend at most on β and q.

Proof. The proof is an adapted version of the one given in [BG06]. The outline is as follows: we translate representations of integers coprime to the conductor to ideals in the order $\mathcal{O}_D$. We use special factorizations of these ideals into pairs of simpler factors and then estimate the possibilities for each factor. Afterwards we define a counting function that is close to the count we are interested in but easier to compute and then bound the difference, in each range of β, by the estimates achieved above.

First we introduce the useful notion of primitive ideals.

Definition 4.14. An ideal $\mathfrak{a} \subset \mathcal{O}_D$ is called *primitive* if it is not divisible by any rational integer other than 1 or, equivalently, if $e^{-1}\mathfrak{a} \not\subseteq \mathcal{O}_D$ for any $e \in \mathbb{Z}, |e| > 1$.

Now let $\mathfrak{G}(D) \subset \mathfrak{C}(D)$ be the set of ambiguous classes in the order $\mathcal{O}_D$, let $\mathfrak{A}$ be the set of primitive ideals coprime to D[3], and let

$$\mathfrak{X}_G = \{\mathfrak{c} \in G \mid \mathfrak{c} \subset \mathcal{O}_D, \mathfrak{c} \neq \overline{\mathfrak{c}}\}$$

for $G \in \mathfrak{G}(D)$. First we gather some basic properties of ideals in $\mathfrak{A}$.

Lemma 4.15. *If $\mathfrak{a} \in \mathfrak{A}$, then $\mathfrak{a}^2 \in \mathfrak{A}$. For $\mathfrak{a}_1, \mathfrak{a}_2 \in \mathfrak{A}$ in the same class, the principal ideal $\mathfrak{a}_1\overline{\mathfrak{a}_2}$ is generated by a rational integer if and only if $\mathfrak{a}_1 = \mathfrak{a}_2$.*

Proof. We prove both assertions by contradiction. For the first claim suppose that $\mathfrak{a}^2 = k\mathfrak{b}$ for some $k \in \mathbb{Z}$ and an $\mathcal{O}_D$-ideal $\mathfrak{b}$. Assume without loss of generality that k is prime. Then k is coprime to the discriminant D since $N(\mathfrak{a})^2 = k^2 N(\mathfrak{b})$ and $(N(\mathfrak{a}), D) = 1$. By Corollary 3.26 we have the prime ideal decomposition $k\mathcal{O}_D = \mathfrak{p}_1\mathfrak{p}_2$ or $k\mathcal{O}_D = \mathfrak{p}$. Then $\mathfrak{a}^2 = k\mathfrak{b}$ implies that $k \mid \mathfrak{a}$, which is a contradiction to the assumption $\mathfrak{a} \in \mathfrak{A}$.

For the second claim, suppose $\mathfrak{a}_1\overline{\mathfrak{a}_2} = k\mathcal{O}_D$ for $k \in \mathbb{Z}$. As above, k is coprime to D and has a prime ideal factorization. If an inert prime p divides k, then $p \mid \mathfrak{a}_1$ or $p \mid \mathfrak{a}_2$, which leads to a contradiction. If $p\mathcal{O}_D = \mathfrak{p}\overline{\mathfrak{p}}$ is a split prime dividing k, then either $\mathfrak{p}$ or $\overline{\mathfrak{p}}$ divides $\mathfrak{a}_1$ and then $\overline{\mathfrak{p}}$ or $\mathfrak{p}$ divides $\overline{\mathfrak{a}_2}$, respectively. Doing this recursively for all primes dividing k, it follows that $\mathfrak{a}_1 = \mathfrak{a}_2$. □

Next we recall that, by Theorem 3.33, a pair $(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{Z}^2 \times \mathbb{Z}^2$ of inequivalent solutions to $g(\mathbf{x}_1) = g(\mathbf{x}_2) = n \in \mathbb{N}$ corresponds to a pair of different ideals in the class C_g having norm n. For $(n, f) = 1$, these are exactly the pairs of ideals $(\mathfrak{b}\mathfrak{c}, \mathfrak{b}\overline{\mathfrak{c}})$ with $N(\mathfrak{b}\mathfrak{c}) = n$, where $\mathfrak{c} \in \mathfrak{X}_G$ for some $G \in \mathfrak{G}$ and $\mathfrak{b} \in \mathfrak{A}$ is in the class $C_g G$.

Proof. Indeed, we can construct $\mathfrak{b}$ and $\mathfrak{c}$ using the prime ideal factorization as in Corollary 3.26. Two ideals $\mathfrak{a}_1, \mathfrak{a}_2 \in C_g$ with norm n have the form $\mathfrak{a}_1 = \prod \mathfrak{p}_i \prod \mathfrak{q}_j \prod (r_k)$ and $\mathfrak{a}_2 = \prod \mathfrak{p}_i \prod \mathfrak{s}_j \prod (r_k)$, where $\{\mathfrak{p}_i\}_i$ correspond to the ramified primes, $\{\mathfrak{q}_j, \mathfrak{s}_j\}_j$ to the split primes, and $\{r_k\}_k$ are the inert primes. Thus, the two outer products $\prod \mathfrak{p}_i$ and $\prod (r_k)$ are identical in both factorizations, respectively, since $N(\mathfrak{a}_1) = N(\mathfrak{a}_2)$. We want $\mathfrak{b}$ to be primitive and coprime to D, so that $\prod \mathfrak{p}_i \prod (r_k)$ needs to divide $\mathfrak{c}$.

For the rest, let $Q = \{q \text{ prime} : q \mid n, (\frac{D_0}{q}) = 1\}$ and for each $q \in Q$ choose a prime ideal denoted by $\mathfrak{q}$ that contains q. Then the middle products are of the form

$$\prod \mathfrak{q}_j = \prod_{q \in Q} \mathfrak{q}^{i_q} \overline{\mathfrak{q}}^{j_q}, \quad \prod \mathfrak{s}_j = \prod_{q \in Q} \mathfrak{q}^{k_q} \overline{\mathfrak{q}}^{l_q}, \tag{4.2}$$

[3] By definition, an ideal $\mathfrak{a}$ is coprime to D if $\mathfrak{a} + D\mathcal{O}_D = \mathcal{O}_D$. This is equivalent to $\gcd(N(\mathfrak{a}), D) = 1$, as in Remark 3.19.

where $i_q + j_q = k_q + l_q$ for all $q \in Q$. To see how to construct $\mathfrak{b}$ and $\mathfrak{c}$, let us assume without loss of generality that $i_q = \min(i_q, j_q, k_q, l_q)$ for a prime $q \in Q$. Then the factor in (4.2) corresponding to q is of the form

$$\mathfrak{q}^{i_q}\overline{\mathfrak{q}}^{j_q} = (q^{i_q})\overline{\mathfrak{q}}^{j_q - i_q}, \quad \mathfrak{q}^{k_q}\overline{\mathfrak{q}}^{l_q} = (q^{i_q})\mathfrak{q}^{k_q - i_q}\overline{\mathfrak{q}}^{l_q - i_q},$$

using that $\mathfrak{q}\overline{\mathfrak{q}} = q$. Since we want $\mathfrak{b} \in \mathfrak{A}$, the principal ideal (q^{i_q}) should divide $\mathfrak{c}$. Denoting $\mathfrak{b}_q = \overline{\mathfrak{q}}^{l_q - i_q} \subset \mathcal{O}_D$ and $\mathfrak{c}_q = (q^{i_q})\overline{\mathfrak{q}}^m \subset \mathcal{O}_D$ where $m = j_q - i_q - (l_q - i_q) = k_q - i_q \in \mathbb{N}_0$, we have

$$\mathfrak{q}^{i_q}\overline{\mathfrak{q}}^{j_q} = \mathfrak{b}_q\mathfrak{c}_q, \quad \mathfrak{q}^{k_q}\overline{\mathfrak{q}}^{l_q} = \mathfrak{b}_q\overline{\mathfrak{c}_q}.$$

Defining $\mathfrak{b}_q$ and $\mathfrak{c}_q$ analogously for all $q \in Q$ and denoting

$$\mathfrak{b} = \prod_{q \in Q} \mathfrak{b}_q, \quad \mathfrak{c} = \prod_i \mathfrak{p}_i \prod_{q \in Q} \mathfrak{c}_q \prod_k (r_k),$$

we find that $\mathfrak{a}_1 = \mathfrak{b}\mathfrak{c}$ and $\mathfrak{a}_2 = \mathfrak{b}\overline{\mathfrak{c}}$. Lemma 4.15 shows that $\mathfrak{b} \in \mathfrak{A}$, since each $\mathfrak{b}_q$ is primitive by Theorem 3.24. Moreover, $\mathfrak{a}_1 \neq \mathfrak{a}_2$ implies that $\mathfrak{a}_1\mathfrak{b}^{-1} = \mathfrak{c} \neq \overline{\mathfrak{c}} = \mathfrak{a}_2\mathfrak{b}^{-1}$. Since $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are in the same class, there exists $\xi \in \mathcal{O}_D$ such that $\xi\mathcal{O}_D = \mathfrak{a}_1\overline{\mathfrak{a}_2} = N(\mathfrak{b})\mathfrak{c}^2$, and it follows that $\mathfrak{c}$ is in an ambiguous class. Finally, the considerations that lead to the explicit construction make clear that $\mathfrak{b}$ and $\mathfrak{c}$ are unique (here we make implicit use of the uniqueness of prime factorization). □

Let $\mathfrak{u}$ be the ideal in some class C_gG_0 of the coset $C_g\mathfrak{G}$ having smallest possible norm $N\mathfrak{u} =: u_g$ coprime to the conductor f. Then $\mathfrak{u} \in \mathfrak{A}$ since we may divide out from $\mathfrak{u}$ any rational integer and any ideal dividing (D) and still obtain an ideal in a class $C_g\mathfrak{G}$ with even smaller norm (notice that ideals dividing (D_0) are contained in ambiguous classes). For $n \in \mathbb{N}, G \in \mathfrak{G}$, and $x \in \mathbb{R}$ define

$$\rho_1(n, G) := \#\{\mathfrak{a} \in \mathfrak{X}_G \mid N\mathfrak{a} = n\}, \quad R_1(x, G) := \sum_{\substack{n \le x \\ (n,f)=1}} \rho_1(n, G),$$
$$\rho_2(n, G) := \#\{\mathfrak{a} \in C_gG \cap \mathfrak{A} \mid N\mathfrak{a} = n\}, \quad R_2(x, G) := \sum_{\substack{n \le x \\ (n,f)=1}} \rho_2(n, G).$$

Since the number of pairs of different ideals in C_g having norm n is equal to

$$\sum_{m|n} \sum_{G \in \mathfrak{G}} \rho_1(G, m)\rho_2(G, n/m),$$

the following estimates will become very useful later, when we will essentially multiply them together.

Lemma 4.16. *For all $G \in \mathfrak{G}$, we have*

$$R_1(x,G) \leq \frac{16x}{\sqrt{D}}, \quad R_2(x,G) \begin{cases} \ll xD^{-1/2} + \sqrt{x} & \textit{for all } x, \\ \leq 1 & \textit{for } x < \sqrt{D/4}, \\ = 0 & \textit{for } x < (D/4)^{1/4}, \end{cases}$$

with absolute implied constants.

Proof. First, we note that if the vector $\mathbf{x} \in \mathbb{Z}^2$ corresponds to the ideal $\mathfrak{a}$ as in Theorem 3.33, then for $r \in \mathbb{Z}$ the vector $r\mathbf{x}$ corresponds to $(r)\mathfrak{a}$. By Lemma 3.38, each ambiguous class $G \in \mathfrak{G}$ contains a form of the shape

$$h_G(x_1,x_2) = ax_1^2 + cx_2^2, \quad 4ac = D,$$
$$h_G(x_1,x_2) = ax_1^2 + ax_1x_2 + cx_2^2 = a\left(x_1 + \frac{1}{2}x_2\right)^2 + \left(c - \frac{1}{4}a\right)x_2^2, \quad a(4c-a) = D$$

or

$$h_G(x_1,x_2) = ax_1^2 + bx_1x_2 + ax_2^2 = \left(\frac{a}{2} + \frac{b}{4}\right)(x_1+x_2)^2 + \left(\frac{a}{2} - \frac{b}{4}\right)(x_1-x_2)^2, \qquad 4a^2 - b^2 = D,$$

with positive integers $b < a \leq c$.

In the first case, the vectors $(0,k),(k,0)$ correspond to ideals $(k)\mathfrak{a}$ with $N(\mathfrak{a}) \mid D$, which are either not coprime to the conductor or are equal to their conjugate, by Corollary 3.22; so they are not in $\mathfrak{X}_G$. Therefore,

$$R_1(x,G) \leq \#\left\{(x_1,x_2) \in \mathbb{Z}^2 \mid x_1x_2 \neq 0, |x_1| \leq \sqrt{\frac{x}{a}}, |x_2| \leq \sqrt{\frac{x}{c}}\right\} \leq \frac{8x}{\sqrt{D}}.$$

In the second case, the vectors $(x_1,0)$ and $(x_1,-2x_1)$ correspond to ideals that are equal to their conjugate, using similar arguments as above. Consequently,

$$
\begin{aligned}
R_1(x,G) &\leq \# \left\{ (y_1,y_2) \in \mathbb{Z}^2 \mid y_1y_2 \neq 0, |y_1| \leq \sqrt{\frac{x}{a}}, |y_2| \leq \sqrt{x}\left(c - \frac{1}{4}a\right)^{-\frac{1}{2}} \right\} \\
&\leq \frac{8x}{\sqrt{D}}
\end{aligned}
$$

In the third case, the vectors $(x_1, \pm x_1)$ correspond to ideals that are equal to their conjugate. Thus,

$$
\begin{aligned}
R_1(x,G) &\leq \# \left\{ (y_1,y_2) \in \mathbb{Z}^2 \mid y_1y_2 \neq 0, |y_1| \leq \sqrt{x}\left(\frac{a}{2} + \frac{b}{4}\right)^{-\frac{1}{2}}, |y_2| \leq \sqrt{x}\left(\frac{a}{2} - \frac{b}{4}\right)^{-\frac{1}{2}} \right\} \\
&\leq \frac{16x}{\sqrt{D}}.
\end{aligned}
$$

This proves the first part of the lemma.

For the second assertion, we first have the estimate $R_2(x,G) \ll xD^{-1/2} + \sqrt{x}$ from Lemma 4.11, which considers all representations and, equivalently, all ideals. Next note that, if $\mathfrak{a}$ is a principal ideal with $N(\mathfrak{a}) < D/4$, then $\mathfrak{a}$ is generated by a rational integer, since any element in $\mathcal{O}_D$ is of the form $(x + y\sqrt{D})/2$, so that $N((x + y\sqrt{D})/2) = (x^2 + y^2D)/4$. Therefore an ideal $\mathfrak{v} \in C_f G \cap \mathfrak{A}$ with $N(\mathfrak{v}) < (D/4)^{1/4}$ would produce a principal ideal $\bar{\mathfrak{u}}^2\mathfrak{v}^2$ with norm smaller than $D/4$, since $N(\mathfrak{u}) \leq N(\mathfrak{v})$ by minimality. Thus $\bar{\mathfrak{u}}^2\mathfrak{v}^2$ would be generated by a rational integer, but if $\mathfrak{u} \neq \mathfrak{v}$, then this is impossible by Lemma 4.15. Similarly, two different ideals $\mathfrak{v}_1, \mathfrak{v}_2 \in C_f G \cap \mathfrak{A}$ with $N(\mathfrak{v}_1), N(\mathfrak{v}_2) < (D/4)^{1/2}$ would produce a principal ideal $\overline{\mathfrak{v}_1}\mathfrak{v}_2$ with norm smaller than $D/4$, which is thus generated by a rational integer. This is impossible by the same argument. □

Define $A_1(n) := \#\{\mathfrak{a} \in C_g \mid N(\mathfrak{a}) = n, \mathfrak{a} \notin \mathfrak{u}\mathfrak{X}_{G_0}\}$ and $A_2(n) := \#\{\mathfrak{a} \in C_g \mid N(\mathfrak{a}) = n, \mathfrak{a} \in \mathfrak{u}\mathfrak{X}_{G_0}\}$. By our definition of $\mathfrak{X}_{G_0}$ we see that $A_2 = 0$ or $A_2 \geq 2$. Now define $B := \{\mathfrak{c} \in G_0 \mid \mathfrak{c} = \bar{\mathfrak{c}}, N(\mathfrak{c}) \leq x/u_g, (N(\mathfrak{c}), f) = 1\}$. Using the unique prime ideal factorization, we notice that all ideals in B are of the form $\mathfrak{c} = \xi\mathfrak{p}_1 \ldots \mathfrak{p}_k$, where $k \in \mathbb{N}_0$, $\xi \in \mathbb{Z}$ and $\mathfrak{p}_1, \ldots \mathfrak{p}_k$ are distinct ramified primes. We see that there are at most two primitive ideals in G_0 dividing D_0. Indeed, if $\mathfrak{c}_1, \mathfrak{c}_2$, and $\mathfrak{c}_3$ were three such distinct ideals, then their pairwise products are principal ideals. Since these products are distinct, we can find one, say $\mathfrak{c}_2\mathfrak{c}_3$, of the form $(\xi)\mathfrak{p}_1 \ldots \mathfrak{p}_k$, where $\xi \in \mathbb{Z}$ and $\mathfrak{p}_1, \ldots \mathfrak{p}_k$ are distinct ramified primes with $N(\mathfrak{p}_1 \ldots \mathfrak{p}_k) < D/4$. It follows that $\mathfrak{c}_2\mathfrak{c}_3$ is the principal ideal generated by a rational prime and this implies that $\mathfrak{c}_2$ and $\mathfrak{c}_3$ are equal by the uniqueness of prime ideal factorization.

If $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are two (possibly equal) primitive ideals in G_0 dividing D_0, then the above arguments show that all elements of B are integers times $\mathfrak{c}_1$ or times $\mathfrak{c}_2$. If we assume (without loss of generality) that $N(\mathfrak{c}_1) \leq N(\mathfrak{c}_2)$, then $|B| \ll \sqrt{x/(u_g N(\mathfrak{c}_1))}$. Since $\mathfrak{u}\mathfrak{c}_1 \in C_g$ and $(N(\mathfrak{u})N(\mathfrak{c}_1), f) = 1$, we have $u_g N(\mathfrak{c}_1) \geq a_g$ by minimality of a_g, so that $|B| = O(\sqrt{x/a_g})$.

If we define $r_g^*(n, \beta) = A_1(n) + 2^{\beta-1}A_2(n)$, then by Lemma 4.11 and the correspondence between representations and ideals we have

$$\begin{aligned}\sum_{\substack{n\leq x\\(n,f)=1}} r_g^*(n,\beta) &= \sum_{\substack{n\leq x\\(n,f)=1}} (A_1(n)+A_2(n)) + (2^{\beta-1}-1)\Bigg(|B| + \sum_{\substack{n\leq x\\(n,f)=1}} A_2(n)\Bigg) + O_\beta(|B|)\\
&= |\{\mathfrak{a}\in C_g \mid N\mathfrak{a}\leq x, (N\mathfrak{a}, f)=1\}| +\\
&+ (2^{\beta-1}-1)\cdot|\{\mathfrak{c}\in G_0 \mid N\mathfrak{c}\leq x/u_g, (N\mathfrak{c}, f)=1\}| + O\left(\sqrt{\frac{x}{a_g}}\right)\\
&= \frac{\varphi(f)}{f}\cdot\frac{2}{w_D}\left(1+\frac{2^{\beta-1}-1}{u_g}\right)\frac{\pi x}{\sqrt{D}} + O\left(2^{\omega(f)}\left(1+\sqrt{\frac{x}{a_g}}\right)\right) \qquad (4.3)\end{aligned}$$

Let us now assume $\beta \leq 2$. A short computation using the first derivative yields that $\xi(\beta) := (A_1 + A_2)^\beta - (A_1 + 2^{\beta-1}A_2)$ satisfies $|\xi(\beta)| \leq \xi(2)$ for $0 \leq \beta \leq 2$ and $A_1 \in \mathbb{N}_0, A_2 \in \mathbb{N}_0 \setminus \{1\}$ as above. By Lemma 4.16 and the fact that $A_2(n) \in \mathbb{N}_0 \setminus \{1\}$ for all $n \in \mathbb{N}$, as noted above, we have

$$\begin{aligned}&\sum_{\substack{n\leq x\\(n,f)=1}} |r_g(n)^\beta - r_g^*(n,\beta)| \leq \sum_{\substack{n\leq x\\(n,f)=1}} r_g(n)^2 - r_g^*(n,2)\\
&= \sum_{G\in\mathfrak{G}(D)} \#\{(\mathfrak{b}\mathfrak{c}, \mathfrak{b}\bar{\mathfrak{c}}) \mid \mathfrak{c}\in\mathfrak{X}_G, \mathfrak{b}\in C_gG\cap\mathfrak{A}\setminus\{\mathfrak{u}\}, N\mathfrak{b}\mathfrak{c}\leq x, (N\mathfrak{b}\mathfrak{c}, f)=1)\}\\
&= \sum_{G\in\mathfrak{G}(D)} \sum_{\substack{k\leq x\\(k,f)=1}} \rho_1(k,G) \sum_{\substack{l\leq x/k\\(l,f)=1}} \rho_2(l,G)\\
&\ll \sum_{G\in\mathfrak{G}(D)}\Bigg(\sum_{\substack{k\ll xD^{-1/4}\\(k,f)=1}} \rho_1(k,G) + \sum_{\substack{k\leq xD^{-1/2}\\(k,f)=1}} \rho_1(k,G)\left(\frac{x}{k\sqrt{D}} + \sqrt{\frac{x}{k}}\right)\Bigg)\end{aligned}$$

Using partial summation (for the statement of the formula see for instance [Ten95, Thm. 1, Chap I.0]), we can estimate

$$
\begin{aligned}
&\sum_{\substack{k\le xD^{-1/2}\\(k,f)=1}} \rho_1(k,G)\left(\frac{x}{k\sqrt{D}}+\sqrt{\frac{x}{k}}\right)\\
&= \sum_{\substack{k\le xD^{-1/2}\\(k,f)=1}} \rho_1(k,G)(1+D^{1/4}) + \int_1^{x/\sqrt{D}} R_1(t,G)\left(\frac{x}{t^2\sqrt{D}}+\frac{\sqrt{x}}{2t^{3/2}}\right)dt\\
&\ll \frac{x}{D^{3/4}} + \int_1^{x/\sqrt{D}} \frac{t}{\sqrt{D}}\cdot\frac{1}{t}\left(\frac{x}{t\sqrt{D}}+\frac{\sqrt{x}}{2t^{1/2}}\right)dt\\
&= \frac{x}{D^{3/4}} + \left(\frac{x}{D}\log t+\sqrt{\frac{x}{D}}\sqrt{t}\right)\Bigg|_1^{x/\sqrt{D}} \ll \frac{x}{D^{3/4}}+\frac{x\log x}{D}.
\end{aligned}
$$

Summing over the ambiguous classes and using that $|\mathfrak{G}(D)| \ll 2^{\omega(D)}$ as in Theorem 3.39, we arrive together with (4.3) at the theorem in the case $\beta \le 2$.

If $\beta > 2$, then $0 \le r_g(n)^\beta - r_g^*(n,\beta) \le 3r_g(n)^{\beta-2}(r_g(n)^2 - r_g^*(n,2))$, that is, $(A_1 + A_2)^\beta - (A_1 + 2^{\beta-1}A_2) \le 3(A_1+A_2)^{\beta-2}((A_1+A_2)^2-(A_1+2A_2))$. Indeed if $A_1+A_2 \ge 3$, then $(A_1+A_2)^2 \ge 3(A_1+A_2) \ge (3/2)(A_1+2A_2)$ and the claim follows. If $A_1+A_2 \le 2$ with $A_2 \ne 1$, then both sides of the inequality are zero. Therefore, by Lemma 4.7:

$$
\begin{aligned}
\sum_{\substack{n\le x\\(n,f)=1}} |r_g(n)^\beta - r_g^*(n,\beta)| &\le 3 \sum_{\substack{n\le x\\(n,f)=1}} \tau(n)^{\beta-2}(r_g(n)^2 - r_g^*(n,2))\\
&\le 3 \sum_{G\in\mathfrak{G}(D)} \sum_{\substack{k\le x\\(k,f)=1}} \tau(k)^{\beta-2}\rho_1(k,G) \sum_{\substack{l\le x/k\\(l,f)=1}} \tau(l)^{\beta-2}\rho_2(l,G).
\end{aligned}
$$

For $q > 1$ choose $p > 1$ such that $1/q + 1/p = 1$. By Hölder's inequality and Lemma 4.8 we get

$$
\begin{aligned}
\sum_{\substack{k\le x\\(k,f)=1}} \tau(k)^{\beta-2}\rho_1(k,G) &\le \Bigg(\sum_{\substack{k\le x\\(k,f)=1}} \rho_1(k,G)\Bigg)^{1/p}\Bigg(\sum_{k\le x}\tau(k)^{((p-1)/p+\beta-2)q}\Bigg)^{1/q}\\
&\ll \frac{x^{1/p}}{D^{1/2p}}\cdot x^{1/q}\cdot\left((\log x)^{2^{((p-1)/p+\beta-2)q}+1}\right)^{1/q}\\
&\ll \frac{x}{D^{(1/2)(1-1/q)}}(\log x)^{(1/q)(2^{(\beta-2)q+1}-1)},
\end{aligned}
$$

and similarly,

$$\sum_{l\leq x}\tau(l)^{\beta-2}\rho_2(l,G) \ll \frac{x}{D^{(1/4)(1-1/q)}}(\log x)^{(1/q)(2^{(\beta-2)q+1}-1)},$$

where we use the bound $\sum_{n\leq x}\rho_2(n,G) \ll xD^{-1/4}$ given by Lemma 4.16. Collecting these estimates, we find by partial summation that

$$\sum_{\substack{n\leq x\\(n,f)=1}} r_g(n)^\beta - r_g^*(n,\beta) \ll 2^{\omega(D)}\frac{x(\log x)^{(2/q)(2^{(\beta-2)q+1}-1)}}{D^{(3/4)(1-1/q)}}$$

for any $q > 1$. □

Remark 4.17. In the special case when $f = 1$, we obtain Theorem 2 in [BG06] by using the improved error term mentioned in Remark 4.12.

Theorem 4.18 (Blomer-Granville). *For a given binary quadratic form g with fundamental discriminant $-D = D_0 < 0$, let a_g be the smallest positive integer that is represented by g, and let u_g be the smallest positive integer that can be represented by some form in the coset $g\mathfrak{G}(D)$. For any $\beta \geq 0$ we have:*

$$\sum_{n\leq x} r_g(n)^\beta = \left(1+\frac{2^{\beta-1}-1}{u_g}\right)\frac{\pi x}{\sqrt{D}} + E_\beta(x,D),$$

where

$$E_\beta(x,D) \ll \begin{cases} \sqrt{\dfrac{x}{a_g}} + 2^{\omega(D)}\left(\dfrac{x\log x}{D}+\dfrac{x}{D^{3/4}}\right), & 0\leq\beta\leq 2,\\ \sqrt{\dfrac{x}{a_g}} + 2^{\omega(D)}\dfrac{x(\log x)^{(2/q)(2^{(\beta-2)q+1}-1)+1}}{D^{(3/4)(1-1/q)}}, & \beta > 2,\end{cases}$$

for any real $q > 1$. The implied constants depend at most on β and q.

Notice also that Blomer and Granville tacitely assume the discriminant to be $-D < -4$, so that $w_D = 2$, since for very small D the main term in the theorem is smaller than the error term.

Remark 4.19 (Pedantic remark). It is useful for the next proof to notice that the $\log x$ terms in the error do not appear if $x < 1$. This is trivial in itself, since the left hand side in the theorem is zero automatically if $x < 1$, but in the following we shall apply Theorem

4.13 to quotients of x by divisors of the conductor, which may make the logarithm negative and we wish to avoid such formal errors and confusions they may bring about.

Proof of the Main Theorem 4.1. We reduce the case of arbitrary numbers that may not be coprime to the conductor by the Reduction Theorem 4.5 and then we apply Theorem 4.13. We have

$$\begin{aligned}\sum_{n\leq x} r_g(n) &= \sum_{n\leq x}\frac{1}{w_D}R(g,n) = \sum_{d|f}\sum_{\substack{n\leq x\\(n,f^2)=d^2}}\frac{1}{w_D}R(g,n)\\ &= \sum_{d|f}\frac{w_{D/d^2}}{w_D}\sum_{\substack{k\leq x/d^2\\(k,(f/d)^2)=1}} r_{\theta_d(g)}(k).\end{aligned}$$

Applying Theorem 4.13, we first compute the main term:

$$\begin{aligned}&\sum_{d|f}\frac{w_{D/d^2}}{w_D}\cdot\frac{\varphi(f_{D/d^2})}{f_{D/d^2}}\cdot\frac{2}{w_{D/d^2}}\left(1+\frac{2^{\beta-1}-1}{u_{\theta_d(g)}}\right)\frac{\pi x/d^2}{\sqrt{D/d^2}}\\ &=\frac{2}{w_D}\cdot\frac{\pi x}{\sqrt{D}}\sum_{d|f}\frac{\varphi(f/d)}{f}\left(1+\frac{2^{\beta-1}-1}{u_{\theta_d(g)}}\right)\\ &=\frac{2}{w_D}\cdot\frac{\pi x}{\sqrt{D}}\left(1+(2^{\beta-1}-1)\sum_{d|f}\frac{\varphi(f/d)}{fu_{\theta_d(g)}}\right).\end{aligned}$$

Now for the error term, we first recall the convolution identities of Lemma 4.9 and obtain

$$\sum_{d|f}2^{\omega(f/d)}\left(1+\sqrt{\frac{x}{d^2a_{\theta_d(g)}}}\right)=\tau(f^2)+\sum_{d|f}\frac{2^{\omega(f/d)}}{d}\sqrt{\frac{x}{a_{\theta_d(g)}}}.$$

Next we show that $\sum_{d|f}2^{\omega(D/d^2)}\leq\tau(D)$. For simplicity we assume that 2 does not divide the fundamental discriminant D_0. Then the discriminant has the following prime factor decomposition:

$$D=f^2D_0=\prod_{p\in P}p^{2\alpha_p}\prod_{q\in Q}q^{2\alpha_q+1}\prod_{r\in R}r,$$

where P,Q and R are disjoint sets of primes and

$$f=\prod_{p\in P}p^{\alpha_p}\prod_{q\in Q}q^{\alpha_q}\quad\text{and}\quad D_0=\prod_{q\in Q}q\prod_{r\in R}r.$$

It follows, again by Lemma 4.9, that

$$\begin{aligned}
\sum_{d|f} 2^{\omega(D/d^2)} &= \sum_{d|\prod p^{\alpha_p}} \sum_{e|\prod q^{\alpha_q}} 2^{\omega\left(\frac{\prod p^{2\alpha_p}}{d^2}\cdot\frac{\prod q^{2\alpha_q+1}}{e^2}\prod r\right)} \\
&= \sum_{e|\prod q^{\alpha_q}} \sum_{d|\prod p^{\alpha_p}} 2^{\omega\left(\frac{\prod p^{2\alpha_p}}{d^2}\right)} \cdot 2^{|Q|+|R|} \\
&= \tau\left(\prod q^{\alpha_q}\right) \tau\left(\prod p^{2\alpha_p}\right) \cdot 2^{|Q|+|R|} \\
&= \prod_q 2(\alpha_q+1) \cdot \prod_p (2\alpha_p+1) \cdot 2^{|R|} = \tau(D).
\end{aligned}$$

If $2 \mid D_0$, then one can easily adapt the computations above to check that

$$\sum_{d|f} 2^{\omega(D/d^2)} \leq \tau(D).$$

Finally we bound $\log(x/d^2)$ by $\log x$ and recall Remark 4.19. Moreover, we have

$$\frac{x/d^2}{(D/d^2)^\alpha} \leq \frac{x}{D}$$

for any $\alpha \leq 1$. Therefore,

$$\sum_{d|f} 2^{\omega(D/d^2)} \left(\frac{(x/d^2)\log(x/d^2)}{D/d^2} + \frac{x/d^2}{(D/d^2)^{3/4}} \right) \ll \tau(D) \left(\frac{x \log x}{D} + \frac{x}{D^{3/4}} \right),$$

which gives the error term in the case $0 \leq \beta \leq 2$. The estimate for $\beta > 2$ is analogous. □

5 Considerations, variations and applications

5.1 A brief critical analysis of the result

We have seen that the main term in Theorem 4.1 includes a sum over divisors of the conductor. We would like to simplify this sum, since computing all of the numbers $u_{\theta_d(g)}$ is far from trivial and can be indeed very expensive. We are able to do so satisfyingly if the form is in an ambiguous class, but in the general case it seems that the value of the sum is not very easy to describe.

If g is in an ambiguous class, then $u_g = 1$, since the coset $g\mathfrak{G}(D)$ includes the class of the principal form, i.e. the neutral element in the group, which obviously represents 1 (recall Remark 3.31). Analogously, for all divisors d of the conductor we have $u_{\theta_d(g)} = 1$, since the images of g under the maps θ_d are ambiguous classes. Indeed, Theorem 2.1 in [SW06] states that θ_d is in fact a surjective homomorphism[4], so that images of elements with order 1 or 2, i.e. the ambiguous classes, have order 1 or 2 as well. Therefore, the sum simplifies to

$$\sum_{d|f} \frac{\varphi(f/d)}{f u_{\theta_d(g)}} = \sum_{d|f} \frac{\varphi(f/d)}{f} = 1,$$

by the convolution identity of φ from Lemma 4.9. Since all $u_{\theta_d(g)}$ are at least 1, this is the maximal value of the sum. Now taking $\beta = 0$ in the main theorem, so that $2^{\beta-1} - 1 = -1/2 < 0$, we note that the constant in front of the main term is smallest for ambiguous forms. Consequently, ambiguous forms represent fewer small integers than non-ambiguous forms. On the other hand, taking for instance $\beta = 1$ shows that ambiguous forms represent small integers with higher multiplicity.

In contrast, for general classes of forms we cannot expect the sum to be 1. Indeed, the values of $u_{\theta_d(g)}$ usually vary with d and the sum does not necessarily equal a fraction of the form $1/u$, as one would naively try to generalize Theorem 4.18. Table 1, computed using the computer algebra system *Magma*, shows a few examples of non-ambiguous forms of discriminant $-3600 = -4 \cdot (2 \cdot 3 \cdot 5)^2$. Here, the number $\tilde{u}_g$ denotes the smallest integer represented by the coset $g\mathfrak{G}(D)$ without the condition of coprimality to the conductor. For the straight-forward (and non-optimized) algorithm used for these computations see the Appendix (section 6).

We observe from the table that, unlike in the case of ambiguous forms, the smallest

[4] See alternatively Theorem 6.4.14 of [HK13]. It is shown there that the map θ_d corresponds through the isomorphism between form classes $\mathfrak{F}_D$ and ideal classes $\mathfrak{C}(D)$ to the homomorphism induced by the map $\mathfrak{C}(D) \longrightarrow \mathfrak{C}(D/d^2), \mathfrak{a} \mapsto \mathfrak{a}\mathcal{O}_{D/d^2}$, which can be seen to be well-defined (see [HK13, Theorem 5.9.7]). Since the composition of homomorphisms is again a homomorphism, θ_D preserves the group structure.

g	$\tilde{u}_g$	u_g	$u_{\theta_2(g)}$	$u_{\theta_3(g)}$	$u_{\theta_5(g)}$	$u_{\theta_6(g)}$	$u_{\theta_{10}(g)}$	$u_{\theta_{15}(g)}$	$u_{\theta_{30}(g)}$	$\sum \frac{\varphi(f/d)}{f u_{\theta_d(g)}}$
$[8,4,113]$	8	17	1	13	5	1	1	1	1	$\frac{9491}{16575}$
$[9,6,101]$	9	29	13	1	5	1	1	1	1	$\frac{12527}{28275}$
$[13,12,72]$	13	13	13	13	1	1	1	1	1	$\frac{5}{13}$

Table 1: Computed examples for forms of discriminant -3600

number represented by the coset need not be coprime to the conductor, so that $u_{\theta_d(g)}$ must be bigger. It is also not obvious how to simplify the sum. One possible idea is to use the Reduction Theorem 4.5 and try to relate the numbers u_g and $u_{\theta_d(g)}$. This strategy would at least require the homomorphisms θ_d to map ambiguous classes to ambiguous classes surjectively, so that the cosets $g\mathfrak{G}(D)$ and $\theta_d(g)\mathfrak{G}(D/d^2)$ are correlated. Unfortunately, this is not necessarily true for even discriminants. For instance, using the formula from Theorem 3.39, we can compute that there are only two ambiguous classes of discriminant $-256 = -4 \cdot 8^2$, namely $[4,4,17]$ and $[1,0,64]$ (recall Lemma 3.38). Note that, by applying the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, we have $[4,4,17] = [17,-4,4]$. Now

$$\theta_2([1,0,64]) = [1,0,16] \quad \text{and} \quad \theta_2([17,-4,4]) = [17,-2,1] = [1,0,16],$$

where the last equality can be seen by applying the matrix $\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. On the other hand, there are two ambiguous classes of discriminant -64, so that the map θ_2 restricted to ambiguous classes is not surjective in this case.

To conclude, further study is needed to understand the constant in front of the main term in our main theorem. Nevertheless, by estimating this constant, the result may be good enough for some applications, as we shall see in section 5.3.

5.2 Proper representations

There are counting problems involving binary quadratic forms where it is only the primitive representations that need to be considered. I have been able to obtain partial results in this setting, but failed to prove a general asymptotic as in the main theorem of this thesis. At the end of this section, the main difficulty and a possible strategy will presented.

Lemma 5.1. *For $n \in \mathbb{N}$ and a binary quadratic form g, we have*

$$R^*(g,n) = \sum_{d^2|n} \mu(d) R(g, n/d^2).$$

Proof. We first note that

$$R(g,n) = \sum_{d^2} R^*(g, n/d^2), \tag{5.1}$$

since if $g(x,y) = n$ and $\gcd(x,y) = d$, then $g(x/d, y/d) = n/d^2 \in \mathbb{Z}$ and $\gcd(x/d, y/d) = 1$; conversely, $g(x,y) = n/d^2$ implies that $g(dx, dy) = n$. Equation (5.1) is a Dirichlet convolution identity of the form $R = \mathbf{1}_\square * R^*$, where $\mathbf{1}_\square$ is the indicator function of square integers. Since the Dirichlet series associated to $\mathbf{1}_\square$ is $\zeta(2s)$, we find that the inverse of $\mathbf{1}_\square$ in the ring of arithmetic functions is the generating function of $\zeta^{-1}(2s)$, that is

$$\nu(n) = \begin{cases} \mu(m), & n = m^2, m \in \mathbb{Z}, \\ 0, & \text{if } n \text{ is not a square.} \end{cases}$$

Using this variant of Möbius inversion, we find that $R^* = \nu * R$, which is a restatement of the claim. □

Lemma 5.2. *For a binary quadratic form g with discriminant $-D < 0$ and conductor f, let a be the smallest integer represented by g. Then we have the asymptotics*

$$\sum_{n \le x} R^*(g,n) = \frac{1}{\zeta(2)} \cdot \frac{2\pi x}{\sqrt{D}} + O\left(\sqrt{\frac{x}{a}} \log x\right)$$

and

$$\sum_{\substack{1 \le n \le x \\ (n,f)=1}} R^*(g,n) = \frac{1}{\zeta(2)} \cdot \frac{\phi(f)}{f} \cdot \frac{2\pi x}{\sqrt{D}} + O\left(2^{\omega(f)}\left(\sqrt{x} + \sqrt{\frac{x}{a}} \log x\right)\right).$$

Proof. By the convolution relation between R and R^*, i.e. Lemma 5.1, and the asymptotics for R, i.e. Lemma 4.11 together with Remark 4.12, we compute

$$\begin{aligned} \sum_{n \le x} R^*(g,n) &= \sum_{n \le x} \sum_{d^2 | n} \mu(d) R(g, n/d^2) \\ &= \sum_{d \le \sqrt{x}} \mu(d) \sum_{k \le x/d^2} R(g,k) \\ &= \frac{2\pi x}{\sqrt{D}} \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} + O\left(\sum_{d \le \sqrt{x}} \frac{1}{d} \cdot \sqrt{\frac{x}{a}}\right) \\ &= \frac{1}{\zeta(2)} \cdot \frac{2\pi x}{\sqrt{D}} + O\left(\sqrt{\frac{x}{D}} + \sqrt{\frac{x}{a}} \cdot \log x\right). \end{aligned}$$

Here we used the identity $\sum \mu(n)/n^2 = 1/\zeta(2)$ and the simple estimate $\sum_{n \le x} \mu(n)/n^2 = 1/\zeta(2) + O(1/x)$. Since $a \le D$ (recall (2.3)), the error term condenses to $O(\sqrt{x/a} \log x)$.

The proof for the sum over numbers coprime to the conductor is completely analogous. □

The previous lemma gives us one of the fundamental ingredients for the proof of the main theorem in the case of primitive representations. We may use almost the same proof to obtain the special result for the proper representation of numbers coprime to the conductor.

Corollary 5.3. *For a given binary quadratic form g with discriminant $-D = D_0 f_D^2 < 0$ and conductor f_D, let a_g be the smallest positive integer that is represented by g, and let u_g be the smallest positive integer coprime to f_D that can be represented by some form in the coset $g\mathfrak{G}(D)$. For any $\beta \geq 0$ we have:*

$$\sum_{\substack{n \leq x \\ (n, f_D)=1}} r_g^*(n)^\beta = \frac{1}{\zeta(2)} \cdot \frac{\varphi(f_D)}{f_D} \cdot \frac{2}{w_D} \left(1 + \frac{2^{\beta-1}-1}{u_g}\right) \frac{\pi x}{\sqrt{D}} + E_\beta(x, D),$$

where

$$E_\beta(x, D) \ll \begin{cases} 2^{\omega(f)} \left(\sqrt{x} + \sqrt{\dfrac{x}{a_g}} \log x\right) + 2^{\omega(D)} \left(\dfrac{x \log x}{D} + \dfrac{x}{D^{3/4}}\right), & 0 \leq \beta \leq 2, \\ 2^{\omega(f)} \left(\sqrt{x} + \sqrt{\dfrac{x}{a_g}} \log x\right) + 2^{\omega(D)} \dfrac{x (\log x)^{(2/q)(2^{(\beta-2)q+1}-1)+1}}{D^{(3/4)(1-1/q)}}, & \beta > 2, \end{cases}$$

for any real $q > 1$. The implied constants depend at most on β and q.

Proof. Note first that u_g and a_g are properly represented in virtue of their minimality. Moreover, proper representations correspond by Theorem 3.33 to primitive ideals. Indeed, if $\gcd(x, y) > 1$, then $(xa + ya\tau)\mathfrak{a}^{-1} = \gcd(x, y)(\frac{x}{\gcd(x,y)} a + \frac{y}{\gcd(x,y)} a\tau)\mathfrak{a}^{-1}$ is not primitive. Conversely, if $kI = (xa + ya\tau)\mathfrak{a}^{-1}$ for some $\mathcal{O}_D$-ideal I and $k \in \mathbb{Z}$, then $kI\mathfrak{a} = (xa + ya\tau)\mathcal{O}_D$. Thus $(xa + ya\tau) = k(x'a + y'a\tau)$, which is easily seen to imply that $k \mid x$ and $k \mid y$.

We can now adapt the proof of Theorem 4.13 simply by requiring all ideals to be primitive. We can still factorize each pair of distinct primitive ideals into products of the form $\mathfrak{b}\mathfrak{c}$ and $\mathfrak{b}\bar{\mathfrak{c}}$, where now both $\mathfrak{b}$ and $\mathfrak{c}$ must be primitive. Indeed, a primitive ideal can only be written as a product of primitive ideals. For the corresponding sums considering only primitive ideals, $R_1^*(x, G) =$ and $R_2^*(x, G)$, we may use the same bounds from Lemma 4.16, since $R_{1,2}^*(x, G) \leq R_{1,2}(x, G)$. These bounds are used in estimating the error terms and are essentially multiplied together to produce all pairs of the form

$(\mathfrak{bc}, \mathfrak{b}\bar{\mathfrak{c}})$. Although multiplying two primitive ideals does not necessarily yield a primitive one, we are only interested in upper bounds, so that for $G \in \mathfrak{G}(D)$ we may estimate

$$\begin{aligned}
&\#\{(\mathfrak{bc}, \mathfrak{b}\bar{\mathfrak{c}}) \mid \mathfrak{bc} \text{ and } \mathfrak{b}\bar{\mathfrak{c}} \text{ primitive}, \mathfrak{c} \in \mathfrak{X}_G, \mathfrak{b} \in C_g G \cap \mathfrak{A} \setminus \{\mathfrak{u}\}, N\mathfrak{bc} \leq x, (N\mathfrak{bc}, f) = 1)\} \\
&\leq \sum_{\substack{k \leq x \\ (k,f)=1}} \rho_1^*(k, G) \sum_{\substack{l \leq x/k \\ (l,f)=1}} \rho_2^*(l, G)
\end{aligned}$$

Here, ρ_1^* and ρ_2^* only count primitive ideals. The rest of the computations for the error term remain the same.

For calculating the main term we use the asymptotics for $R^*(g, n)$ from Lemma 5.2 and the observations above. The rest is almost identical. □

The other crucial ingredient which allowed us to generalize Theorem 4.18 to non-fundamental discriminants, i.e. the reduction of representation numbers from Theorem 4.5, is not available any more in this case. We may easily find a counter-example.

Remark 5.4. The form $g(x, y) = x^2 + 36y^2$ has discriminant $-4 \cdot (2 \cdot 3)^2 = -144$. Its image under θ_3 is the form $\tilde{g}(x, y) = x^2 + 4y^2$ of discriminant $-4 \cdot 2^2$. The number $n = 37$ has, up to equivalence, a single representation, namely $\tilde{g}(1, 3) = \tilde{g}(-1, -3) = 37$, which is proper. On the other hand, the number $m = n \cdot 3^2 = 333$ has only one equivalence class of representations, namely $g(3, 3) = g(-3, -3) = 333$, which is not proper.

The asymptotic for proper representations of forms with non-fundamental discriminant remains a topic for further study. The solution seems to require a better understanding of the reduction of representation numbers for proper representations or a new technique for handling ideals not coprime to the conductor. A possible idea for the latter would be to use factorization into primary ideals. The recent preprint [BGR19] gives in Theorem 3.6 a precise description and count of the proper primitive primary ideals with radical dividing the conductor.

5.3 Apollonian circle packings

Following Elena Fuchs' survey [Fuc13], this last section presents one of the many and newer applications of the theory of binary quadratic forms. The topic is a very old one: it stems from a theorem proved by the ancient Greek geometer Apollonius of Perga (c. 262 BC – c. 190 BC), who was studying straight edge and compass constructions of mutually tangent circles and lines.

Theorem 5.5. *To any three mutually tangent circles or lines there are precisely two other circles or lines which are tangent to all three.*

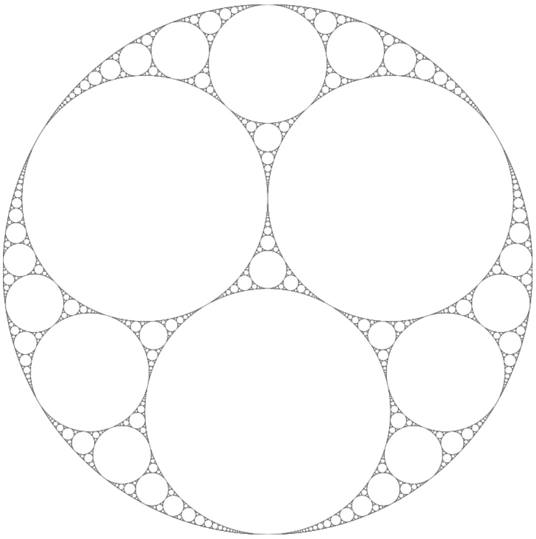


Figure 1: An Apollonian circle packing

Let us start with four mutually tangent circles (one of the circles contains the other three). In each interstice between them we can inscribe by Theorem 5.5 a unique circle, resulting in a configuration of eight mutually tangent circles. We can continue this process for the newly formed interstices and so on ad infinitum, obtaining a so-called *Apollonian circle packing* (ACP) as in Figure 1.[5] In 1643 Descartes found a key property of the curvatures, i.e. inverses of the radii, of circles in such configurations. They satisfy the following relation.

Theorem 5.6. *Let a, b, c and d denote the curvatures of four mutually tangent circles, where a circle is taken to have negative curvature if it is internally tangent to the other three. Then*

$$Q(a, b, c, d) := 2(a^2 + b^2 + c^2 + d^2) - (a + b + c + d)^2 = 0. \tag{5.2}$$

We refer to the quaternary quadratic form Q in (5.2) as the *Descartes quadratic form* and to the curvatures (a, b, c, d) of any four mutually tangent circles as a *Descartes quadruple.* From Theorem 5.6 we easily obtain a remarkable corollary: If any of the Descartes quadruples (a, b, c, d) in an Apollonian circle packing is integral, i.e. $a, b, c, d \in \mathbb{Z}$, then all circles in the packing must have integer curvature. This follows by noting

[5] All figures in this section are reprinted under the Creative Commons licence from the website https://en.wikipedia.org/wiki/Apollonian_gasket.

that, if we view (5.2) as an equation in d, then we have the two solutions:

$$d, d' = a + b + c \pm 2\sqrt{ab + bc + ac}.$$

Therefore, given four mutually tangent circles with integer curvatures a, b, c, d, the only other circle tangent to all three circles with curvatures a, b and c must have curvature

$$d' = 2a + 2b + 2c - d \in \mathbb{Z}. \tag{5.3}$$

The argument is similar for the rest of the circles in an ACP containing the quadruple (a, b, c, d), using the reasoning above recursively and the fact that Q is symmetric in all four variables.

ACP's in which all circles have integer curvature are called *integer* ACP's. A few examples are illustrated in Figure 2 (recall the negative curvature convention from Theorem 5.6). As noted above, we obtain integer ACP's by starting with a generating integer quadruple. Though there are many quadruples generating the same ACP, it is useful to pick out the one with the smallest components. Equivalently, we would like to extract the four largest generating circles in an ACP. This has been done by Graham, Lagarias, Mallows, Wilks and Yan in [Gra+03].

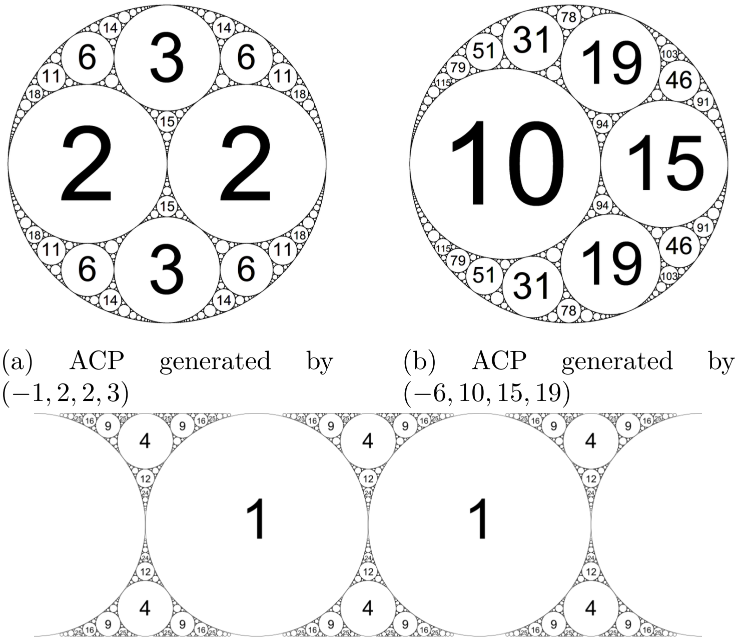
Theorem 5.7. *Define a Descartes quadruple (a, b, c, d) with $a + b + c + d > 0$ to be a root quadruple if $a \leq 0 \leq b \leq c \leq d$ and $a + b + c \geq d$. Then every integer ACP has a unique root quadruple. However, the packing may contain more than one quadruple of mutually tangent circles which yields the root quadruple.*

Now that we have integer ACP's, it is very natural to start asking number theoretic questions. In this thesis we will only touch upon Question 3 in [Fuc13]:

Question. *Do the integers which come up as curvatures in a given ACP make up a positive fraction of $\mathbb{N}$?*

The answer was suspected to be affirmative in [Gra+03] and the conjecture was first proved in [BF11] by Jean Bourgain and Elena Fuchs. To understand the proof, we first need to introduce the Apollonian group, which produces all Descartes quadruples in an integer ACP by acting on the root quadruple.

Starting with four mutually tangent circles with curvatures a, b, c and d, we may produce new circles in the packing by fixing three of the given ones and using the formula (5.3) (and its analogues obtained by switching variables, since the Descartes quadratic form Q is symmetric) to get the second solution for the fourth circle. Extending this



(a) ACP generated by $(-1, 2, 2, 3)$

(b) ACP generated by $(-6, 10, 15, 19)$

(c) Unbounded ACP generated by $(0, 0, 1, 1)$

Figure 2: ACP's with integer curvatures inscribed in the corresponding circles.

process we see that, given a Descartes quadruple $\mathbf{v}_P = (a, b, c, d)^t$ in a packing P, the collection of Descartes quadruples in P is precisely the orbit $A\mathbf{v}_P$, where A is the group generated by the four matrices

$$S_1 = \begin{pmatrix} -1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$S_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 2 & -1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & -1 \end{pmatrix}.$$

The group A is called the *Apollonian group*.

We now state the main theorem in [BF11], which gives a positive answer to our question.

Theorem 5.8. *For an integer Apollonian circle packing P, let $\kappa(P, X)$ denote the number of distinct integers up to X occuring as curvatures in the packing. Then for X large* *we have*

$$\kappa(P, X) \gg X,$$

where the implied constant depends on the packing P.

Since it turns out that working with the full Apollonian group is too difficult, this theorem is proven by considering certain subgroups of it, thus counting curvatures in different ”subpackings” of the given ACP. We define A_i to be the group generated by all but the i-th generator of A:

$$A_i := \langle \{S_1, S_2, S_3, S_4\} \setminus \{S_i\} \rangle.$$

Since S_i is the only generator acting on the i-th component of a vector in $\mathbb{R}^4$, we see that A_i fixes the i-th circle in the root quadruple and produces circles which are tangent to the fixed one.

In [Sar07], Peter Sarnak showed that the set of integers occurring in the orbits of such a subgroup contains the set of integers represented by a certain binary quadratic form with coefficients expressed in terms of the root quadruple of the packing. More precisely, let $\mathbf{v}_P = (a_0, b, c, d)^t$ be the root quadruple of a bounded packing P and C_{a_0} a circle of curvature a_0. For $X \in \mathbb{N}$ let

$$\mathcal{P}_1 := \{n \in \mathbb{N} \mid n \leq X, n = |x_j| \text{ for some } 1 \leq j \leq 4, \text{ for some } (x_1, x_2, x_3, x_4)^t \in A_1 \mathbf{v}_P\}$$

and let

$$f_{a_0}(x, y) = Ax^2 + 2Bxy + Cy^2,$$

where

$$A = b + a_0, \quad B = \frac{a_0 + b + d - c}{2}, \quad C = d + a_0.$$

The discriminant of f_{a_0} can be seen to be $-4a_0^2$. Sarnak showed that $\mathcal{P}_1$ contains the set

$$\mathcal{A}(a_0) = \{a \in \mathbb{N} \mid a \leq X, a = f_{a_0}(x, y) - a_0 \text{ for some } x, y \in \mathbb{Z}, \gcd(x, y) = 1\}.$$

Therefore, since the orbit of A_1 is contained in the orbit of the full group A, a lower bound on the integers less than X represented by the shifted quadratic form $f_{a_0} - a_0$ will serve as a lower bound for $\kappa(X, P)$ as well. We may now finally apply our theory of binary quadratic forms.

Unfortunately, the direct approach does not suffice: it can only show that $\kappa(P, X) \gg X/\sqrt{\log X}$ (see [Sar07]). This is essentially because the orbit of a particular group A_i

gives us only the curvatures of the circles tangent to a fixed circle in P. To obtain some of the missing curvatures, we can associate to each integer $a \in \mathcal{A}(a_0)$ a circle C_a of curvature a tangent to C_{a_0} in the packing P. Applying the same method as above, we find a binary quadratic form f_a such that the set

$$\{\alpha \in \mathbb{N} \mid \alpha \leq X, \alpha \text{ is the curvature of a circle tangent to } C_a \text{ in } P\}$$

contains the set

$$S_a := \{\alpha \in \mathbb{N} \mid \alpha \leq X, \alpha = f_a(x,y) - a \text{ for some } x, y \in \mathbb{Z}, \gcd(x,y) = 1\}.$$

Thus, if we take these sets for all $a \in \mathcal{A}(a_0)$ into consideration, the new count will again reflect circles tangent to a fixed circle in P, but also the circles tangent to all of the already examined ones. Taking care that we do not count curvatures twice, this is enough to show Theorem 5.8.

Indeed, as explained above, since

$$\kappa(P,X) \gg \left| \bigcup_{a \in \mathcal{A}(a_0)} S_a \right|,$$

the theorem is proved if we are able to show that

$$\left| \bigcup_{a \in \mathcal{A}(a_0)} S_a \right| \gg X.$$

This lower bound is achieved using the first step of inclusion-exclusion:

$$\left| \bigcup_{a \in \mathcal{A}(a_0)} S_a \right| \geq \sum_a |S_a| - \sum_{a \neq a'} |S_a \cap S_{a'}|.$$

There are many technical obstacles in the proof, which we will ignore in this thesis. Essentially, we can only achieve good bounds on $|S_a|$ when a is neither too small, nor too big in relation to X. Moreover, one must also obtain a balance between bounds on $\sum |S_a|$ and on $\sum |S_a \cap S_{a'}|$. To achieve this, Bourgain and Fuchs restrict the sum over $\mathcal{A}(a_0)$ to the sum over a carefully chosen smaller set, which we denote by $\mathbf{S}$. More precisely, for a parameter $0 < \eta < 1$ we define

$$\mathbf{S} := \bigcup_k (\mathcal{A}(a_0) \cap [2^k, 2^k + \eta 2^k/\sqrt{k}]),$$

where k ranges over all positive integers satisfying $(\log X)^2 < 2^k < (\log X)^3/2$. In particular $\mathbf{S} \subset \mathcal{A}(a_0) \cap [(\log X)^2, (\log X)^3]$. Bourgain and Fuchs show that

$$|\mathcal{A}(a_0) \cap [2^k, 2^k + \eta 2^k/\sqrt{k}]| \gg \eta \frac{2^k}{k}. \tag{5.4}$$

We may now state and prove Lemma 3.2 from [BF11].

Lemma 5.9. *Let $0 < \eta < 1$. With the notation above we have*

$$\Big| \bigcup_{a \in \mathbf{S}} S_a \Big| \gg \eta X.$$

Proof. The proof given here is an application of Theorem 4.1 and our considerations on proper representations. The article [BF11] applies Blomer and Granville's Theorem 4.18 on the binary quadratic form f_a, which has discriminant $-D = -4a^2$. Since we need to count only proper representations and because $-D$ is fundamental if and only if $a = 1$, Theorem 4.18 is generally not applicable. We may easily mend this by employing essentially the same arguments given by Bourgain and Fuchs, but using the more general main theorem of this thesis.

Although the analogue of Theorem 4.1 for proper representations is not yet available, we can use the Cauchy-Schwarz inequality to reduce the problem to known cases:

$$\sum_{n \le x} R^*(f_a, n)^0 \ge \frac{\left(\sum_{n \le x} R^*(f_a, n)\right)^2}{\sum_{n \le x} R^*(f_a, n)^2} \ge \frac{\left(\sum_{n \le x} R^*(f_a, n)\right)^2}{\sum_{n \le x} R(f_a, n)^2}, \tag{5.5}$$

since $R^*(f_a, n) \le R(f_a, n)$.

First we bound the denominator. Since $(\log X)^2 \le a \le (\log X)^3$ by our definition of $\mathbf{S}$, we find that $(\log X)^4 \ll D \ll (\log X)^6$. We now apply Theorem 4.1 for the form f_a with $\beta = 2$. Since we can crudely bound $u_{\theta_d(f_a)} \ge 1$ for all $d \mid a$ (note that the conductor of f_a is a), we find that

$$1 + (2^{2-1} - 1) \sum_{d \mid a} \frac{\varphi(a/d)}{a u_{\theta_d(f_a)}} \le 1 + \sum_{d \mid a} \frac{\varphi(a/d)}{a} = 2,$$

by Lemma 4.9. Consequently, the main term is bounded by an absolute constant times $X/\sqrt{D}$.

Next, we can also estimate the error term as follows. The first summand in $E_2(X, D)$

is

$$\sum_{d|a} \frac{2^{\omega(a/d)}}{d} \sqrt{\frac{X}{a_{\theta_d(g)}}} \leq \sqrt{X} \sum_{d|a} 2^{\omega(a/d)} = \tau(D^2)\sqrt{X}.$$

Notice that for any $\varepsilon > 0$, we can bound $\tau(D^2) \ll_\varepsilon D^{\varepsilon/6} \ll (\log X)^\varepsilon$ (see [Ten95, Cor. 1.1, Chap. I.5] for the first bound). The last summand can be bounded by

$$D^\varepsilon \left(\frac{X \log X}{D} + \frac{X}{D^{3/4}} \right) \ll \frac{X}{D^{3/4-\varepsilon}},$$

where we used that $\log X \ll D^{1/4}$. Since $D^\varepsilon \sqrt{X} \ll X^{1/2+\varepsilon} \ll X/(\log X)^3 \ll X/D^{3/4}$ we obtain

$$E_2(X, D) \ll_\varepsilon \frac{X}{D^{3/4-\varepsilon}},$$

for any $\varepsilon > 0$. Fixing $\varepsilon \leq 1/4$, it follows that

$$\sum_{n \leq X} R(f_a, n)^2 \ll \frac{X}{\sqrt{D}}.$$

To bound the numerator we use Lemma 5.2. As above, $\sqrt{X} \log X \ll X/(\log X)^4 \ll X/D$, so that the error term is smaller than the main term. Consequently, we have $\sum_{n \leq X} R^*(f_a, n) \gg X/\sqrt{D}$. Applying the bounds obtained above to (5.5), we find that

$$\sum_{n \leq x} R^*(f_a, n)^0 \gg \frac{X}{\sqrt{D}} \gg \frac{X}{a},$$

where the implied constant does not depend on $D = 4a^2$. Finally, this implies that

$$\sum_{a \in \mathcal{S}} S_a \gg \sum_{a \in \mathcal{S}} \frac{X}{a} \gg \eta X \sum_{\substack{2^{k+1} \leq (\log X)^3 \\ 2^k > (\log X)^2}} \frac{1}{k} \gg \eta X,$$

by (5.4) and comparing the sum to the integral. □

The next step in proving the conjecture is to bound the sum over the intersections. This is done in Proposition 3.4 of [BF11] using different techniques, which will not be discussed here. The result states that there exists a positive constant c'', independent of the parameter $0 < \eta < 1$, such that

$$\sum_{a \neq a'} |S_a \cap S_{a'}| \leq c'' \eta^2 X.$$

Choosing η small enough, so that $\eta - c''\eta^2 > 0$, we obtain the desired bound

$$\kappa(P, X) \gg \left| \bigcup_{a \in \mathcal{A}(a_0)} S_a \right| \gg (\eta - c''\eta^2)X \gg X.$$

6 Appendix

For the computations done for the examples in Table 1 the following *Magma* code was used.

```
// Input
D_0 := -4;
f := 5 * 3 * 2;

// Definitions
D := D_0 * f^2;
Q := QuadraticForms(D);
G := AmbiguousForms(Q);
C := ReducedForms(Q);
m := Divisors(f);

// Algorithm
// We cycle through classes
for j := 1 to #C do
       // We are only interested in non-ambiguous classes
       if Order(C[j]) ge 3 then
       print "F =", C[j];
       S := [];
       // S will contain the smallest numbers represented
       // by the product of C[j] with ambiguous forms
       for i := 1 to #G do
              F := C[j]*G[i];
              n := 1;
              while RepresentationNumber(F, n) eq 0 do
                     n := n + 1;
              end while;
              Append(~S, n);
       end for;
       u, ind := Min(S);
       print "the smallest number represented by coset: ", u;

       suma := 0;

       for d := 1 to #m do
              // Definitions for the reduction
              Q_d := QuadraticForms(D / m[d]^2);
```

```
                      G_d := AmbiguousForms(Q_d);
                      C_d := ReducedForms(Q_d);
                      theta_d := QuotientMap(Q, Q_d);

                      S := [];
                      for i:= 1 to #G_m do
                             F := psi_m(C[j])*G_m[i];
                             n := 1;
                             while Gcd(n, m[d]) ne 1 or RepresentationNumber(F,
                                 ↪  n) eq 0 do
                             n := n + 1;
                             end while;
                             Append(~S, n);
                      end for;
                      u_m, ind_m := Min(S);
                      print "m =", m[d], "and u_m =", u_m;
                      suma := suma + EulerPhi(m[d]) / u_m;
               end for;
               print "sum of phi(m)/u_m =", suma, "and then divided by f:", suma
                   ↪ /f;

               end if;
        end for;
```

References

[BF11] Jean Bourgain and Elena Fuchs. “A proof of the positive density conjecture for integer Apollonian circle packings”. In: *J. Amer. Math. Soc.* 24 (2011), pp. 945–967.

[BG06] Valentin Blomer and Andrew Granville. “Estimates for representation numbers of quadratic forms”. In: *Duke Mathematical Journal* 135.2 (2006), pp. 261–302.

[BGR19] Johannes Brantner, Alfred Geroldinger, and Andreas Reinhart. “On monoids of ideals of orders in quadratic number fields”. In: (2019). arXiv: `1901.04528v2`.

[Brü95] Jörg Brüdern. *Einführung in die analytische Zahlentheorie.* Springer, 1995. ISBN: 3-540-58821-3.

[BS66] Z. I. Borevich and I. R. Shafarevich. *Number theory.* Academic Press, 1966. ISBN: 0-12-117850-1.

[Cona] Keith Conrad. *Factoring in quadratic fields.* URL: `https://kconrad.math.uconn.edu/blurbs/gradnumthy/quadraticgrad.pdf` (visited on 06/01/2019).

[Conb] Keith Conrad. *Ideal factorization.* URL: `https://kconrad.math.uconn.edu/blurbs/gradnumthy/idealfactor.pdf` (visited on 06/01/2019).

[Conc] Keith Conrad. *Modules over a PID.* URL: `https://kconrad.math.uconn.edu/blurbs/linmultialg/modulesoverPID.pdf` (visited on 06/01/2019).

[Cox13] David A. Cox. *Primes of the form $x^2 + ny^2$. Fermat, class field theory, and complex multiplication.* 2nd ed. Wiley, 2013. ISBN: 978-1-118-39018-4.

[Del71] F. Delmer. “Sur la somme de diviseurs $\sum_{k\leq x} d[f(k)]^s$”. In: *C. R. Acad. Sci. Paris Sér. A* 272 (1971), pp. 849–852.

[Fuc13] Elena Fuchs. “Counting problems in Apollonian packings”. In: *Bulletin of the American Mathematical Society* 50.2 (2013), pp. 229–266.

[Gau01] Carl Friedrich Gauß. *Disquisitiones Arithmeticae.* Leipzig, 1801. German translation: *Untersuchungen über höhere Arithmetik.* Berlin, 1889.

[Gra+03] R. L. Graham et al. “Apollonian Circle Packings: Number Theory”. In: *J. Number Theory* 100 (2003), pp. 1–45.

[HK13] Franz Halter-Koch. *Quadratic irrationals. An introduction to classical number theory.* CRC Press, 2013. ISBN: 978-1-4665-9183-7.

[Lan02] Serge Lang. *Algebra*. Third Revised Edition. Springer, 2002. ISBN: 978-1-4612-6551-1.

[Mag] *Magma documentation: Binary quadratic forms*. URL: `https://magma.maths.usyd.edu.au/magma/handbook/text/345` (visited on 06/15/2019).

[Sar07] Peter Sarnak. *Letter to Lagarias*. 2007. URL: `http://web.math.princeton.edu/sarnak/AppolonianPackings.pdf`.

[SW06] Zhi-Hong Sun and Kenneth S. Williams. "On the number of representations of n by $ax^2 + bxy + cy^2$". In: *Acta Arithmetica* 122.2 (2006), pp. 101–171.

[Ten95] Gérald Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*. Cambridge University Press, 1995. ISBN: 0-521-41261-7.

[Zag81] Don B. Zagier. *Zetafunktionen und quadratische Körper*. Springer, 1981. ISBN: 3-540-10603-0.